CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 109, No. 4 April 2001

ACOUSTICAL NEWS—USA 1265

USA Meetings Calendar 1270

ACOUSTICAL NEWS—INTERNATIONAL 1271

International Meetings Calendar 1271

ABSTRACTS FROM ACOUSTICS RESEARCH LETTERS ONLINE 1273

BOOK REVIEWS 1274

TECHNICAL PROGRAM SUMMARY 1277

REVIEWS OF ACOUSTICAL PATENTS 1279

MATHEMATICAL ACOUSTICS

Generalized optical theorem for scatterers having inversion symmetry: Applications to acoustic backscattering — Philip L. Marston — 1291

GENERAL LINEAR ACOUSTICS [20]

On the modeling of narrow gaps using the standard boundary element method — Vicente Cutanda, Peter Møller Juhl, Finn Jacobsen — 1296

Wave reflections from duct terminations — A. Selamet, Z. L. Ji, R. A. Kach — 1304

Modeling of high-frequency scattering from objects using a hybrid Kirchhoff/diffraction approach — J. A. Fawcett — 1312

Use of nonsingular boundary integral formulation for reducing errors due to near-field measurements in the boundary element method based near-field acoustic holography — Sung-Chon Kang, Jeong-Guon Ih — 1320

NONLINEAR ACOUSTICS [25]

A generalized Westervelt equation for nonlinear medical ultrasound — G. Taraldsen — 1329

AEROACOUSTICS, ATMOSPHERIC SOUND [28]

Vortex sound generation due to a flow impedance discontinuity on a flat surface — S. K. Tang, K. M. Li — 1334

A computational approach for flow–acoustic coupling in closed side branches — Paul M. Radavich, Ahmet Selamet, James M. Novak — 1343

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 109, NO. 4, APRIL 2001

CONTENTS—*Continued from preceding page*

UNDERWATER SOUND [30]

Matched field processing with data-derived modes — Paul Hursky, W. S. Hodgkiss, W. A. Kuperman — 1355

Acoustic transmission across a roughened fluid–fluid interface — Raymond Lim, Iris C. Paustian, Joseph L. Lopes — 1367

Modeling of bottom backscattering from three-dimensional volume inhomogeneities and comparisons with experimental data — Dan Li, George V. Frisk, Dajun Tang — 1384

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

The explicit secular equation for surface acoustic waves in monoclinic elastic crystals — Michel Destrade — 1398

TRANSDUCTION [38]

Finite-element modeling of lead magnesium niobate electrostrictive materials: Dynamic analysis — Jocelyne Coutte, Jean-Claude Debus, Bertrand Dubus, Régis Bossut — 1403

New piezoelectric polymer for air-borne and water-borne sound transducers — Reiner Kressmann — 1412

STRUCTURAL ACOUSTICS AND VIBRATION [40]

Wave transmission through structural inserts — B. R. Mace, R. W. Jones, N. R. Harland — 1417

Time-domain simulation of damped impacted plates. I. Theory and experiments — Antoine Chaigne, Christophe Lambourg — 1422

Time-domain simulation of damped impacted plates. II. Numerical model and results — Christophe Lambourg, Antoine Chaigne, Denis Matignon — 1433

The influence of substructure modeling on the structural-acoustic response of a plate system — W. Steve Shepard, Jr., Kenneth A. Cunefare — 1448

Noise transmission from a curved panel into a cylindrical enclosure: Analysis of structural acoustic coupling — James K. Henry, Robert L. Clark — 1456

NOISE: ITS EFFECTS AND CONTROL [50]

Psychoacoustic correlates of individual noise sensitivity — Wolfgang Ellermeier, Monika Eigenstetter, Karin Zimmer — 1464

ARCHITECTURAL ACOUSTICS [55]

Predicting speech metrics in a simulated classroom with varied sound absorption — Sylvio R. Bistafa, John S. Bradley — 1474

ACOUSTICAL MEASUREMENTS AND INSTRUMENTATION [58]

Structure-borne noise reduction for an infinite, elastic cylindrical shell — Sung-Hwan Ko, Woojae Seong, Sangwoo Pyo — 1483

PHYSIOLOGICAL ACOUSTICS [64]

Suppression of distortion product otoacoustic emissions and hearing threshold — Martin Pienkowski, Hans Kunov — 1496

DPOAE group delays versus electrophysiological measures of cochlear delay in normal human ears — Ruurd Schoonhoven, Vera F. Prijs, Sandra Schneider — 1503

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 109, NO. 4, APRIL 2001

CONTENTS—*Continued from preceding page*

Wave and place fixed DPOAE maps of the human ear	Richard D. Knight, David T. Kemp	1513
Effects of stimulus frequency and complexity on the mismatch negativity and other components of the cortical auditory-evoked potential	Julia L. Wunderlich, Barbara K. Cone-Wesson	1526

PSYCHOLOGICAL ACOUSTICS [66]

Monaural masking release in random-phase and low-noise noise	David A. Eddins	1538
Measurement of auditory temporal processing using modified masking period patterns	David A. Eddins	1550
Functional magnetic resonance imaging measurements of sound-level encoding in the absence of background scanner noise	Deborah A. Hall, Mark P. Haggard, A. Quentin Summerfield, Michael A. Akeroyd, Alan R. Palmer, Richard W. Bowtell	1559
The effect of basilar-membrane nonlinearity on the shapes of masking period patterns in normal and impaired hearing	Magdalena Wojtczak, Anna C. Schroder, Ying-Yee Kong, David A. Nelson	1571
Gap detection for similar and dissimilar gap markers	John H. Grose, Joseph W. Hall III, Emily Buss, Debora Hatch	1587
The contribution of two ears to the perception of vertical angle in sagittal planes	Masayuki Morimoto	1596
Sensitivity to brief changes of interaural time and interaural intensity	Leslie R. Bernstein, Constantine Trahiotis, Michael A. Akeroyd, Klaus Hartung	1604

SPEECH PRODUCTION [70]

Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees	Ronald C. Scherer, Daoud Shinwari, Kenneth J. De Witt, Chao Zhang, Bogdan R. Kucinschi, Abdollah A. Afjeh	1616
Stochastic models of jitter	Jean Schoentgen	1631
The relationship of vocal tract shape to three voice qualities	Brad H. Story, Ingo R. Titze, Eric A. Hoffman	1651
Influence of emotion and focus location on prosody in matched statements and questions	Marc D. Pell	1668
A study of sentence stress production in Mandarin speakers of American English	Yang Chen, Michael P. Robb, Harvey R. Gilbert, Jay W. Lerman	1681

SPEECH PERCEPTION [71]

Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners	Winifred Strange, Reiko Akahane-Yamada, Rieko Kubo, Sonja A. Trent, Kanae Nishi	1691

BIOACOUSTICS [80]

Evidence for an analytic perception of multiharmonic sounds in the bat, *Megaderma lyra*, and its possible role for echo spectral analysis	Katrin Krumbholz, Sabine Schmidt	1705
Assessment of dolphin (*Tursiops truncatus*) auditory sensitivity and hearing loss using jawphones	Randall L. Brill, Patrick W. B. Moore, Lois A. Dankiewicz	1717
Comparison between visual and passive acoustic detection of finless porpoises in the Yangtze River, China	Tomonari Akamatsu, Ding Wang, Kexiong Wang, Zhou Wei	1723
The acoustic calls of blue whales off California with gender data	Mark A. McDonald, John Calambokidis, Arthur M. Teranishi, John A. Hildebrand	1728

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 109, NO. 4, APRIL 2001

CONTENTS—*Continued from preceding page*

LETTERS TO THE EDITOR

Surface acoustic impedance and causality [28]	Yves H. Berthelot	1736
An updated perspective on basin-scale tomography [30]	John L. Spiesberger	1740
On the group velocity of guided waves in drill strings [40]	Flavio Poletto, José M. Carcione	1743
On the modeling of the vent path in hearing aid systems [66]	B. Rafaely, J. L. Hayes	1747

CUMULATIVE AUTHOR INDEX 1750

NOTES CONCERNING ARTICLE ABSTRACTS

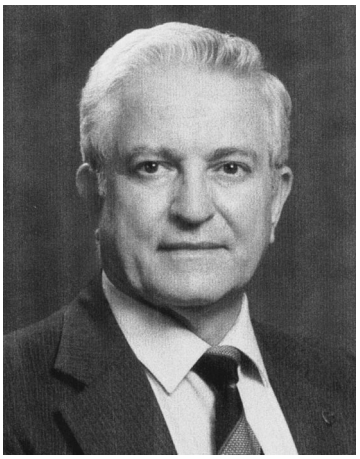
1. **The number following the abstract copyright notice is a Publisher Item Identifier (PII) code that provides a unique and concise identification of each individual published document. This PII number should be included in all document delivery requests for copies of the article.**
2. **PACS numbers are for subject classification and indexing. See June and December issues for detailed listing of acoustical classes and subclasses.**
3. **The initials in brackets following the PACS numbers are the initials of the JASA Associate Editor who accepted the paper for publication.**

ACOUSTICAL NEWS—USA

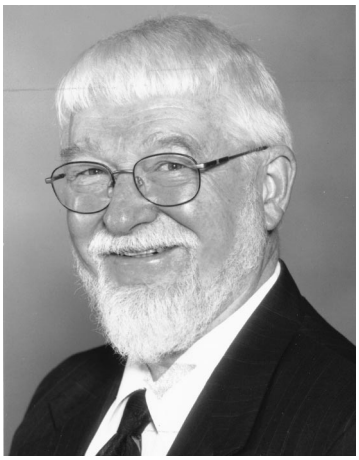
Elaine Moran
Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502

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



Angelo J. Campanella—For contributions to acoustical standards, noise control and architectural acoustics.



W. Jack Hughes—For contributions to the development of electroacoustic transducers.

Notice of Annual Meeting of the Society

The Annual Meeting of the Membership of the Acoustical Society of America will be held on Wednesday, 6 June 2001, at 3:30 p.m. at the Palmer House Hilton Hotel, 17 East Monroe Street, Chicago, IL. All members are encouraged to attend.

CHARLES E. SCHMID
Executive Director

Preliminary Notice: 141st Meeting of the Acoustical Society of America

The 141st Meeting of the Acoustical Society of America (ASA) will be held Monday through Friday, 4–8 June 2001 at the Palmer House Hilton Hotel in Chicago, Illinois. Further information about the meeting can be found on the ASA Home Page at http://asa.aip.org/meetings.html.

Technical Program

The technical program will consist of lecture and poster sessions in all branches of acoustics. Technical sessions will be scheduled Monday through Friday, 4–8 June.

List of Special Sessions

The special sessions described below have been organized by the ASA Technical Committees.

Acoustical Oceanography (AO)

Acoustical instrumentation for water column measurements
Examine new developments in instrumentation for water column oceanography

Benchmarking range dependent numerical models (Joint with Underwater Acoustics)
Looking at the problems created for state-of-the-art numerical models by complicated oceanography and geology

Inverse methods for sub-bottom surveys (Joint with Underwater Acoustics)
Examine inverse techniques for obtaining the geoacoustic structure of the top kilometer of the seabed

Animal Bioacoustics (AB)

Comparative aspects of auditory system development (Joint with Psychological and Physiological Acoustics)
Development of hearing in various species from tadpoles to humans

Information theory analysis of animal acoustic communication (Joint with Signal Processing in Acoustics)
Application of information theory to study animal acoustic communications

Role of signal processing in understanding echolocation (Joint with Signal Processing in Acoustics)
Focuses on direct application of signal-processing theories to behavioral and physiological data from bats and dolphins in an effort to enhance understanding of echo processing

Architectural Acoustics (AA)

Acoustical test facilities (Joint with Engineering Acoustics and Noise)
Poster session showcasing facilities involved in acoustical materials testing, product development, and underwater acoustics

 0001-4966/2001/109(4)/1265/6/$18.00

Concert hall measurements
Measurements of halls, stages, coupled spaces in performance spaces

Halls for music performance . . Another two decades of experience, 1982-2002
Halls completed since the publication of ASA's poster book that covered the period 1962-1982

The first 80 milliseconds in auditoria
On the objective measures and subjective effects of early reflections and diffraction on the perception of sound in concert halls

The sound of baseball (Joint with Noise, Physical Acoustics and Engineering Acoustics)
The game from an acoustics point of view: From the crack of the bat and the roar of the crowd to stadium acoustics and noise in the neighborhood

Walking tour of three Chicago theatres—An acoustical perspective
A look into theatres in the Chicago Loop with a brief technical presentation in each venue

Biomedical Ultrasound/Bioresponse to Vibration (BB)

Beamforming and adaptive aberration correction (Joint with Signal Processing in Acoustics)
Methods to adjust ultrasound imaging systems to maximize image quality

Brad Sturtevant memorial session in lithotripsy (Joint with Physical Acoustics)
Mechanisms, bioeffects, and practical considerations in stone-breaking honoring the memory of Brad Sturtevant

Novel imaging techniques in biomedical ultrasound (Joint with Signal Processing in Acoustics)
Pulse inversion, (sub) harmonic, elasticity, contrast agents, and other new ultrasound imaging methods

Ultrasound bioeffects
Heat- and-cavitation related effects of ultrasound on biological tissues

Education in Acoustics (ED)

Hands on demos with high school students
Demonstrations in many areas of acoustics presented to high school students with opportunities for individual interaction and learning

Take fives for high school presentations; Sharing ideas for teaching acoustics
Five-minute presentations on demonstrations and other educational methods appropriate for high schools

Undergraduate research poster session
Posters by undergraduate students reporting on their research

Engineering Acoustics (EA)

Magnetostrictive materials and transducers (Joint with Underwater Acoustics)
Recent developments and studies of modern magnetostrictive materials and transducers

Transducers for underwater vehicle sonars (Joint with Underwater Acoustics)
Recent developments in transduction technologies for underwater vehicle sonars

Musical Acoustics (MU)

Experimental musical instruments
Studies of acoustic instruments outside the categories of standard concert and folk instruments

New synthesis techniques
Musical sound synthesis for real-time and non-real-time applications

3-D spatialization for music applications
Issues in production, delivery, and composition in spatial formats: Multimedia, home listening, theatre, concerts or art installations

Noise (NS)

Community noise ordinances - Topics for Chicago
The status of local and state noise ordinances is reviewed at the location where one of the first was written: Chicago, Illinois

International noise initiatives
Discussion of I-INCE technical study group initiatives, including noise labels, noise from outdoor recreational activities, noise policies and regulations and noise control for classrooms

Measurement procedures in social studies on soundscapes and annoyance (Joint with Architectural Acoustics and Psychological and Physiological Acoustics)

Motor vehicle interior noise control (Joint with Architectural Acoustics)
The acoustics of cars, vans and buses

Physical Acoustics (PA)

Extraterrestrial acoustics
Applications of acoustical concepts in the solar system and beyond, including cosmology, helioseismology, comet and meteor impacts, etc., as well as global acoustics

Psychological and Physiological Acoustics (PP)

Behavioral studies and physiological correlates
Research using simultaneous psychophysical and physiological measurement

Exploring the dynamics of auditory perception through neuroimaging
Overview of neuroimaging and addressing topics such as within- and cross-modality processes, cognitive influences, and plasticity following pathology

Signal Processing in Acoustics (SP)

Bayesian signal processing approach in acoustics (Joint with Underwater Acoustics)
Bayesian techniques evolve in many estimation and detection problems

Coded signals for acoustics applications (Joint with Underwater Acoustics, Biomedical Ultrasound/Bioresponse to Vibration and Architectural Acoustics)
Development of coded sequences for localization, imaging and communications

Speech Communication (SC)

Dynamics of speech production and perception
One relatively neglected body of theories and data pictures the system of speech communication as one in which the information is conveyed by dynamic changes in the production stage that generates a dynamically changing acoustic signal; the receiver stage retrieves this information by employing dynamic processes optimally adapted for these acoustic signals

Session honoring Harry Levitt (Joint with Psychological and Physiological Acoustics)
Topics covered that Harry Levitt was involved in and to which he contributed

Speech analysis and synthesis techniques and their applications to acoustics (Joint with Signal Processing in Acoustics)
Update on signal processing techniques and how they apply to speech analysis and/or synthesis problems

Structural Acoustics and Vibration (SA)

Complexity and random matrix theory
Response of complex structures, and random matrix techniques that can be applied to this problem

Machinery prognostics (Joint with Signal Processing in Acoustics, Noise and Engineering Acoustics)
Various machinery conditions and faults as well as the processing techniques for detection and classification of faults

David Crighton memorial session (Joint with Physical Acoustics)
Topics covered that David Crighton was involved in and to which he contributed

Underwater Acoustics (UW)

Underwater acoustic communications: Algorithm design and analysis for realistic propagation conditions (Joint with Signal Processing in Acoustics)
Design and analysis (not presentation of data) of algorithms that take complex propagation into consideration

Other Technical Events

History Lectures and Hots Topics Session

A ''Hot Topics'' session sponsored by the Tutorials Committee is scheduled. Also, the Committee on Archives and History will jointly sponsor the next two in a series of lectures on the history of acoustics with the Technical Committee on Psychological and Physiological Acoustics. The lecturers will be Ira J. Hirsh for Psychological Acoustics and Murray Sachs for Physiological Acoustics.

Distinguished Lecture

Bernard Widrow of Stanford University, who is a pioneer in developing algorithms for adaptive beamforming will present a Distinguished Lecture titled ''A microphone array for hearing aids.''

Student Design Competition

The Technical Committee on Architectural Acoustics of the Acoustical Society of America and the National Council of Acoustical Consultants are sponsoring a student design competition which will be judged at the meeting. The purpose of this design competition is to encourage students enrolled in Architecture, Architectural Engineering, and other University curriculums that involve building design to express their knowledge of architectural acoustics and building noise control in the schematic design of portions of a building where acoustical considerations are of primary importance. The submitted designs, which will be displayed at the Chicago ASA Meeting, will be judged by a panel of professional architects and acoustical consultants. An award of $1,000 will be made to the submitter(s) of the design judged ''first honors.'' Four awards of $500 each will be made to the submitters of four entries judged ''commendation.'' Entries may be by individual students or teams of students with a maximum of three members. Students intending to enter this competition must make their intentions known by sending an E-mail message to Bob Coffeen on or before 16 April 2001. Please indicate your name(s), school, faculty advisor, and E-mail address. Entries must be received no later than 18 May 2001. Full details can be found on the ASA web site at asa.aip.org/chicago/design.pdf.

Meeting Program

An advance meeting program summary appears in this issue of JASA and a complete meeting program will be mailed as Part 2 of the May issue. Abstracts will be available on the ASA Home Page (http://asa.aip.org) in April.

Tutorial Lecture

A tutorial presentation on Demonstrations in Acoustics will be given by Thomas D. Rossing of Northern Illinois University, Uwe J. Hansen of Indiana State University, and a team of Chicago-area physics teachers at 7:00 p.m. on Monday, 4 June. Lecture notes will be available at the meeting in limited supply. Those who register by 2 May 2001 are guaranteed to receive a set of notes. To partially defray the cost of the lecture a registration fee is charged. The fee is $15 for registration received by 2 May and $25 thereafter including on-site registration at the meeting. The fee for students with current ID cards is $7 for registration received by 2 May and $12 thereafter, including on-site registration at the meeting. Use the meeting registration form that appeared in the call for papers or online at <http://asa.aip.org/chicago/chicago.html> to register for the Tutorial Lecture.

Short Course on Applied Digital Signal Processing in Acoustics

This short course will be held Friday Afternoon, 8 June, 1:00 to 6:00 p.m. and Saturday, 9 June, 8:00 a.m. to 1:00 p.m.

This short course is designed to develop digital signal processing (DSP) techniques that are applicable to acoustical signal processing problems. The discussions range from basic digital signal processing techniques such as digital filtering and the fast Fourier transform (FFT), to more advanced approaches such as spectral estimation and adaptive processing, to some of the even more sophisticated adaptive and model-based techniques-.The intent is to provide an overview which will provide the participant with additional details of a particular approach to investigate them further. At each stop along the way we will apply the various DSP techniques to an acoustical application. The participant will obtain a basic understanding of the approaches and their applicability discussed from the practictioner's perspective, rather than that of a DSP expert.

The course instructor is James V. Candy is the Chief Scientist for Engineering and Director of the Center for Advanced Signal and Image Sciences at the University of

California, Lawrence Livermore National Laboratory. He received his B.S.E.E. degree from the University of Cincinnati and his M.S.E. and Ph.D. degrees in Electrical Engineering from the University of Florida, Gainesville. Dr. Candy is a Fellow of the IEEE and a Fellow of the Acoustical Society of America (ASA) as well as a member of Eta Kappa Nu and Phi Kappa Phi honorary societies. He has published over 125 journal articles, book chapters, and technical reports as well as written two texts in signal processing, ''Signal Processing: the Model-Based Approach,'' (McGraw-Hill, 1986) and ''Signal Processing: the Modern Approach'' (McGraw-Hill, 1988).

The registration fee is $250.00 and covers attendance, instructional materials and coffee breaks. The number of attendees will be limited so please register early to avoid disappointment. This course was first offered at the last meeting of the Society in Newport Beach, CA and was oversubscribed. Only those who have registered by 2 May will be guaranteed receipt of instructional materials. There will be a $50 discount for registration made prior to 2 May. Full refunds will be made for cancellations prior to 6 November. Any cancellation after 2 May will be charged a $50 processing fee. Use the meeting registration form that appeared in the meeting call for papers or online at <http://asa.aip.org/chicago/chicago.html> to register for this short course.

Student Transportation Subsidies

A student transportation subsidies fund has been established to provide limited funds to students to partially defray transportation expenses to meetings. Students presenting papers who propose to travel in groups using economical ground transportation will be given first priority to receive subsidies, although these conditions are not mandatory. No reimbursement is intended for the cost of food or housing. The amount granted each student depends on the number of requests received. To apply for a subsidy, submit a proposal (e-mail preferred) to be received by 16 April: Elaine Moran, ASA, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502, 516-576- 2360, Fax: 516-576-2377, E-mail: asa@aip.org. The proposal should indicate your status as a student, whether you have submitted an abstract, whether you are a member of ASA, method of travel, whether you will travel alone or with other students, names of those traveling with you, and approximate cost of transportation.

Young Investigator Travel Grants

The Committee on Women in Acoustics is sponsoring a Young Investigator Travel Grant to help with travel costs associated with presenting a paper at the Chicago ASA meeting. This award is designed for young professionals who have completed the doctorate in the past five years (not currently enrolled as a student), who plan to present a paper at the Chicago meeting. Up to $500 in support for travel and lodging costs is available. Applicants should submit a request for support, a copy of the abstract they have submitted for the meeting and a current resume/vita to Eliza (Z.-H) Michalopoulou, Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, NJ 07102; Tel: 973-596-8395; email: elmich@m.njit.edu. Deadline for receipt of applications is 16 April.

Students Meet Members for Lunch

The Education Committee has established a program for students to meet with members of the ASA over lunch. Students who are interested in participating in this program are strongly encouraged to contact Elaine Moran, Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502, USA; Tel: 516-576-2360; E-mail: asa@aip.org prior to the meeting. Please include the area of acoustics you are studying and the days you will be at the meeting. There will also be a sign up sheet available at the registration desk for those students who have not responded

prior to the meeting. Members who wish to participate are also encouraged to contact Elaine Moran. Participants are responsible for the cost of their own meal.

Plenary Session, Awards Ceremony, Fellows Luncheon and Social Events

Complimentary buffet socials with cash bar will be held early on Tuesday and Thursday evenings. The Plenary session will be held on Wednesday afternoon when Society awards will be presented and recognition of Fellows will be announced. A Fellows' Luncheon will be held on Thursday. Each ASA Fellow may bring one guest to the luncheon and tickets may be purchased using the registration form that appeared in the call for papers or online at <http://asa.aip.org/chicago/chicago.html>.

Paper Copying Service

Authors are requested to provide one paper copy of their projection material and/or paper(s) to the Paper Copies Desk upon arrival. The copy should contain material on one side only on 8-1/2 x 11 inch or A4 paper suitable for photocopy reproduction. Copies of available papers will be made for a nominal charge.

Technical Tours

There will be technical tours to The Riverbank Acoustical Laboratores (RAL) in Geneva, Illniois, a Walking Tour of three theatres in the Chicago Loop, and a presentation at Orchestra Hall. Additional information and sign-up sheets will be available at the meeting.

Accompanying Persons Program

Accompanying persons are welcome. A hospitality room, specifically designated for accompanying persons, will be open each day at the Palmer House Hilton. Information will be available about activities in the Chicago area. Spouses and accompanying persons may register in advance using the registration form. The accompanying person registration fee includes the buffet socials on Tuesday and Thursday evenings and refreshments in the accompanying persons hospitality suite during the conference technical sessions. An Accompanying Persons Reception is scheduled for Monday, 4 June at 9:00 a.m. in the hospitality room. A local tour representative will give a brief talk about Chicago and its major points of interest. Advice on tour options and maps will be available. The Chicago area is rich with homes and public buildings designed by the world's most notable architects. The wide variety of museums and cultural institutions throughout the area have something to pique every interest. Outdoor activities are also available with walks through any of the many parks and miles of lakefront walking paths. Places of interest within walking distance of the Palmer House include the Art Institute of Chicago, Grant Park and Buckingham Fountain, Orchestra Hall at Symphony Center, Auditorium Theater, Chicago's theater district, Civic Opera House, Sears Tower, State Street shopping district, Chicago Architecture Foundation and many architectural tours of the city on foot or bus. The Michigan Avenue shopping area, John Hancock Building, historic Water Tower, Navy Pier, boat tours, and the Museum Campus including the Shedd Aquarium/Oceanarium, Field Museum of Natural History, and Adler Planetarium can be reached easily by taxi, public transportation or longer walks. A number of universities and colleges as well as the Museum of Science and Industry can also be reached easily by taxi or public transportation. Further information about Chicago can be found at the following web site: www.chicago.il.org/.

Air and Ground Transportation

Chicago's O'Hare International Airport is served by all major airlines. The airport designation is ORD. Other airports serving Chicago are Midway Airport (MDW) and Meigs Field (CGX).

The Palmer House Hilton Hotel is located 18 miles from Chicago O'Hare International Airport (35-60 minutes). Transportation from the airport to the hotel may be by car, taxi, shuttle service, or public transportation (The El). Taxi fare: $35-$49 (30 to 60 minutes); ''El:'' the elevated train (blue line) leaves from inside the O'Hare terminal parking garage, fare: $1.50 (45 to 60 minutes); Hotel shuttle: Continental Airport Express (60 minutes), fare: $17.50 (serves all the downtown hotels). Midway Airport is located 12 miles from the Palmer House Hilton (30-45 minutes). Taxi fare: $22; ''El:'' the elevated train (orange line) leaves from the parking area at Midway Airport, fare: $1.50 (45 minutes); Hotel shuttle: Continental Airport Express (45 to 60 minutes), fare: $13.50 (serves all the downtown hotels).

Room Sharing

ASA will complete a list of those who wish to share a hotel room and its cost. To be listed, send your name, telephone number, e-mail address, gender, smoker or nonsmoker, by 9 April, preferable by e-mail to asa@aip.org. The responsibility for completing any arrangements for room sharing rests solely with the participating individuals.

Hotel Accommodations

The meeting will be held at the Palmer House Hilton Hotel in Chicago. Please make your reservations directly with the hotel and ask for one of the rooms being held for the Acoustical Society of America. The reservation cut-off date for the special discounted ASA rates is 2 May; after this date the conference rate will not be available. Guest rooms at discounted rates have been reserved for conference participants at the Palmer House Hilton. To reserve a room, please contact the hotel directly. Early reservations are strongly recommended. Note that the conference rates are not guaranteed after 2 May. You must mention the Acoustical Society of America when making your reservation to obtain the special ASA conference rates.

The room rates are $165 Single, $175 Double, $199 Tower Rooms (high-speed internet access, complimentary beverage, bathrobes) plus tax. Send reservations to: The Palmer House Hilton Hotel, 17 East Monroe Street, Chicago, IL 60603, Tel.: 312-726-7500, Toll-free reservations: 800-HILTONS, Fax: 312-917-1707.

Weather

The temperatures in Chicago in June range from 57 to 80 degrees F. It is the second highest month of the year for rainfall, so it might be prudent to bring an umbrella.

Assistive Listening Devices

Anyone planning to attend the meeting who will require the use of an assistive listening device, is requested to advise the Society in advance of the meeting by writing to: Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502, asa@aip.org.

Preregistration and Registration

The registration desk at the meeting will open on Monday morning, 4 June at 7:30 a.m. at the Palmer House Hilton. The preregistration deadline is 2 May. If your registration is not received at the ASA headquarters by 23 May you must register on-site.

Registration fees are as follows:

Category	Preregistration by 2 May	Registration after 2 May
Acoustical Society Members	$225	$275
Acoustical Society Members One-Day	$115	$140
Nonmembers	$275	$325
Nonmembers One-Day	$140	$165
Nonmember Invited Speakers (Note: The fee is waived for these speakers if they attend the meeting on the day of their presentation only)	$225	$275
Students (with current ID cards)	Fee waived	Fee waived
Emeritus members of ASA (Emeritus status pre-approved by ASA)	$35	$45
Accompanying Persons (Spouses and other accompanying persons who will not attend or participate in the technical sessions)	$35	$45

Nonmembers who simultaneously apply for Associate Membership in the Acoustical Society of America will be given a $50 discount off their dues payment for the first year of membership (2001). (Full price for dues: $100). Invited speakers who are members of the Acoustical Society of America are expected to pay the registration fee, but nonmember invited speakers who participate in the meeting for one day only may register without charge. Nonmember invited speakers who wish to participate in the meeting for more than one day will be charged the member registration fee, which will include a one- year membership in the ASA upon completion of an application form at the meeting. Note: A $25 processing fee will be charged for cancelations of registration after 2 May.

To register use the registration form in the Call for Papers or online at <http://asa.aip.org/chicago/chicago.html>

ASA Student Design Competition Awards

The Acoustical Society of America (ASA) Technical Committees on Engineering Acoustics and Architectural Acoustics, in association with the American Loudspeaker Association, JBL Professional, JBL Consumer, Infinity Systems, and Revel sponsored a Loudspeaker Student Design Competition, which was judged by a group of six audio professionals at the 4–8 December 2000 meeting of the ASA in Newport Beach, CA. Four entrants submitted samples and a ''poster board'' presentation detailing the design. Judging was based on a double-blind listening test performed by Sean Olive, a series of standard objective measurements, sample build quality, and the poster board presentation. Three of the entries met all of the design criteria. One of the loudspeakers had a minimum impedance below the allowable limit but complied otherwise.

Ara Baghdassarian (California State University—Northridge) submitted a two-way loudspeaker system with a 6-in. woofer and a 1-in. dome tweeter. Shawn Devantier (University of Victoria, British Columbia) also submitted a two-way system with a 6-in. woofer, but featured a 1-in. tweeter fitted to a waveguide. These two speakers performed exceptionally well in the double-blind listening test. Geoff Christopherson (University of Southern California) submitted the only three-way system with a 10-in. woofer, 5-in. midrange and a 1-in. dome tweeter. Due to the three-way approach this was the most complicated loudspeaker submitted. This model did well in the double-blind test, and the poster presentation explained the design approach well. Dave Tremblay (University of Colorado) submitted a design that was significantly more sensitive than the other entrants. Given its high sensitivity it is not surprising that this sample had the least bass of all the systems and therefore did not perform well in the double-blind test. Otherwise, the design was very well executed and documented. This system featured two 8-in. woofers and a horn tweeter and was fitted with a switch to allow smooth operation on a stand or flush mounted in a wall.

All of the designs impressed the judges and the final scoring was very close. Shawn Devantier was awarded the $1000 first prize. Ara Baghdassarian, Geoff Christopherson, and Dave Tremblay all received $500 ''commendation'' awards.

Addendum: Reviewers of Manuscripts, 1998 and 1999

The Journal previously published in the November 2000 issue (pp. 1955–1960) a list of 1223 persons who had reviewed manuscripts during the calendar years 1998 and 1999. In the preamble to that list it was acknowledged that it might not be complete and readers were asked to send corrections to either Elaine Moran (elaine@aip.org) or Allan Pierce (adp@bu.edu). Subsequently, several associate editors and reviewers have brought to our attention some names that were missing from the original published list and these names are listed here.

Åbom, M.
Ashford, N. A.
Atalla, N.
Balachandran, B.
Berger, E. H.
Berglund, B.
Berry, A.
Blanc-Benon, P.
Bodén, H.
Bolton, J. S.
Botteldooren, D.
Burdisso, R.
Carello, C.
Champoux, J.
Chandler-Wilde, S. N.
Crawford, D. H.
Dancer, A. L.
Di, X.
Don, C. G.
Duhamel, D.
Egli, M.
Engdahl, B.
Erdreich, J.
Everbach, E. C.
Fields, J. M.
Flindell, I. H.
Foker, W.
Franchek, M. A.
Frederickson, C. K.
Gauthier, F.
Gjestland, T. T.
Gottlieb, H. P. W.
Habault, D. M. L.
Hamernik, R. P.
Hansen, C. H.
Heng, C. C.
Heutschi, K.
Horne, J. A.
Horoshenkov, K.
Hothersall, D. C.
Huang, L.
Jain, V. K.
Jean, P.
Job, R. F. S.
Johnson, D. L.
Johnson, M. E.
Kang, Y. J.
Kittinger, E.
Koopman, G. H.
Kostek, T. M.
Landsberger, B. J.
Laroche, C.
Makarewicz, R.
Marino, G.
McGee, J.
Mechel, F. P.
Micheau, P.
Miedema, H. M. E.
Morgan, D. R.
Nelson, P. A.
Öhrström, E.
Oldham, D. J.
Omoto, A.
Pascal, J.-C.
Pearsons, K. S.
Poulsen, T.
Prieto, T. E.
Prince, M. M.
Qiu, X.
Rafaely, B.
Rasmussen, K. B.
Reichard, K. M.
Remington, P. J.
Rienstra, S. W.
Ronneberger, D.
Royster, L. H.
Rylander, R.
Sabatier, J. M.
Saunders, W. R.
Schulte-Fortkamp, B.
Shepherd, K. P.
Silvati, L.
Snyder, S. D.
Solberg, S.

Stusnick, E.
Sullivan, B. M.
Swanson, D. C.
Taherzadeh, S.
Tarnow, V.
Taylor, S. M.
Thompson, D. J.
Timmerman, N. S.
Truman, C. E.
Umnova, O.
Velea, D.
Vos, J.
Yoshida, T.

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.

2001

30 April–3 May — 2001 SAE Noise & Vibration Conference & Exposition, Traverse City, MI [Patti Kreh, SAE Int'l., 755 W. Big Beaver Rd., Suite 1600, Troy, MI 48084, Tel.: 248-273-2474; Fax: 248-273-2494; E-mail: pkreh@sae.org].

4–8 June — 141st Meeting of the Acoustical Society of America, Chicago, IL, Palmer House Hilton Hotel [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

7–9 June — International Hearing Aid Conference VI. Novel Processing and Fitting Strategies, Iowa City, IA [Rich Tyler, Tel.: 319-356-2471, E-mail: rich-tyler@uiowa.edu, WWW: www.medicine.uiowa.edu/otolaryngology/news/news].

9–13 July — 2001 SIAM Annual Meeting, San Diego, CA [Society for Industrial and Applied Mathematics (SIAM), Tel.: 215-382-9800; Fax: 215-386-7999; E-mail: meetings@siam.org; WWW: www.siam.org/meetings/an01/].

15–19 Aug. — ClarinetFest 2001, New Orleans, LA [Dr. Keith Koons, ICA Research Presentation Committee Chair, Music Dept., Univ. of Central Florida, P.O. Box 161354, Orlando, FL 32816-1354, Tel.: 407-823-5116; E-mail: kkoons@pegasus.cc.ucf.edu].

19–24 Aug. — Asilomar Conference on Implantable Auditory Prostheses, Pacific Grove, CA [Michael Dorman, Dept. of Speech and Hearing Science, Arizona State Univ., Tempe, AZ 85287-0102, Tel.: 480-965-3345; Fax: 480-965-0965; E-mail: mdorman@asu.edu].

4–6 Oct. — Ninth Annual Conference on the Management of the Tinnitus Patient, Iowa City, IA [Rich Tyler, Tel.: 319-356-2471; E-mail: rich-tyler@uiowa.edu; WWW: www.medicine.uiowa.edu/otolaryngology/news/news].

7–10 Oct. — 2001 IEEE International Ultrasonics Symposium Joint with World Congress on Ultrasonics, Atlanta, GA [W. O'Brien, Electrical and Computer Engineering, Univ. of Illinois, 405 N. Mathews, Urbana, IL 61801; Fax: 217-244-0105; WWW: www.ieeeuffc.org/2001].

3–7 Dec. — 142nd Meeting of the Acoustical Society of America, Ft. Lauderdale, FL [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org]. Abstract submission deadline: 3 August 2001.

2002

3–7 June — 143rd Meeting of the Acoustical Society of America, Pittsburgh, PA [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

2–6 Dec. — Joint Meeting: 144th Meeting of the Acoustical Society of America, 3rd Iberoamerican Congress of Acoustics, and 9th Mexican Congress on Acoustics, Cancun, Mexico [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org/cancun.html].

ACOUSTICAL NEWS—INTERNATIONAL

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

JASJ(E) changes name

The English language edition of the Journal of the Acoustical Society of Japan, JASJ(E), has changed its name as of 1 January 2001. The journal will be published as *Acoustical Science and Technology* with no further reference to the country of origin. The first issue will not be Vol. 1, No. 1 but, keeping the sequence of the former JASJ(E), Vol. 22, No. 1.

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

The January 2001 issue, Vol. 22, No. 1, contains the following contributions:

J. Hasegawa and K. Kobayashi, ''Blood flow noise transducer for detecting intracranial vascular deformations''
T. Hoshishiba and S. Horiguchi, ''Improved DP matching between a musical score and its performance using interpolation''
F. N. Ucar, Y. Yamakoshi, and E. Yazgan, ''3D image reconstruction algorithm based on subaperture processing for medical ultrasonic imaging''
E. Obataya, Y. Ohno, M. Norimoto, and B. Tomita, ''Effects of oriental laquer (Urishi) coating on the vibrational properties of wood used for the soundboards of musical instruments''

International Meetings Calendar

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

April 2001

9–11 **Acoustical Oceanography**, Southampton. (Fax: +44 1727 850553; Web: www.ioa.org.uk) *8/00*

23–25 **1st International Workshop on Thermoacoustics**, s'Hertogenbosch. (Web: www.phys.tue.nl/index.html) *12/00*

May 2001

21–25 **5th International Conference on Theoretical and Computational Acoustics (ICTCA2001)**, Beijing. (Fax: +1 303 497 3577; Web: www.etl.noaa.gov/ictca01) *12/00*

22–25 **8th International School on Acousto-Optics and Application**, Gdańsk-Jurata. (Fax: +48 58 341 3175; e-mail: fizao@univ.gda.pl) *02/01*

28–31 **3rd EAA International Symposium on Hydroacoustics**, Jurata. (Fax: +48 58 625 4846; Web: www.amw.gdynia.pl/pta/sha2001.html) *12/00*

July 2001

2–5 **Ultrasonics International Conference (UI01)**, Delft. (Fax: +1 607 255 9179; Web: www.ccmr.cornell.edu/~ui01/) *12/00*

2–6 **8th International Congress on Sound and Vibration**, Kowloon, Hong Kong. (Fax: +852 2365 4703; Web: www.iiav.org) *8/00*

23–24 ***2nd Symposium on Underwater Bio-Sonar Systems and Bioacoustics**, Loughborough, UK. (D. Goodson, Dept. of Electronics and Electrical Engineering, Loughborough University, Loughborough, Leicestershire LE11 3TU, UK; Fax: +44 1509 22 7053; Web: sonar-fs.lboro.ac.uk/uag/ioa)

August 2001

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

September 2001

2–7 **17th International Congress on Acoustics (ICA)**, Rome. (Fax: +39 6 4976 6932; Web: www.ica2001.it) *10/98*

10–14 **International Symposium on Musical Acoustics (ISMA 2001)**, Perugia. (Fax: +39 75 577 2255; Web: www.cini.ve.cnr.it/ISMA2001) *10/99*

30–5 ***Conference on Microgravity Transport Processes in Fluid, Thermal, Materials, and Biological Sciences**, Banff, Canada. (Fax: +1 212 591 7441; Web: www.engfnd.org/engfnd/lay.html)

October 2001

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

25–26 **Fall Meeting of the Swiss Acoustical Society**, Wallis/Valias. (Web: www.sga-ssa.ch) *02/01*

November 2001

21–23 **Australian Acoustical Society Annual Meeting**, Canberra. (e-mail: m.burgess@adfa.edu.au; Web: www.users.bigpond.com/Acoustics) *02/01*

March 2002

4–8 **German Acoustical Society Meeting (DAGA 2002)**, Bochum. (Web: www.ika.ruhr-uni-bochum.de) *10/00*

June 2002

4–6 **6th International Symposium on Transport Noise and Vibration**, St. Petersburg. (Fax: +7 812 127 9323; e-mail: noise@mail.rcom.ru) *02/01*

10–14 **Acoustics in Fisheries and Aquatic Ecology**, Montpellier. (Web: www.ices.dk/symposia/) *12/00*

August 2002

19–23 **16th International Symposium on Nonlinear Acoustics (ISNA16)**, Moscow. (Fax: +7 095 126 8411; Web: acs366b.phys.msu.su/isna) *12/00*

September 2002

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

December 2002

2–6 **Joint Meeting: 9th Mexican Congress on Acoustics, 144th Meeting of the Acoustical Society of America, and 3rd Iberoamerican Congress on Acoustics**. Cancún. (e-mail: sberista@maya.esimez.ipn.mx; Web: asa.aip.org) *10/00*

April 2004

5–9 ***18th International Congress on Acoustics (ICA2004)**, Kyoto, Japan. (Web: ica2004.or.jp)

 0001-4966/2001/109(4)/1271/1/$18.00

BOOK REVIEWS

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

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

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

Acoustics of Wood

Voichita Bucur

CRC Press, Boca Raton, FL, 1995.
xi+284 pp. Price: $264.95 ISBN: 0849348013.

This is a valuable book, covering many aspects of wood and its acoustical properties. From the Preface, we learn that the book is intended to ''be helpful for scientifically educated persons wishing to know more about wood, as a natural anisotropic composite material.'' Its aim is ''to present a comprehensive account of progress and current knowledge in wood acoustics,'' covering the last 30 years or so. (Note that the book was published in 1995.) In fact, there are about 800 bibliographic items, although (inconveniently) these are grouped (alphabetically) at the end of each chapter. This bibliography is an excellent resource in itself because the literature is very scattered.

This is neither a textbook nor a research monograph. It is written as an extended survey, giving brief descriptions of the cited literature, together with many tables and figures. This compilation of data and acquired experience is also very useful.

The book is divided into three parts: Environmental Acoustics (2 chapters, 24 pages), Material Characterization (3 chapters, 100 pages), and Quality Assessment (5 chapters, 132 pages). There is also a brief introductory chapter, an Appendix listing the wood species cited in the book, and a subject index.

Part I considers the propagation of acoustic waves in forests and the effects of wood on architectural acoustics. Typical applications are the use of trees to attenuate sound and the use of wood as an acoustic insulator in concert halls.

Part II begins with a summary of wave propagation in a homogeneous anisotropic elastic solid. Wood is usually modeled as an orthotropic solid, with three distinguished symmetry planes. The elastic constants for various woods are given. They can be obtained using dynamic experiments, and such methods are discussed. Much of this is not specific to wood, but there are exceptions. For example, how do wave speeds depend on moisture content? Wave attenuation through wood is also considered, using a viscoelastic model.

The orthotropic model is local; the symmetry planes may vary with position. One way to deal with this is to attempt to give a global characterization based on the assumption that the properties of a piece of wood ''can be represented by an equivalent homogeneous anisotropic continuum'' (p. 91); this is taken up in Chapter 5. Many results from the literature are summarized. Another way to proceed is to change the local model, assuming cylindrical orthotropy, for example, or a layered composite structure. Applications of acoustic microscopy to determine local structure are described. At the local level, wood is certainly a complicated anisotropic material. We do not have good continuum models in place. Chapter 6 concludes by noting that there is ''a whole hierarchy of questions concerning the interaction of ultrasonic waves with the complex structural organization of wood species'' (p. 130). We can hope that this book will stimulate further research on these questions.

As wood is a natural product, its properties will vary from sample to sample. Part III is concerned with assessing the quality of wooden objects, with emphasis on acoustical properties and methods of nondestructive evaluation. Chapter 7 is devoted to wood as used in musical instruments, including violins, guitars, and pianos; most work in this area is experimental. Applications of ultrasonic methods to inspect live trees, logs, planks, poles, and wooden composites are reviewed in Chapter 8; the aim here is usually to search for defects of various kinds.

The properties of wood can change significantly due to environmental variations. For example, wood can decay and it can crack. Chapter 9 opens with the observation that ''the scientific literature lacks sufficient data on the dynamic, nondestructive characterization of wood'' (p. 197). This chapter reviews how wood is changed by the environment and how these changes can be detected using ultrasonics. Acoustic-emission techniques can also be used (Chapter 10) to monitor changes in wooden structures, such as the growth of cracks or termite activity. The last chapter discusses briefly how high-energy ultrasonic waves can be used to modify the internal structure of wood.

In summary, the author has done an admirable job, collecting, organizing, and reviewing the disparate literature on most aspects of the *Acoustics of Wood*.

PAUL A. MARTIN
Department of Mathematical and Computer Sciences
Colorado School of Mines
Golden, Colorado 80401

Fundamentals of Physical Acoustics

David T. Blackstock

John Wiley & Sons, 2000.
560 pp. Price: $90.00 hc ISBN: 0471319791.

This text is an outgrowth of a two-semester course that has introduced a generation of beginning graduate students in engineering and physics at the University of Texas at Austin to acoustics. Its 15 chapters cover the fundamentals of acoustics of fluid media in the clear and deliberate manner that has come to be expected of the author. A natural question is how does this book compare to standards such as Kinsler, Frey *et al.*; Morse; Morse and Ingard; or Pierce? Although some general comments along these lines will be offered at the end, the bulk of this review will be devoted to describing the content of the book in some detail so that the readers might answer this question for themselves.

After a brief, nonmathematical discussion of what constitutes a wave, Chapter 1 moves rapidly to a statement (without derivation) of the 1D wave equation and to its general, time-domain solutions. Time-domain methods are emphasized to a large degree in this book. In fact, with a few exceptions, use of steady-state, complex exponential solutions is postponed until Chapter 4. Blackstock explains that his preference for time-domain methods is founded in his experience in nonlinear acoustics research, in which the time-domain approach is sometimes the simplest. While this approach may be more difficult for many beginning students than steady-state harmonic analysis, it does offer the opportunity to introduce concepts, such as characteristics, that many presentations omit. In addition, it emphasizes the physical basis of some phenomena that tends to get lost in steady-state analyses, e.g., whether radiated pressure is proportional to the velocity of a source or its acceleration.

The chapter continues with several examples of waves on strings, including d'Alembert's general solution to 1D problems with general bound-

 0001-4966/2001/109(4)/1274/3/$18.00

ary conditions. Following this discussion, fairly traditional derivations of the 1D linear wave equation are presented for the electrical transmission line, the string, and sound. The concept of characteristic impedance is introduced here. Next comes a brief discussion of 1D spherical and cylindrical waves, including the introduction of the velocity potential. This section ends with sound generated by a pulsating sphere. The chapter concludes with a section on various concepts commonly encountered in acoustics, including the Fourier transform, impedance, intensity and power, and levels.

Blackstock suggests that the logical progression of topics would skip Chapter 2 and proceed to Chapters 3–9 which continue the development of plane-wave acoustics. Chapter 2 contains a rigorous derivation of the three-dimensional wave equation in dissipative fluids, including body forces. It begins with a detailed derivation of the equation of continuity, conservation of momentum, conservation of energy, the equation of state, and the entropy equation. The next section is devoted to developing the nonlinear wave equation. The three-dimensional nonlinear wave equation for a lossless ideal gas is derived in the first part of this section, followed by the exact solution for finite-amplitude plane progressive waves. The section continues with a discussion of second harmonic distortion of plane waves in a lossless tube driven by a sinusoidal excitation. Chapter 2 concludes with the derivation of the small-signal wave equation. Cases considered are the wave equation in a lossless medium at rest, moving with constant velocity, and in a gravitational field; a thermoviscous medium; and a relaxing medium.

The development of 1D systems resumes with Chapter 3, which treats reflection and transmission of normally incident plane waves. The first half of the chapter introduces the pressure and sound power reflection and transmission coefficients, and discusses rigid and pressure release surfaces, and changes in cross sectional area. The remainder of the chapter is devoted to three examples, including a rectangular pulse in finite-length tube with different terminations, the pressure disturbance generated in a shock tube, and the pressure disturbance generated by a bursting balloon. The treatment in this chapter is in the time-domain, so the results are valid for arbitrary waveforms. Characteristic diagrams are used to facilitate the analysis.

The discussion of normal incidence continues in Chapter 4, which is restricted to time harmonic waves. This chapter contains a wide range of topics, some of which one might not expect to find in a chapter devoted to normally incident waves. The discussion begins with standing waves in tubes with arbitrary termination impedance. Impedance tube calculations are used as an example. The next topic, perhaps unexpectedly, is lumped-element approximations. The analysis is introduced by taking the short (compared to the wavelength) tube limit of results found in the previous section. Particular cases treated are the short closed cavity, the end correction for an open tube, the short open cavity, the Helmholtz resonator, and the orifice. The chapter continues with examples of application of lumped-element analysis to side branches, filters, and the probe tube microphone. The discussion then turns to reflection and transmission of harmonic waves in three-medium problems. The case of constant cross section and different cross sections are both treated. The chapter concludes with a lumped-element treatment of wall transmission loss.

Chapter 5 deals with transmission phenomena at oblique incidence. Two derivations of Snell's law and the law of specular reflection are offered. The first is based on requiring the trace velocities of the incident, reflected and transmitted waves to all be equal. The second derivation is based on continuity of pressure and normal component of velocity at the interface. Both derivations assume arbitrary waveforms. At the end of the second derivation, the formulas for the pressure reflection and transmission coefficients for both normal (discussed in Chapter 4) and oblique incidence are shown to have a common form if written in terms of acoustic impedance. Next, special cases including angle of intromission, critical angle, and grazing incidence are addressed. Harmonic time variations are assumed for the discussion of critical angle and evanescent waves. Chapter 5 continues with the transmission of sound through panels at oblique incidence. The discussion assumes that the medium is the same on both sides of the panel, the more general case being saved for an end-of-chapter problem. The first case treated is the one in which the mass of the panel dominates the transmission phenomena. The mass law for time harmonic waves is derived. The second case, leading to the coincidence effect, takes the panel stiffness into account by modeling the panel as a lossless, Euler–Bernoulli plate. A very clear derivation of the transmission coefficient of time harmonic waves is presented. The chapter concludes with a brief discussion of sound transmission in composite walls.

Chapter 6 is devoted to normal modes in Cartesian coordinates. The four sections in this chapter treat the vibrating string, the vibrating rectangular membrane, the rectangular enclosure, and the rectangular waveguide, respectively. The discussion emphasizes the use of the method of separation of variables for solving the wave equation, and the use of boundary conditions to find the eigenfunctions and eigenvalues. Fourier analysis is used to determine the arbitrary coefficients from initial conditions. The waveguide section, intended to be only a brief introduction, includes phase and group velocity, cutoff, traveling and evanescent waves, and modal excitation. A more complete discussion of waveguides is postponed until Chapter 12.

The treatment of one-dimensional phenomena continues with horns in Chapter 7. The chapter begins with a derivation of the equation of continuity and conservation of momentum in ducts of variable cross section, leading to Webster's horn equation, followed by a discussion of the exponential horn. The next section deals with impedance, transmitted power and transmission factor. The conical horn is treated as a spherical wave system and its properties are compared and contrasted to those of the exponential horn. Chapter 7 continues by introducing the WKB method and using it to treat the catenoidal horn. The derivation of Webster's horn equation at the beginning of the chapter was accomplished by eliminating particle velocity from the fundamental equations, the result being a differential equation for pressure. Chapter 7 concludes by noting that had pressure been eliminated in favor of particle velocity, the resulting differential equation would not have had the form of Webster's horn equation. However, a change of variables from particle velocity to volume velocity does lead to an equation in the correct form.

The subject of Chapter 8 is propagation in stratified media. After a brief overview of the static properties of the atmosphere and ocean, the wave equation for vertical propagation of plane waves in derived including the effects of gravity. Vertical propagation through an isothermal atmosphere is discussed and its relation to propagation in an exponential horn pointed out. Having related the wave equation for vertical propagation and Webster's horn equation, a general solution is obtained through use of the WKB method. From there, the chapter turns to an elementary discussion of ray theory. First, the radius of curvature of a ray is found in terms of the sound speed gradient and Snell's constant. This result is applied to find the paths of rays in a medium with a linear sound speed profile and in a medium comprised of two regions with different linear profiles. The chapter concludes with a discussion of travel time along a ray path.

The discussion of plane-wave phenomena concludes in Chapter 9 with a detailed discussion of absorption and dispersion in dissipative fluids. After introducing the complex propagation constant, the development addresses bulk absorption in thermoviscous and relaxing fluids, and thermoviscous boundary layer losses. The treatment of bulk absorption begins with the derivation of the wave equation for plane waves in a viscous fluid. Use of time harmonic solutions leads to the dispersion relation. The usual (relatively) low frequency approximation is used to arrive at expressions for the absorption coefficient and phase speed. The viscous fluid is a simple enough medium so that this section can demonstrate how, by introducing the appropriate physics into the conservation equations, the dispersion relation can be found in a straightforward manner. The same procedure is applied next to a thermally conducting fluid. However, the discussion is complicated by the need to deal with energy conservation in a much more detailed manner than in the case of a viscous fluid. Perseverance leads to a fourth order dispersion relation, where it is pointed out that two of the roots correspond to ''thermal'' waves and two correspond to sound waves. After taking the appropriate roots, the discussion arrives at the absorption coefficient and the phase speed. The viscous and thermal effects are then combined to lead to the absorption coefficient for a thermoviscous fluid. Attention turns to absorption and dispersion in a fluid with a single relaxation process. The physics is truncated in this section (for instance, the equation of state is stated, not derived) in favor of moving briskly to the dispersion relation. The section concludes with a discussion of the frequency dependence of absorption and dispersion in a relaxing fluid. An overview of boundary layer absorption and dispersion follows. Viscous and thermal boundary layers and hydraulic diameter are introduced. The dispersion relation and expressions for the absorption coefficient and phase speed are given in the limit that the boundary layers are small compared to the hydraulic diameter. Chapter 9 concludes with a summary of the important equations developed in the chapter. The discussion of relaxation phenomena is supplemented with Appendix B, which contains formulas and graphs for absorption of sound in the atmo-

sphere and the ocean. The discussion of boundary layer absorption is supplemented with Appendix C, which contains a detailed treatment of the viscous boundary layer.

The remainder of the text, Chapters 10 through 15, addresses nonplane-wave phenomena. As mentioned previously, these chapters would follow sequentially after the development of the three-dimensional wave equation in Chapter 2. Chapters 10 and 11 are similar to Chapter 6 (dealing with problems in rectangular coordinates), in that they address a particular coordinate system, spherical and cylindrical, respectively. Each begins with the solution of the wave equation using the separation of variables method, and by introducing the needed special functions and their relevant properties. Each chapter then proceeds to address bounded systems and radiation problems. After introducing Legendre polynomials, spherical Bessel and Hankel functions, and spherical harmonies in Chapter 10, the bipolar pulsating sphere is used to demonstrate the expansion of functions in a Legendre series. The solution of Legendre's equation and discussion of Legendre polynomials is supplemented with Appendix D, which includes the solution of Legendre's equation by power series. The next section treats standing waves in spherical enclosures, using the pressure release sphere as the primary example. The remainder of Chapter 10 is devoted to a detailed discussion of monopole and dipole radiation introduced through the multipole expansion of the solution of the wave equation in spherical coordinates. Monopole radiation from both simple sources and finite-radius spherical sources is treated, the latter leading to discussion of radiation mass loading. The dipole is modeled as both two simple sources of opposite phase and as a translating sphere.

Chapter 11 includes a discussion of (cylindrical) Bessel and Hankel functions and the Fourier–Bessel series. The free and driven vibrations of a circular membrane are treated, including discussion of the pie-shaped membrane. The chapter moves on to standing waves in cylindrical enclosures and concludes with radiation problems, including the infinite pulsating cylinder, the infinite vibrating wire, and solution for a finite line source (neglecting end effects).

Chapter 12 extends some of the material in Chapters 6 and 11 in a more thorough treatment of waveguides. Three cases are treated: the rectangular and circular cross section waveguides, and a waveguide formed between two parallel infinite planes. This chapter considers the driven waveguide and modal excitation in more detail than was done previously.

Chapter 13 is devoted to radiation from a baffled piston. The Rayleigh integral is introduced and specialized to time harmonic piston vibration. The example of a baffled ring piston is used as an illustrative example. The treatment is extended to the farfield and on-axis radiation from a baffled circular piston. The analysis of the time harmonic circular piston concludes with the derivation of the radiation impedance and piston functions, which are tabulated along with the directivity function for a baffled circular piston in Appendix E. Transient radiation from a circular piston is addressed from the time-domain perspective. Two central points are made regarding the on-axis pressure. One is that the spatial dependence can be explained in terms of the superposition of two plane waves, one emanating from the center of the piston and the other from the edge. The other is that the pressure is proportional to the acceleration of the piston, not the velocity. An operator formulation is used to analyze the farfield behavior of transient radiation. It is shown that the off-axis radiation is proportional to higher-order derivatives of the piston velocity. Chapter 13 concludes with a discussion of the farfield radiation from a shaded circular piston.

Chapter 14 deals with diffraction. The Helmholtz–Kirchhoff integral is derived in the frequency-domain, starting from Green's theorem. The integral is then converted into the time-domain by taking its inverse Fourier transform. The result is then used to analyze diffraction of normally incident plane waves through a circular aperture. The discussion is reminiscent of parts of Chapter 13, in that it describes the diffracted pressure field in terms of the superposition of a direct wave and an edge wave. Diffraction of spherical waves is discussed briefly. Attention turns next to the reflection of plane waves from a rigid circular disk. The chapter concludes with diffraction from complementary objects through use of Babinet's principle. The discussions are aided with reference to measurements of diffraction of spark-generated N-waves, clearly motivated from the author's research experience. Neither Fresnel diffraction nor Fourier approaches to diffraction are addressed.

The final chapter is a brief discussion of time harmonic, far field radiation from arrays. The chapter begins with definitions of the amplitude directivity function, the intensity directivity factor, and the directivity index. The example of a monopole next to a rigid plane is used to illustrate these concepts. Next an array composed of two equal-amplitude point sources is presented. The discussion includes the case of arbitrary relative phase to introduce beam steering. The results for the two-element array are generalized for a linear, N-element array, with particular emphasis on the directivity function. The continuous line array is discussed next by taking the results for the N-element linear array in the limit of very large N and very small element spacing. The chapter concludes with a statement of the product theorem.

The text contains approximately 270 problems of varying degrees of difficulty and approximately 140 references. The only appendix yet to be mentioned is Appendix A, which tabulates the elastic properties of selected solids, liquids and gases.

While some readers may appreciated a broader scope, from the student's perspective the limited scope is likely to be one of the books' strengths. Rather than fitting more topics into a given number of pages, space is devoted to a thorough development of fundamental concepts and analytical methodology. Derivations are complete enough so that intermediate steps are easily scaled. In addition, examples are interspersed to demonstrate the analytical techniques. The problems provide ample opportunity for readers to challenge their understanding and to extend their investigation of acoustics beyond the topics presented in the text. As such, this text fills the gap between undergraduate-level books and advanced graduate-level texts that assume thorough knowledge of advanced mathematics. At the risk of being overly brief, the level of presentation is somewhat higher than that of Kinsler and Frey, although not as much so with the addition of new material in the new fourth edition. However, Blackstock's book does not contain the introduction to vibrations, nor the discussions of transduction, hearing, environmental noise, or architectural acoustics included in K&F. The level is somewhat lower than that of either Morse and Ingard or Pierce, and the scope is certainly narrower. Of the four ''standards'' mentioned at the outset, this book is closest in level and scope to Morse. In many respects, Blackstock is the contemporary version of Morse. It can serve much the same function as an introduction to fundamental concepts and mathematical formalism. As such, Blackstock is a valuable source for students beginning their study of advanced acoustics.

ANTHONY A. ATCHLEY
Graduate Program in Acoustics
The Pennsylvania State University
State College, Pennsylvania 16804

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
IBRAHIM M. HALLAJ, *Wolf Greenfield & Sacks P.C., 600 Atlantic Avenue, Boston, Massachusetts 02210*
DAVID PREVES, *Songbird Hearing, Inc., 5 Cedar Brook Drive, Cranbury, New Jersey 08512*
WILLIAM THOMPSON, JR., *601 Glenn Road, State College, Pennsylvania 16803*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*

6,104,670

43.25.Ba ULTRASONIC HARMONIC IMAGING SYSTEM AND METHOD

John A. Hossack *et al.*, assignors to Acuson Corporation
15 August 2000 (Class 367/7); filed 2 March 1995

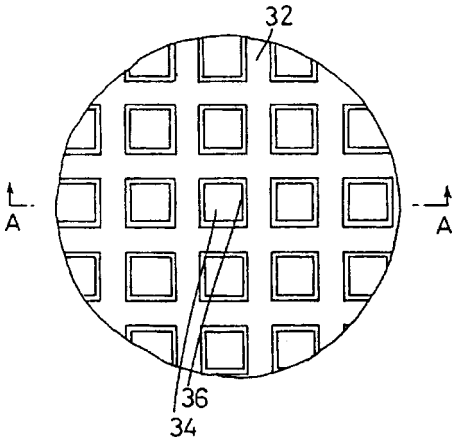
A method and apparatus for harmonic imaging is given that calls for transmission of ultrasound at a fundamental frequency *f*, focused at two depths along a scan line, then reception of a harmonic of *f* from a target in the scan line associated with the two depths. The target may or may not include contrast agents.—IMH

6,095,978

43.30.Jx METHOD OF MANUFACTURING AN ULTRASONIC PROBE AND THE ULTRASONIC PROBE AND AN ULTRASONIC IMAGING APPARATUS

Yasuhito Takeuchi, assignor to GE Yokogawa Medical Systems, Limited
1 August 2000 (Class 600/443); filed in Japan 11 November 1997

A low-impedance ultrasonic probe is described, having a plurality of holes **34** in the thickness dimension of the piezoelectric material **32**. The



interior surfaces of the holes are coated to form electrodes **36**. An imager using an array of such structures is provided.—IMH

6,097,670

43.30.Sf METHOD AND APPARATUS FOR TRACKING OBJECTS UNDERWATER

Robert L. Johnson, assignor to Battelle Memorial Institute
1 August 2000 (Class 367/120); filed 27 May 1998

The position of an underwater object, e.g., a fish, is tracked by tagging the object with an acoustic transmitter that preferably operates at about 200 kHz. The position of the transmitter is then determined by processing the outputs from one or more fixed-position, split-beam hydrophones whose positions are known.—WT

6,108,270

43.30.Tg TORPEDO SEEKER HEAD HAVING DIRECTIONAL DETECTION INDEPENDENT OF FREQUENCY

Martin L. DePoy II, Cleveland, New York
22 August 2000 (Class 367/124); filed 6 July 1999

A torpedo seeker head, which provides directional information about a target relative to three predetermined orthogonal axes, is realized from a combination of three orthogonally oriented directional hydrophones and one omni-directional hydrophone. Phase information from the signals generated by any pair of the directional hydrophones and by the omni-directional hydrophone is used to compute the angle of arrival of an incident acoustical wave in the plane defined by the principal response axes of those two directional hydrophones. Hence the three possible pairs provide orientation directions in three orthogonal planes.—WT

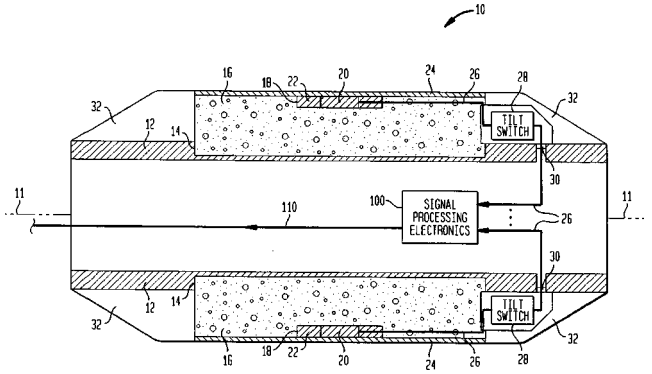
6,128,250

43.30.Yj BOTTOM-DEPLOYED, UPWARD LOOKING HYDROPHONE ASSEMBLY

Robert J. Reid *et al.*, assignors to the United States of America as represented by the Secretary of the Navy
3 October 2000 (Class 367/153); filed 18 June 1999

A hydrophone assembly for deployment on the bottom of a body of water consists of a cylindrical shell **12** that houses appropriate electronics **100** in its interior while a belt of vibration isolation material **16** over the exterior of the shell supports a number of hydrophone elements **20** spaced around the outermost circumference of the assembly. Each element **20** has an associated tilt switch **28** so that, when deployed, only those few elements near the top of the assembly, whose principal acoustic response axes are

 0001-4966/2001/109(4)/1279/10/$18.00



directed towards the surface of the body of water, are activated. Each hydrophone element **20** is a laminated structure wherein the active piezoelectric material is chosen to be a piezorubber composite because of bandwidth considerations.—WT

6,128,251

43.30.Yj SOLID MARINE SEISMIC CABLE

Louis W. Erath *et al.*, assignors to Syntron, Incorporated
3 October 2000 (Class 367/154); filed 16 April 1999

A seismic streamer comprises an interior cable, a surrounding strength member, an overlaying foam flotation member, and a surrounding acoustically transparent jacket. Elongated channels are cut out of the flotation layer and a number of ferroelectric sensor elements are mounted in these channels which are then backfilled with polyurethane potting material.—WT

6,099,670

43.35.Ty ULTRASONIC BONDING METHOD

John W. Louks *et al.*, assignors to 3M Innovative Properties Company
8 August 2000 (Class 156/73.1); filed 11 September 1998

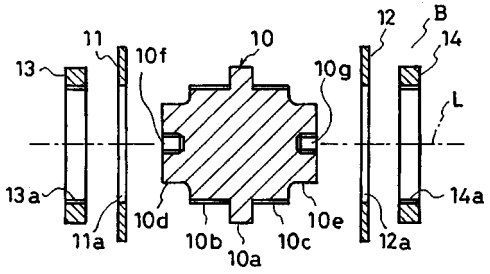
A method is described for bonding a thermoplastic and another material, utilizing an ultrasonic horn and an anvil. Peak pressures of at least 2.068×10^7 N/m^2 and acoustic velocities of no more than 4.72 m/s are used. The materials are quenched prior to reducing the pressure to allow for bonds to form.—IMH

6,098,514

43.35.Zc ULTRASONIC VIBRATION CUTTER

Shigeru Sato and Ryoichi Ishii, assignors to Ultex Corporation
8 August 2000 (Class 83/425.3); filed in Japan 12 April 1996

An ultrasonic vibration cutter having a resonator **10** onto which diamond cutting disks **11** and **12** are changeably-mounted with attachment nuts



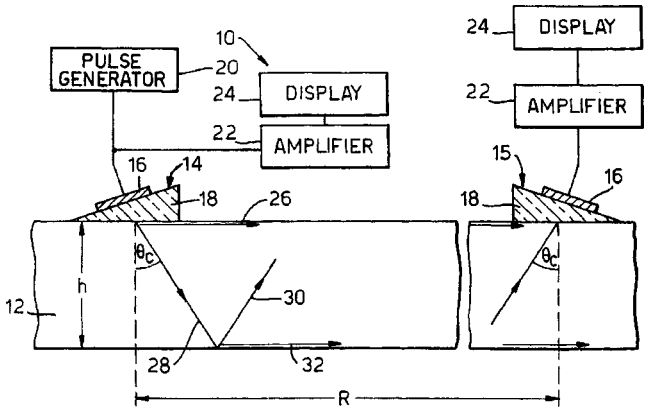
13 and **14**, thus allowing for the exchange of the cutters.—IMH

6,105,431

43.35.Zc ULTRASONIC INSPECTION

Colin Duffill and Maurice Geoffrey Silk, assignors to Aea Technology PLC
22 August 2000 (Class 73/624); filed in the United Kingdom 18 May 1995

A method for nondestructive testing of a plate or a pipe wall is given using two ultrasonic transducers. A series of pulses, emitted from one of the transducers at an angle suitable to induce both compressional and shear



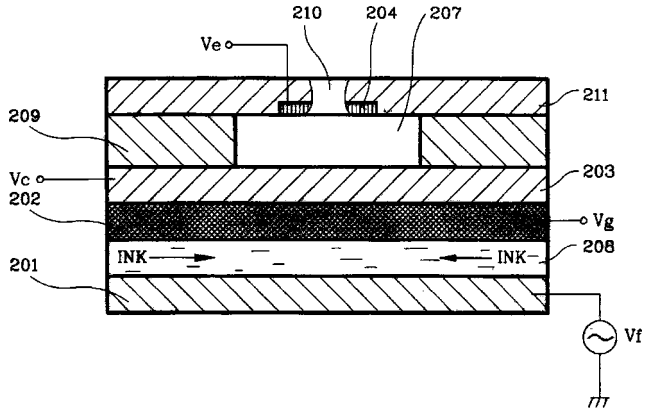
waves in the sample, reaches the second transducer. The received signals may then be used to determine the existence of flaws in the tested sample.—IMH

6,106,103

43.35.Zc INK-JET SPRAYING DEVICE AND METHOD USING ULTRASONIC WAVES

Byung-sun Ahn, assignor to SamSung Electronics Company, Limited
22 August 2000 (Class 347/55); filed in Korea 11 January 1997

An ink-jet printing device nozzle assembly is provided, which uses an ultrasonic vibrating plate **201** to separate pigment particles in the ink. The particles are passed through a charged mesh grid **202** into a holding chamber



207. The particles are then ejected from a nozzle opening **210** under the influence of an electric field between points **203** and **204**.—IMH

6,105,408

43.35.Zc NON-DESTRUCTIVE TESTING OF STEEL DURING ROLLING

Horst Scharlemann, assignor to Georgsmarienhutte GmbH
22 August 2000 (Class 72/31.07); filed in Germany 24 July 1998

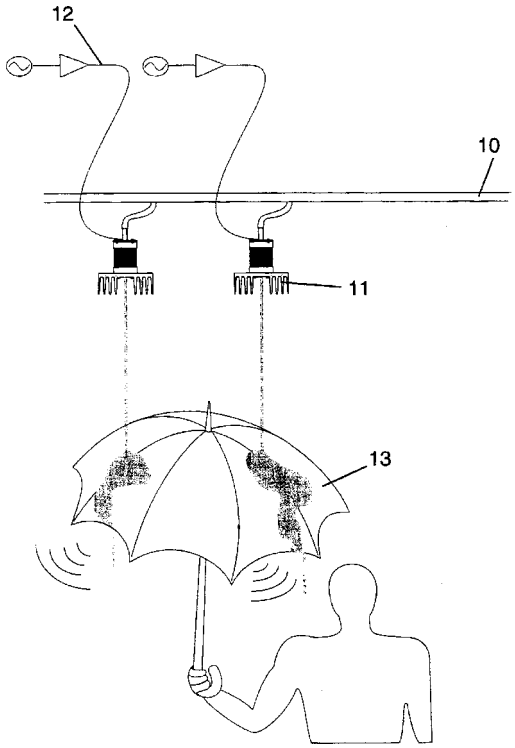
A method is presented for nondestructive testing of hot deformable steel while still being rolled in a roller. An ultrasonic pulse-echo system applied at the point of rolling is used to detect interior flaws by the echoes they produce.—IMH

6,095,889

43.38.Ar INTERACTIVE ENTERTAINMENT DEVICE

Paul M. Demarinis, San Francisco, California
1 August 2000 (Class 446/397); filed 25 January 1999

Nozzle **11** is vibrated by an audio signal from power amplifier **12**.



Modulated raindrops fall on umbrella **13**, thus producing audible sound waves.—GLA

6,102,860

43.38.Hz ULTRASOUND TRANSDUCER FOR THREE-DIMENSIONAL IMAGING

Matthew Mooney, assignor to Agilent Technologies, Incorporated
15 August 2000 (Class 600/443); filed 24 December 1998

An imaging transducer assembly having a two-dimensional array of elements disposed on a surface having curvature in two orthogonal directions is presented. The system also includes a control circuit for sweeping an aperture across the array in two 2-D patterns to yield 3-D imaging data.—IMH

6,073,723

43.38.Ja ACOUSTIC DAMPING MATERIAL

Anthony Gallo, Brooklyn, New York
13 June 2000 (Class 181/151); filed 5 June 1998

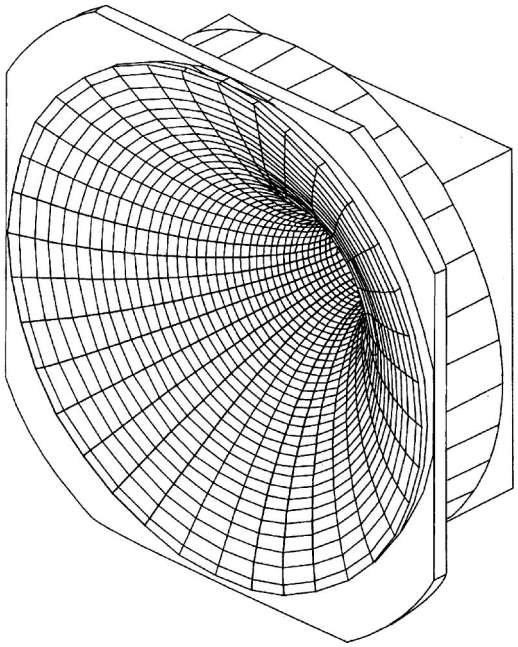
This damping material is intended mainly for use in sealed box loudspeaker systems. Plastic film is loosely rolled into small, multi-layered cylinders. These are used to stuff the speaker back chamber. The result, we are told, is exceptionally flat low-frequency response with a 20% improvement in attack rate and a 30% improvement in decay rate.—GLA

6,079,514

43.38.Ja ACOUSTIC HORN TRANSDUCER WITH A CONIC TYPE DIFFUSER HAVING AND EXPONENTIAL PROFILE IN WOOD

Giuseppe Zingali, assignor to Zingali S.N.C
27 June 2000 (Class 181/152); filed in Finland 30 May 1995

This very short patent is easy to understand. The inventor believes that the ideal high-frequency horn should (a) be made of wood without knots, and (b) have a circular section, fairly rapid exponential flare, and relatively large mouth. With a few more qualifications, the patent could have defined a somewhat vague but legitimate invention. Instead, it is an example of a disturbing trend in writing patent applications: make the first claim as broad



as possible, preferably encompassing everything in the known universe. In the case at hand, Claim 1 omits the reference to wood and simply defines any circular exponential horn having an axial length less than its mouth diameter. Needless to say, prior art is abundant. What purpose this serves, or why the claim was allowed, remains a mystery.—GLA

6,081,602

43.38.Ja ARRAYABLE TWO-WAY LOUDSPEAKER SYSTEM AND METHOD

John D. Meyer *et al.*, assignors to Meyer Sound Laboratories, Incorporated
27 June 2000 (Class 381/99); filed 19 August 1997

Mechanical provisions are described for stacking modular, two-way loudspeaker systems. Individual units can be rotated around the axis of the high-frequency horns' lateral acoustic origin.—GLA

6,094,497

43.38.Ja LOUDSPEAKER SYSTEM HAVING BACK PRESSURE EQUALIZATION

David S. Mohler, assignor to Lucent Technologies, Incorporated
25 July 2000 (Class 381/386); filed 9 April 1999

The chronicles of misguided audio engineering are littered with attempts to ''equalize'' acoustic loading between the front and rear surface of a loudspeaker cone. The first entry for the year 2000 has been assigned to Lucent Technologies Inc.—GLA

6,111,962

43.38.Ja REVERBERATION SYSTEM

Takahashi Akio, assignor to Yamaha Corporation
29 August 2000 (Class 381/63); filed in Japan 17 February 1998

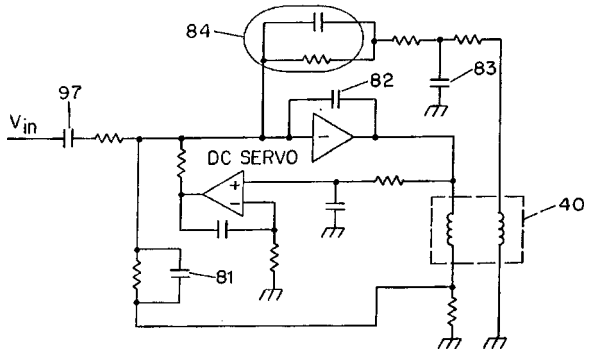
Multiple loudspeakers—say, 12—are spaced around the walls of a listening room. The patent explains how digital delays and filters are computer-adjusted to synthesize a different, virtual acoustic room.—GLA

6,104,817

43.38.Lc SPEAKER AND AMPLIFIER SYSTEM

Chih-Shun Ding, Buena Park, California
15 August 2000 (Class 381/96); filed 12 December 1996

A combination of velocity and current feedback is used to achieve predictable, stable extension of loudspeaker low-frequency response, even



in a vented box system. The patent document includes extensive analysis of loudspeaker loading and feedback circuitry.—GLA

6,111,957

43.38.Md APPARATUS AND METHOD FOR ADJUSTING AUDIO EQUIPMENT IN ACOUSTIC ENVIRONMENTS

Samuel L. Thomasson, assignor to Acoustic Technologies, Incorporated
29 August 2000 (Class 381/15); filed 2 July 1998

Supersonic signals in air have long been used for vehicle detection, intrusion alarms, and the like. In theory, a modulated supersonic carrier could faithfully transmit a high-fidelity audio signal without degradation by loudspeaker distortion or room acoustics. The inventor's earlier two patents propose a kind of simulcast in which a conventional audio signal and its supersonic equivalent are reproduced at the same time by the same loudspeaker. In this latest variant, the two are compared *in situ* and the information used to automatically equalize a playback system.—GLA

6,094,279

43.38.Ne SYSTEM AND PROCESS FOR NON-PERCEPTIBLY INTEGRATING SOUND DATA INTO A PRINTED IMAGE

Peter P. Soscia, assignor to Eastman Kodak Company
25 July 2000 (Class 358/1.9); filed 28 October 1997

Audio signals are invisibly encoded onto a printed image by means of infrared dye. The patent discloses a well thought-out system and process for incorporating sound information on color prints.—GLA

6,067,361

43.38.Vk METHOD AND APPARATUS FOR TWO CHANNELS OF SOUND HAVING DIRECTIONAL CUES

Michael Kohut *et al.*, assignors to Sony Corporation and Sony Electronics, Incorporated
23 May 2000 (Class 381/17); filed 16 July 1997

During the process of motion picture sound mixing, left and right surround signals are processed using head-related transfer functions to provide front–rear cues. Bandwidth is not reduced and no special playback provisions are required.—GLA

6,111,958

43.38.Vk AUDIO SPATIAL ENHANCEMENT APPARATUS AND METHODS

Robert Crawford Maher, assignor to EuPhonics, Incorporated
29 August 2000 (Class 381/17); filed 21 March 1997

Numerous circuits have been developed to enhance the subjective spaciousness of two-channel stereo playback. This clearly written patent describes a method which avoids spectral coloration by providing a generally flat transfer function from input to output. Those involved in this field will want to download a copy.—GLA

6,111,819

43.38.Zp REDUCED MECHANICAL COUPLING INTERLINK FOR SPATIALLY EXTENDED HYDROPHONES

Antonio L. Deus III *et al.*, assignors to the United States of America as represented by the Secretary of the Navy
29 August 2000 (Class 367/173); filed 4 October 1999

An interlink for connecting adjacent fiber optic hydrophone elements, in a coaxial line array of cylindrical shaped elements, is fashioned from a flexible open cell foam material. Hence the connecting optical fiber between adjacent hydrophone elements, which is supported by this cylindrically shaped interlink, should not be susceptible to noise sources associated with mechanical resonances of a more rigid interlink structure. Also, this flexible interlink allows the array to experience considerable bending during handling without damage to the interlink or the optical fiber.—WT

6,111,820

43.38.Zp SEMI-RIGID LOW-NOISE INTERLINK FOR SPATIALLY EXTENDED HYDROPHONES

Gregory H. Ames, assignor to the United States of America as represented by the Secretary of the Navy
29 August 2000 (Class 367/173); filed 4 October 1999

An interlink for connecting adjacent fiber optic hydrophone elements, in a coaxial line array of cylindrical-shaped elements, consists of a semi-rigid tube of a plastic material which surrounds a cylinder of flexible open cell foam material. A helical slot is cut through the wall of the plastic tube so that the optical fiber that connects adjacent hydrophone elements actually is supported by the open cell foam cylinder. This patent can be viewed as a combination of some concepts discussed in two related patents (United States patents 6,118,733 and 6,111,819). The semi-rigid tube enables this compound interlink structure to be a more rigid connection between adjacent hydrophone elements without compromising the desirable features of the foam material interlink.—WT

6,118,733

43.38.Zp INTERLINK FOR SPATIALLY EXTENDED HYDROPHONES

Gregory H. Ames, assignor to the United States of America as represented by the Secretary of the Navy
12 September 2000 (Class 367/173); filed 4 October 1999

An interlink for connecting adjacent fiber optic hydrophone elements, in a coaxial line array of cylindrical-shaped elements, consists of a semi-rigid tube of a plastic material in which a helical slot has been cut through the wall of the tube to make the tube substantially more flexible. An additional helical groove has been cut into the remaining material to accommodate and support the optical fiber that connects adjacent hydrophone elements. The interlink is intended to both protect the optical fiber during handling and deployment and to desensitize the portion of the fiber that rests upon the interlink structure during operation.—WT

6,082,486

43.66.Ts ARTICLE FOR COLLECTING SOUND FOR EARS

Young S. Lee, Los Angeles, California
4 July 2000 (Class 181/136); filed 1 February 1999

Cupping a hand to one's ear improves high-frequency sound pickup. A pair of reflectors on a headband works even better and leaves both hands free.—GLA

6,128,392

43.66.Ts HEARING AID WITH COMPENSATION OF ACOUSTIC AND/OR MECHANICAL FEEDBACK

Hans Leysieffer and Hans Delfs, assignors to Implex Aktiengesellschaft Hearing Technology
3 October 2000 (Class 381/318); filed in Germany 23 January 1998

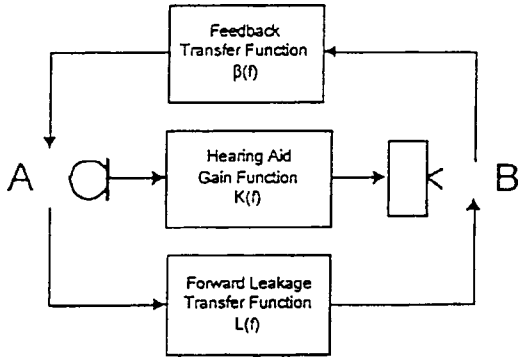
A pulse generator is used to determine the transfer function of the feedback path for either an implanted or conventional hearing aid. The impulse response is used to set the coefficients of an FIR digital filter to form a compensator to prevent the occurrence of mechanical or acoustic feedback oscillation. Assumptions are made that the feedback path transfer function remains relatively constant and that the effect of external signals on the impulse response measurement is negligible.—DAP

6,134,329

43.66.Ts METHOD OF MEASURING AND PREVENTING UNSTABLE FEEDBACK IN HEARING AIDS

Shawn X. Gao and Sigfrid D. Soli, assignors to House Ear Institute
17 October 2000 (Class 381/60); filed 5 September 1997

The open loop transfer function of a hearing aid fitting is derived by three measurements of closed loop transfer functions without breaking the path. The open loop transfer function determines the maximum gain achievable before oscillation and frequencies at which gain might be attenuated or



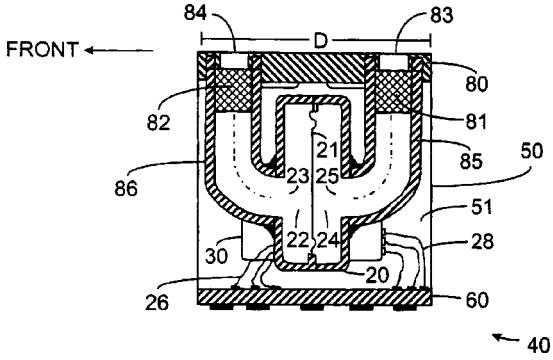
phase shifted to avoid oscillation. The amount of gain attenuation required to avoid oscillation is recommended. The methodology does not require a two-channel signal analyzer and utilizes real ear probe microphone measurement equipment commonly used in hearing aid dispensing practices.—DAP

6,134,334

43.66.Ts DIRECTIONAL MICROPHONE ASSEMBLY

Mead C. Killion *et al.*, assignors to Etymotic Research, Incorporated
17 October 2000 (Class 381/356); filed 31 December 1996

A wearer-switchable omnidirectional/directional hearing aid microphone capsule is described in which the frequency response in the two modes is equalized using two acoustic delay/damping resistors. In addition to the single acoustic resistor normally used in the rear port to make a microphone directional, another acoustic resistor of appropriate value to



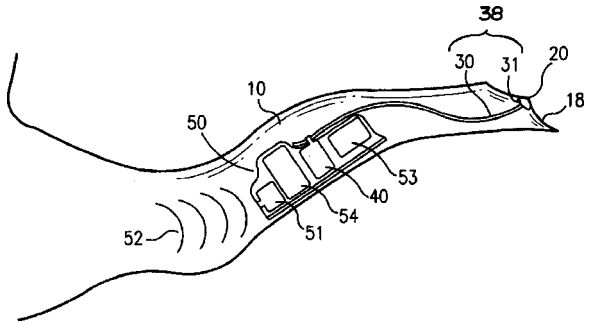
produce the required delay/response is placed in the front sound inlet associated with the front volume of the directional microphone. Electrical equalization circuitry is used in combination with the acoustic resistors to provide the desired smooth frequency response in directional mode.—DAP

6,137,889

43.66.Ts DIRECT TYMPANIC MEMBRANE EXCITATION VIA VIBRATIONALLY CONDUCTIVE ASSEMBLY

Adnan Shennib and Richard C. Urso, assignors to Insonus Medical, Incorporated
24 October 2000 (Class 381/328); filed 27 May 1998

A direct vibrational drive to the tympanic membrane mechanism is described for which a vibratory transducer is packaged with microphone, energy efficient amplifier, and battery. This electronics package is placed for long-term use in the unoccluded ear canal with a thin, vibration-conducting



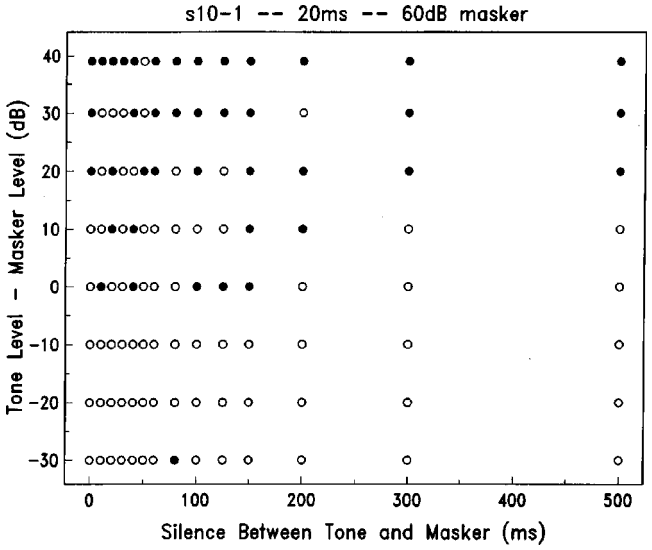
filament connected to the vibratory transducer and extending to the tympanic membrane. The design ensures that only a weak static pressure is exerted by either a removable or permanently attached filament assembly on the tympanic membrane.—DAP

6,109,107

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

Beverly A. Wright *et al.*, assignors to Scientific Learning Corporation
29 August 2000 (Class 73/585); filed 7 May 1997

An impairment in language learning ability known as specific language impairment (SLI) has been related to particular deficits in perceptual processing of rapid transitions in the speech signal. Further, it has been found that these deficits can be reduced with proper training. This patent



presents a system for diagnosing and training individuals with such deficits. The figure shows a performance map with open circles marking perceptual failures in cases which clearly indicate an SLI condition.—DLR

6,109,923

43.70.Dn METHOD AND APPARATUS FOR TEACHING PROSODIC FEATURES OF SPEECH

Martin Rothenberg, assignor to Syracause Language Systems
29 August 2000 (Class 434/185); filed 24 May 1995

The prosodic, or nonsegmental, features of speech include pitch, loudness, rate of speech, and glottal source variations. Such features can be difficult to learn, especially for an adult foreign language learner and, in some cases, for the native-language speaker as well. This patent describes a device to display the prosodic structure of a speech sample and to produce a sound which limits or eliminates the articulatory or segmental content so as to allow the user to concentrate on the prosodic information.—DLR

6,113,393

43.70.Dn RAPID AUTOMATIZED NAMING METHOD AND APPARATUS

Graham Neuhaus, Houston, Texas
5 September 2000 (Class 434/178); filed 29 October 1997

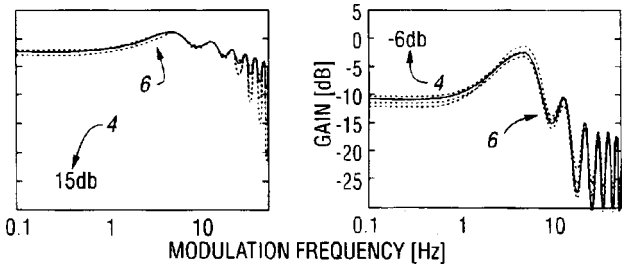
Automated testing is becoming popular as a way to assess reading and other verbal skills, particularly in kindergarten through second grade students. This patent describes a computer system for measuring the test response times based on detected energy levels. Various types of data are presented showing the expected variation in response times for students in different grades and with different levels of ability.—DLR

6,098,038

43.72.Dv METHOD AND SYSTEM FOR ADAPTIVE SPEECH ENHANCEMENT USING FREQUENCY SPECIFIC SIGNAL-TO-NOISE RATIO ESTIMATES

Hynek Hermansky and Carlos M. Avendano, assignors to Oregon Graduate Institute of Science and Technology
1 August 2000 (Class 704/226); filed 27 September 1996

This speech noise reducer separates the signal into subbands and filters each subband ideally according to the SNR value measured in that subband. The figure shows precomputed filter responses for subbands with SNR values of +15 and −6 dB (as shown by **4**). The frequency scales are shown as



differences from the band center frequency. In general, bands with lower SNR levels pass less energy at greater differences from the band center. The filtered subband signals are recombined to produce the output speech signal.—DLR

6,115,687

43.72.Ew SOUND REPRODUCING SPEED CONVERTER

Naoya Tanaka and Hiroaki Takeda, assignors to Matsushita Electric Industrial Company, Limited
5 September 2000 (Class 704/269); filed in Japan 11 November 1996

The basic task in increasing the rate of recorded speech is to locate similar portions of the speech waveform, one or more of which can be cut out without drastically changing the signal. This patent adds little to that background. The search for similarity of waveform segments uses the output of a linear prediction coding system as a starting point.—DLR

6,098,041

43.72.Ja SPEECH SYNTHESIS SYSTEMS

Tatsuro Matsumoto, assignor to Fujitsu Limited
1 August 2000 (Class 704/260); filed in Japan 12 November 1991

This speech synthesis server provides multiple speech synthesis services to network clients, typically on a local-area network. The client may use any or all of the conversions from text to phonetic codes, phonetic codes to acoustic parameters, or acoustic parameters to speech waveforms. A variety of coding techniques are available, such as ADM, ADPCM, PARCOR, line spectral pairs, and various forms of LPC vocoding. Resources such as dictionaries and pronunciation rules are available.—DLR

6,098,042

43.72.Ja HOMOGRAPH FILTER FOR SPEECH SYNTHESIS SYSTEM

Duy Quoc Huynh, assignor to International Business Machines Corporation
1 August 2000 (Class 704/260); filed 30 January 1998

This patent describes a series of processing steps to generate the correct pronunciations for homographic words, such as ''lead,'' ''live,'' or ''wind.'' The first level procedure is based on syntactic patterns, such as whether the word follows an article or an adjective, and the part-of-speech category (or categories) of the word itself. Several layers of rules, known as probability rules, certainty rules, and special rules, deal with varying grammar structures. Additional levels of fallback procedures rely on the statistics of locally occurring words to correct any remaining ambiguities.—DLR

6,101,470

43.72.Ja METHODS FOR GENERATING PITCH AND DURATION CONTOURS IN A TEXT TO SPEECH SYSTEM

Ellen M. Eide and Robert E. Donovan, assignors to International Business Machines Corporation
8 August 2000 (Class 704/260); filed 26 May 1998

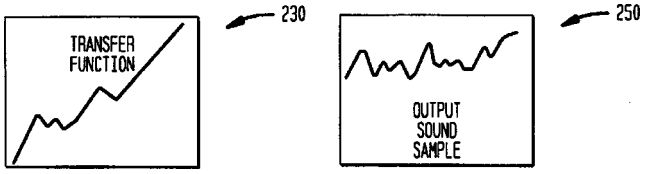
This speech synthesis procedure uses pairs of lexical stress and pitch contour information collected from natural speech to compute appropriate pitch contours for the synthetic speech. Natural speech from several speakers is analyzed to collect pitch contour fragments corresponding to moments of determined lexical stress levels. The pitch levels are normalized such that they can be applied to a synthesized pitch contour at the moments corresponding to lexical stress levels as determined by dictionary lookup of the synthesized words.—DLR

6,101,469

43.72.Ja FORMANT SHIFT-COMPENSATED SOUND SYNTHESIZER AND METHOD OF OPERATION THEREOF

Steven D. Curtin, assignor to Lucent Technologies, Incorporated
8 August 2000 (Class 704/258); filed 2 March 1998

The patent describes a sound synthesizer particularly suited to the synthesis of speech with pitch variations, such as required for inflectional patterns. A periodic input signal is transformed to the frequency domain and modified by some sort of nonlinear operation. The details of these modifi-



cations are rather sketchy, but are characterized as following changes to the fundamental while maintaining a desired spectral envelope. It is hard to see what is novel here.—DLR

6,108,630

43.72.Ja TEXT-TO-SPEECH DRIVEN ANNUNCIATION OF CALLER IDENTIFICATION

Timothy Augustines Keuchler *et al.*, assignors to Nortel Networks Corporation
22 August 2000 (Class 704/270); filed 23 December 1997

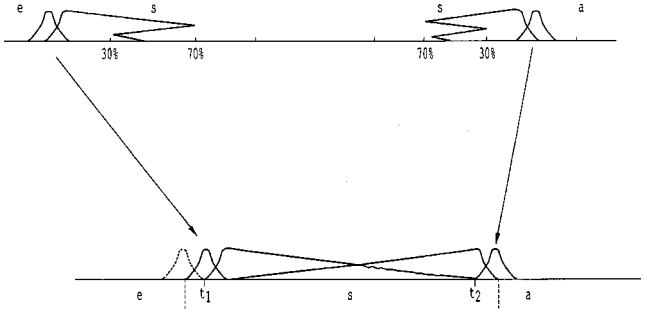
Several previous systems have been proposed for voice announcement of caller information upon identification by a caller ID system on the telephone. Typically, the caller ID information itself is not sufficient for a suitable voice message, requiring a database access for additional information. The system described in this patent includes the database lookup step, but also performs several kinds of ''tricks,'' such as counting vowels and vowels surrounded by consonants, to determine which portions of the information to play as a voice announcement.—DLR

6,112,178

43.72.Ja METHOD FOR SYNTHESIZING VOICELESS CONSONANTS

Jaan Kaja, assignor to Telia AB
29 August 2000 (Class 704/267); filed in Sweden 3 July 1996

One problem which occurs when creating synthetic speech from recorded fragments of natural speech is that if a given fragment is repeated too often, the repetition will be noticeable. This patent presents a novel solution



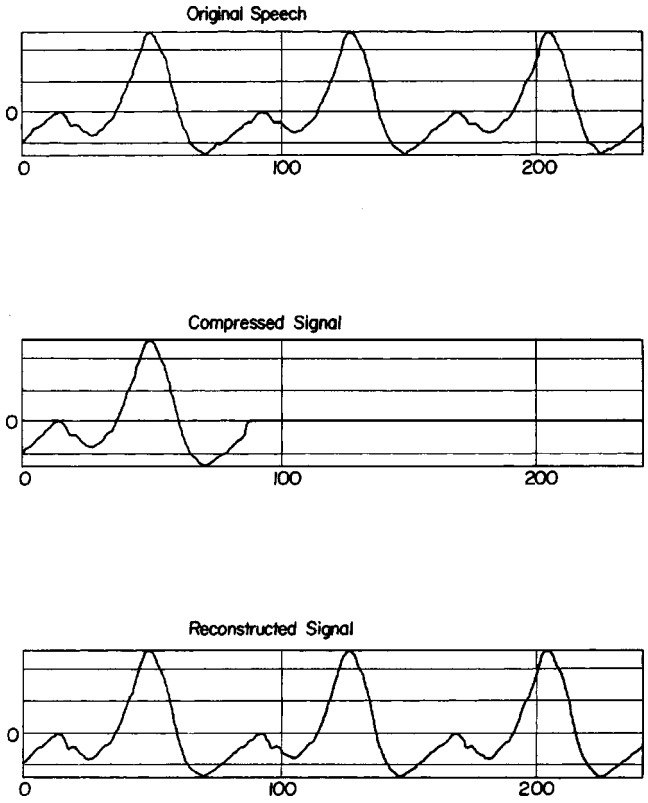
by time-reversing the copied fragment about half of the time. The method is only applicable for noise portions, such as voiceless consonant sounds, but could be useful in those cases.—DLR

6,101,463

43.72.Kb METHOD FOR COMPRESSING A SPEECH SIGNAL BY USING SIMILARITY OF THE F_1/F_0 RATIOS IN PITCH INTERVALS WITHIN A FRAME

Sang Hyo Lee *et al.*, assignors to Seoul Mobile Telecom
8 August 2000 (Class 704/207); filed in Republic of Korea 12 December 1997

The patent describes a novel method of speech compression. For each voice frame of the incoming speech, the ratio of formant 1 to the pitch frequency is computed. A particular F1/F0 ratio implies a more or less repeatable waveform shape during each glottal cycle. If a sequence of



frames has a sufficiently similar set of F1/F0 ratios, then the speech within those frames can be shortened by making cuts at corresponding waveform points.—DLR

6,108,626

43.72.Kb OBJECT ORIENTED AUDIO CODING

Luca Cellario *et al.*, assignors to Cselt-Centro Studi E Laboratori Telecomunicazioni S.p.A.; Robert Bosch GmbH
22 August 2000 (Class 704/230); filed in Italy 27 October 1995

The goal this patent claims to have met is a difficult one, that of segmenting an audio signal into portions corresponding to each of several types of objects. To the extent this goal can be reached, then distinctive routing, bitrate adjustments, and signal coding conditions can be applied for each type of object. In this case, the task has been much simplified by a preset list of available objects, each of which has fairly distinctive sound characteristics. The objects include voice, music, noise, and signaling tones.—DLR

6,098,040

43.72.Ne METHOD AND APPARATUS FOR PROVIDING AN IMPROVED FEATURE SET IN SPEECH RECOGNITION BY PERFORMING NOISE CANCELLATION AND BACKGROUND MASKING

Marco Petroni and Stephen Douglas Peters, assignors to Nortel Networks Corporation
1 August 2000 (Class 704/234); filed 7 November 1997

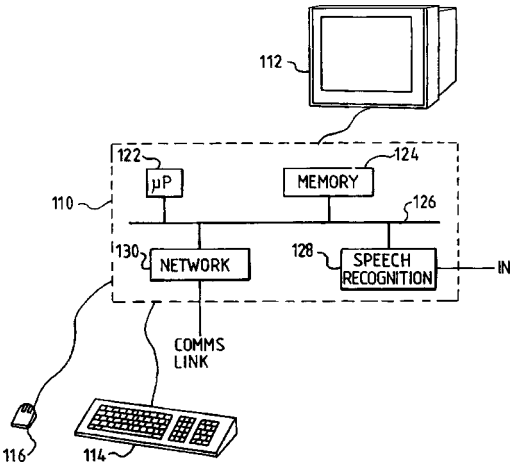
This patent describes the application of a well-known method of improving recognition of noisy speech for use in an automated voice-activated telephone dialing system. Fourier transforms of the input speech frames are separated into barklike bands and classified by energy level into nonspeech, voiceless, and voiced frames. The energy levels of the speech frames are reduced by the amount of the nonspeech levels. These adjusted frames are then used for recognizer training.—DLR

6,100,882

43.72.Ne TEXTUAL RECORDING OF CONTRIBUTIONS TO AUDIO CONFERENCE USING SPEECH RECOGNITION

Richard Anthony Sharman and Paul Stuart Adams, assignors to International Business Machines Corporation
8 August 2000 (Class 345/320); filed in the United Kingdom 19 January 1994

The patent describes a voice-operated workstation suitable for online audio conferences. While the local station is transmitting speech, a recognition system is also converting the outgoing speech to text. The text is dis-



played for the local user, transmitted to the other stations, and also retained as a conference record.—DLR

6,101,241

43.72.Ne TELEPHONE-BASED SPEECH RECOGNITION FOR DATA COLLECTION

Susan J. Boyce *et al.*, assignors to AT&T Corporation
8 August 2000 (Class 379/88.01); filed 16 July 1997

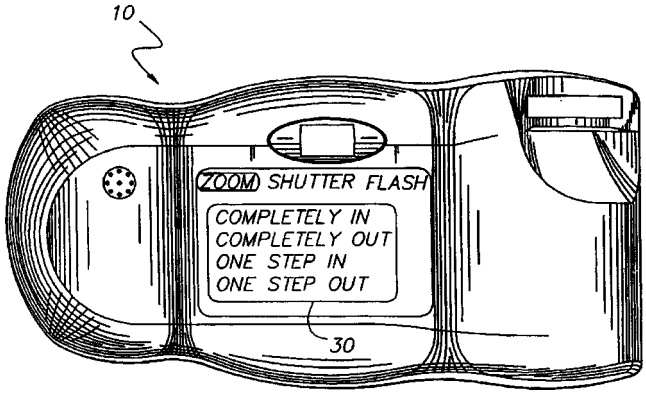
The patent describes an automated system for the collection of large amounts of speech data by telephone. Such data would then be used for training telephone speech recognition systems for various uses. The key elements are the recognition of out-of-grammar items and various verbal cues which tend to indicate speaker confusions or other difficulties. Based on such results, the system retunes the structure of prompts so as to improve the interaction with the person.—DLR

6,101,338

43.72.Ne SPEECH RECOGNITION CAMERA WITH A PROMPTING DISPLAY

Bryan D. Bernardi *et al.*, assignors to Eastman Kodak Company
8 August 2000 (Class 396/287); filed 9 October 1998

This voice-operated digital camera uses a limited vocabulary recognition system in order to reduce battery power needs. To compensate for the limited voice capability, a display provides immediate response to the user,



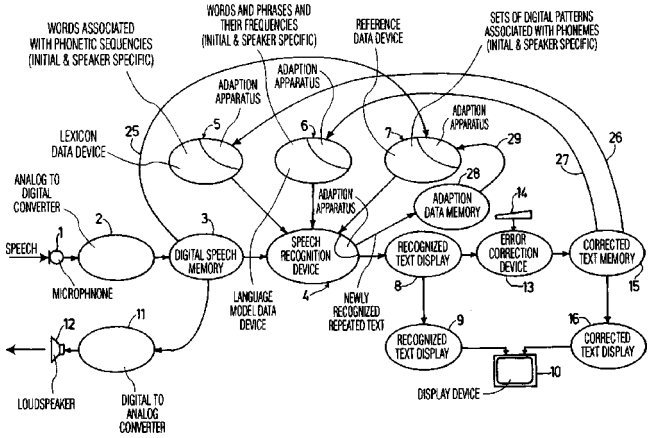
improving the nature of the interaction. There is no discussion of whether the user can reprogram the interactive voice sequences.—DLR

6,101,467

43.72.Ne METHOD OF AND SYSTEM FOR RECOGNIZING A SPOKEN TEXT

Heinrich Bartosik, assignor to U.S. Philips Corporation
8 August 2000 (Class 704/235); filed in European Patent Office 27 September 1996

This is an extreme case of a patent which goes on for several pages and says nothing. The flows of speech and text data through a typical recognition system are described in great detail. Coming to the recognition technology itself, the patent states that the ''technology is known per se and



is not described further hereinafter.'' The system provides a display of recognized and reference texts, allowing the user to make manual corrections, after which a second recognition pass is performed.—DLR

6,101,468

43.72.Ne APPARATUSES AND METHODS FOR TRAINING AND OPERATING SPEECH RECOGNITION SYSTEMS

Joel M. Gould *et al.*, assignors to Dragon Systems, Incorporated
8 August 2000 (Class 704/251); filed 13 November 1992

This patent presents a variety of strategies for improving the operation of a speech recognition system running on a user's personal computer. By typing one or a few letters, lists of alternate words can be suggested, altering the recognition results. During a training phase, additional words can be spoken, again altering the recognition probabilities. Other words identified by moving the text cursor can be used to alter recognition results. Information about other programs running on the same computer can be used to alter word lists. Acoustic models from various sources can be combined to improve results.—DLR

6,101,471

43.72.Ne APPARATUS AND METHOD FOR PROCESSING VOICES, AND STORAGE MEDIUM USING ATTRIBUTE DATA

Yoshihiro Shintaku and Shuzo Kugimiya, assignors to Sharp Kabushiki Kaisha
8 August 2000 (Class 704/270); filed in Japan 9 October 1997

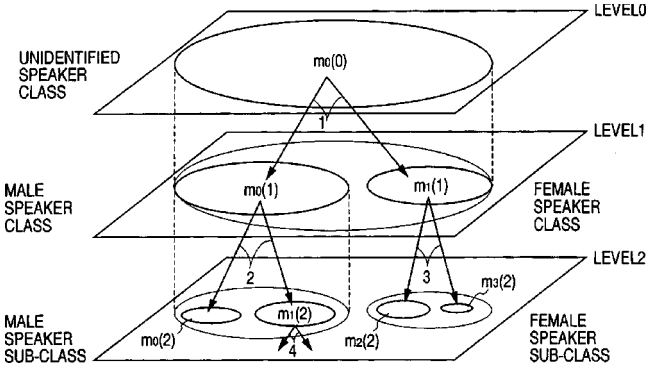
This speech recognition system, intended for use with a personal scheduling device, includes specific capabilities for recognizing date and time information in the spoken input. In addition, it uses the computer system's access to current date and time information, together with spoken time information, to produce a clear record of the person's intended schedule. It is said that the system would reduce errors in scheduling arrangements.—DLR

6,108,628

43.72.Ne SPEECH RECOGNITION METHOD AND APPARATUS USING COARSE AND FINE OUTPUT PROBABILITIES UTILIZING AN UNSPECIFIED SPEAKER MODEL

Yasuhiro Komori *et al.*, assignors to Canon Kabushiki Kaisha
22 August 2000 (Class 704/256); filed in Japan 20 September 1996

This speech recognition system is said to be both fast and accurate as a result of a multilayer strategy of recognition with various language models. A preliminary recognition pass results in a broad assessment of speaker characteristics, particularly gender and a broad speaker-type classification.



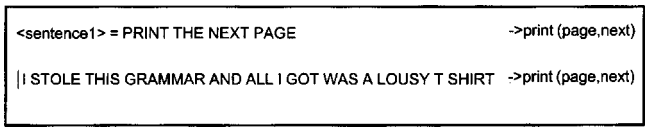
The input speech is then retested using an improved choice of language model. This process is repeated until a high degree of confidence is reached in the final results.—DLR

6,115,820

43.72.Ne DETERMINING THEFT OF GRAMMAR CODE

James R. Lewis, assignor to International Business Machines Corporation
5 September 2000 (Class 713/200); filed 1 June 1998

The trick described in this patent has the intended purpose of revealing whether the speech recognition grammar which contains this trick is being used in an unauthorized application. Embedded in the grammar is a phrase which is very unlikely to be spoken during normal usage. If the recognition



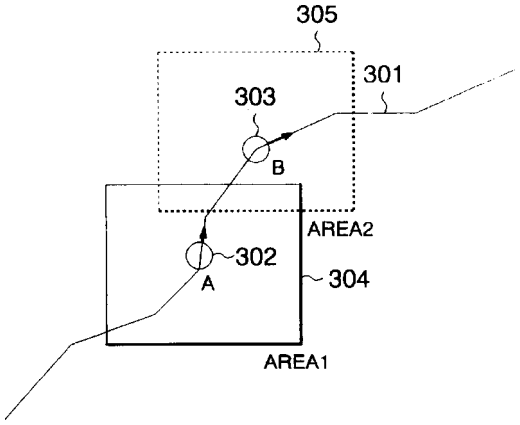
system responds to an utterance of the ''trick'' phrase, cause for a more thorough investigation of the origins of the grammar material is indicated.—DLR

6,112,174

43.72.Ne RECOGNITION DICTIONARY SYSTEM STRUCTURE AND CHANGEOVER METHOD OF SPEECH RECOGNITION SYSTEM FOR CAR NAVIGATION

Shinji Wakisaka *et al.*, assignors to Hitachi Limited; Hitachi Microcomputer System Limited
29 August 2000 (Class 704/251); filed in Japan 13 November 1996

The patent describes the organization of a speech recognition system, such as in an automobile, in which it is desirable to rapidly change the recognition vocabulary. A system of multiple dictionaries allows any one to



be activated depending on external or internal conditions. In the figure, an automobile has moved into a new section of the city. This triggers a switch to a new dictionary with place names for the new district.—DLR

6,111,937

43.72.Ne APPARATUS AND METHOD FOR AUTOMATICALLY DETERMINING WHETHER A TELEPHONE SERVICE USER SPEAKS THE LANGUAGE OF THE SERVICE

Shingo Kuroiwa *et al.*, assignors to Kokusai Denshin Denwa Kabushiki Kaisha
29 August 2000 (Class 379/88.06); filed in Japan 23 August 1996

International telephone services of various types are available to callers who speak a particular language. However, calls to these services from individuals who do not speak the language can be disruptive and, in medical cases, a threat to the proper performance of the services. This patent describes an automated screening process by which the caller is prompted with a well-chosen keyword. Depending on the response, the caller is allowed to continue or denied access to the service.—DLR

6,107,935

43.72.Pf SYSTEMS AND METHODS FOR ACCESS FILTERING EMPLOYING RELAXED RECOGNITION CONSTRAINTS

Liam David Comerford and Stephane Herman Maes, assignors to International Business Machines Corporation
22 August 2000 (Class 340/825.31); filed 11 February 1998

This speaker recognition and verification system includes a speaker-dependent codebook for classification of the provided voice inputs. For each enrolled speaker, the system builds lists of distances to the enrolled speaker's voice patterns from those of each other speaker in the system. When a candidate uses the system and passes the first test of similarity to the voice patterns of the alleged speaker, the system then computes distances from the new speaker to all other enrolled speakers. If all of these distances closely agree with the earlier computed distances, the candidate is accepted as being the alleged person.—DLR

6,112,175

43.72.Pf SPEAKER ADAPTATION USING DISCRIMINATIVE LINEAR REGRESSION ON TIME-VARYING MEAN PARAMETERS IN TRENDED HMM

Rathinavelu Chengalvarayan, assignor to Lucent Technologies, Incorporated
29 August 2000 (Class 704/256); filed 2 March 1998

Recently, several methods of adapting hidden Markov model (HMM) parameters have been used to track speaker characteristics in a speech recognition system. This patent describes such a process, known as trended HMM, to adapt to a speaker. In addition, a process of minimum classification error training is used to optimally estimate the linear transformations to be applied to the trended HMMs.—DLR

Generalized optical theorem for scatterers having inversion symmetry: Applications to acoustic backscattering

Philip L. Marston[a)]
Department of Physics, Washington State University, Pullman, Washington 99164-2814

(Received 22 November 2000; accepted for publication 4 January 2001)

The far-field acoustic scattering amplitudes for the scattering of plane waves by targets having inversion symmetry obey a generalized optical theorem in the absence of dissipation. The theorem allows a component of the complex scattering amplitude in an arbitrary direction to be expressed in terms of an angular integration involving scattering amplitudes evaluated at different angles. The result reduces to the usual optical theorem in the case of forward scattering. The theorem is applied to the backscattering by a perfectly soft sphere as a numerical example. The relevant integrand is shown to be oscillatory. Some potential applications to inverse problems, multiple scattering, and the verification of numerical algorithms are noted. © *2001 Acoustical Society of America.*
[DOI: 10.1121/1.1352082]

PACS numbers: 43.20.Fn, 43.40.Fz, 43.30.Ft [JJM]

I. INTRODUCTION

The optical theorem as it is ordinarily applied to the scattering of sound allows the extinction cross section of an object to be expressed in terms of a component of the complex forward scattering amplitude.[1–3] For the special case where the scatterer is modeled as having no dissipation, the theorem simplifies the calculation of the scattering cross section integrated over all solid angles[4] or the transmission cross section of an aperture.[5,6] The present paper is also concerned with situations where scatterers are modeled as having no dissipation but unlike the ordinary optical theorem, a component of the complex scattering amplitude in an arbitrary direction is related to an angular integration involving scattering amplitudes in other directions (related by supplemental conditions to the direction of interest) provided the scatterer has *inversion symmetry*. Potential applications of this result include certain inverse problems (such as the approximation of backscattering amplitudes from measurements taken at other angles at sufficiently low frequencies) or it may provide a computational test of numerical algorithms for evaluating scattering amplitudes.[7,8]

Symmetry, reciprocity, and energy conservation are widely used in general formulations for acoustic scattering;[9,10] however, the result described here does not appear to be widely utilized in acoustics. For quantum-mechanical scattering amplitudes, Heisenberg[11] is reputed to have derived a theorem similar to the one considered here which is now commonly known as the *generalized optical theorem*.[12] (Newton[12] also reviews the history of the ordinary optical theorem.) In quantum-mechanical scattering, an important inverse problem is the determination of partial-wave phase shifts from scattering data.[13,14] Because quantum-mechanical measurements usually provide only differential cross sections (without phase information), there can be nontrivial ambiguities in the determining the partial-wave phase shifts.[15] The generalized optical theorem has assisted in the understanding and resolution of such ambiguities.[12–14] The importance of inversion symmetry in the generalized quantum-mechanical theorem for scattering by noncentral potentials was noted by Glauber and Schomaker.[16] A generalized optical theorem for the special case of the scattering of sound of perfectly soft targets having inversion symmetry has been noted by Dassios.[17] The theorem has been applied to the derivation of low-frequency expansions of the scattering by soft and hard inversion-symmetric targets.[17,18]

The notation for the complex acoustic scattering amplitude $A(\mathbf{n},\mathbf{n}_i)$ as given by Levine[6] and by Gerjuoy and Saxon[9] is used here. Suppose a scatterer (referred to as the ''target'') is illuminated by a plane wave such that the complex pressure amplitude at the origin O in Fig. 1 is given by $p_i \exp(-i\omega t)$ for the incident wave in the absence of the target. At large distances r from the origin, the pressure amplitude of the scattered wave in the direction denoted by the unit vector $\mathbf{n}$ is $p_s = p_i[A(\mathbf{n},\mathbf{n}_i)/r]\exp(ikr)$, where $\mathbf{n}_i$ is the direction of the incident wave and $k=\omega/c$. From reciprocity[9] (which is equivalent to time-reversal symmetry in the present context), it follows that the amplitude is unchanged for a wave incident along direction $-\mathbf{n}$ and scattered into direction $-\mathbf{n}_i$

$$A(\mathbf{n},\mathbf{n}_i)=A(-\mathbf{n}_i,-\mathbf{n}). \tag{1}$$

For objects which have inversion symmetry about the origin as illustrated in Fig. 1,

$$A(\mathbf{n},\mathbf{n}_i)=A(-\mathbf{n},-\mathbf{n}_i), \tag{2}$$

because nothing is changed when all vectors are inverted through the origin. The relevant directions for arbitrary $\mathbf{n}$ and $\mathbf{n}_i$ are illustrated in Fig. 1. (When the wave is incident along direction $-\mathbf{n}_i$, it is incident from the side opposite the indicated dashed portion.)

Following the notation of Ref. 9, define the complex scattering function (or ''matrix'') $S(\mathbf{n},\mathbf{n}_i)$ such that

$$kA(\mathbf{n},\mathbf{n}_i)/2\pi i=\delta(\mathbf{n}_i-\mathbf{n})-S(\mathbf{n},\mathbf{n}_i), \tag{3}$$

where $\delta(\mathbf{n}-\mathbf{n}_i)$ denotes a two-dimensional delta function defined such that when it is integrated over a solid angle

a)Electronic mail: marston@wsu.edu

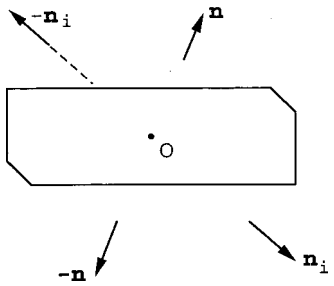


FIG. 1. Scattering by a target having inversion symmetry about the origin *O*. For an incident plane wave directed along an arbitrary direction denoted by $\mathbf{n}_i$, the illumination progresses along the dashed line toward the origin. An arbitrary direction for the scattering is denoted by $\mathbf{n}$. Symmetric vectors are also indicated.

completely containing $\mathbf{n}_i$, the integral gives unity. From the conservation of energy it follows that[9]

$$\int_{4\pi} S(\mathbf{n},\mathbf{n}')S^*(\mathbf{n},\mathbf{n}'')d\Omega = \delta(\mathbf{n}'-\mathbf{n}''). \tag{4}$$

For the purposes of the present paper, the scattering of sound by spheres is used as an example so that the application of Eqs. (3) and (4) to spheres is reviewed in Sec. II. The acoustical generalized optical theorem is given in Sec. III, with the application to spheres given in Sec. IV. Some potential applications, including multiple scattering by arrays, are noted in Sec. V.

Though the numerical example given in Sec. IV is for backscattering by a sphere, the principal result in Sec. III, Eq. (13), is applicable to dissipationless inversion-symmetric scatterers such that Eq. (2) is applicable with the proper choice for the origin. While this may include rigid or soft impenetrable symmetric targets, the target may also have significant scattering contributions associated with (elastic) mechanical responses often attributed to resonances or specific ray contributions.[19–21] Some examples of targets made of isotropic elastic solids of suitable shape having nontrivial elastic responses include cubes[8,22,23] and truncated cylinders.[24] In addition to cubes, most other regular polyhedra (the octahedron, the icosahedron, and the dodecahedron) have the required symmetry along with many other polyhedra.[25] The target may contain defects provided they are inversion symmetric, as is the case with the bevel at each end of the target shown in Fig. 1 which, in the absence of the defect, would be a bluntly truncated circular cylinder. Backscattering enhancements due to elastic responses are also present for spheroids[19] and for tilted disks[26] and bluntly truncated cylindrical shells.[27,28]

II. SCATTERING BY SPHERES

It is instructive to consider the special case of scattering by spheres having neither energy dissipation nor anisotropy in their material properties. The scattering amplitude $A(\mathbf{n},\mathbf{n}_i)$ may be expressed as[19,20,29] $A(\mathbf{n},\mathbf{n}_i)=(a/2)f(\mathbf{n}\cdot\mathbf{n}_i)$, where a is the radius of the sphere and the complex dimensionless form function is given by the partial-wave series

$$f(\mathbf{n}\cdot\mathbf{n}_i)=(-i/ka)\sum_{n=0}^{\infty}(2n+1)P_n(\mathbf{n}\cdot\mathbf{n}_i)(s_n-1), \tag{5}$$

where P_n denotes a Legendre polynomial. The complex s_n depend on ka and are determined by the boundary conditions and material properties. For each term of the series $|s_n|=1$. To verify the decomposition in Eq. (3) and to obtain $S(\mathbf{n},\mathbf{n}_i)$, express each P_n in Eq. (5) with the addition theorem for spherical harmonics[30]

$$P_n(\mathbf{n}\cdot\mathbf{n}_i)=[4\pi/(2n+1)]\sum_{m=-n}^{n}Y_{nm}(\theta,\varphi)Y_{nm}^*(\theta_i,\varphi_i), \tag{6}$$

where θ and φ (θ_i and φ_i) denote the polar and azimuthal angles of $\mathbf{n}(\mathbf{n}_i)$, respectively, with respect to a reference axis. The term of the series in Eq. (5) proportional to $-(2n+1)P_n(\mathbf{n}\cdot\mathbf{n}_i)$ may be summed using the completeness relation[30]

$$\delta(\mathbf{n}_i-\mathbf{n})=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}Y_{nm}(\theta,\varphi)Y_{nm}^*(\theta_i,\varphi_i), \tag{7}$$

which confirms the first term on the right in Eq. (3). Comparison with Eq. (5) gives

$$S(\mathbf{n},\mathbf{n}_i)=(1/4\pi)\sum_{n=0}^{\infty}(2n+1)P_n(\mathbf{n}\cdot\mathbf{n}_i)s_n. \tag{8}$$

This result for S is verified to satisfy Eq. (4) by using Eqs. (6) and (7) with the appropriate substitutions for $\mathbf{n}_i$ and $\mathbf{n}$ and by using the condition $|s_n|=1$.

III. GENERALIZED OPTICAL THEOREM FOR ACOUSTICS

Elimination of S and S^* from the following integral by using Eq. (3) gives

$$\int_{4\pi} S(\mathbf{n},\mathbf{n}')S^*(\mathbf{n},\mathbf{n}'')d\Omega = I_1+I_2+I_3, \tag{9}$$

$$I_1=(k/2\pi)^2\int_{4\pi}A(\mathbf{n},\mathbf{n}')A^*(\mathbf{n},\mathbf{n}'')d\Omega, \tag{10}$$

$$I_2=-k(2\pi i)^{-1}[A(\mathbf{n}'',\mathbf{n}')-A^*(\mathbf{n}',\mathbf{n}'')], \tag{11}$$

$$I_3=\int_{4\pi}\delta(\mathbf{n}-\mathbf{n}')\,\delta(\mathbf{n}-\mathbf{n}'')d\Omega=\delta(\mathbf{n}'-\mathbf{n}''). \tag{12}$$

Comparison with Eq. (4) gives $I_1=-I_2$, which is the special lossless case of a general scattering theorem noted by Dassios and Kleinman.[18] Application of Eqs. (1) and (2) gives $i[A(\mathbf{n}'',\mathbf{n}')-A^*(\mathbf{n}',\mathbf{n}'')]=i[A(\mathbf{n}'',\mathbf{n}')-A^*(\mathbf{n}'',\mathbf{n}')]=-2\,\mathrm{Im}[A(\mathbf{n}'',\mathbf{n}')]$ where Im denotes the imaginary part. From the condition $I_1=-I_2$ and a substitution of $\mathbf{n}_i$ for $\mathbf{n}'$ and $\mathbf{n}'$ for $\mathbf{n}''$,

$$4\pi\,\mathrm{Im}[A(\mathbf{n}',\mathbf{n}_i)]=k\int_{4\pi}A(\mathbf{n},\mathbf{n}_i)A^*(\mathbf{n},\mathbf{n}')d\Omega, \tag{13}$$

which is the generalized optical theorem. The usual optical theorem in the case of a dissipationless target follows by taking $\mathbf{n}'=\mathbf{n}_i$ giving $4\pi\,\mathrm{Im}\,A(\mathbf{n}_i,\mathbf{n}_i)=k\sigma_T$, where σ_T is the

total scattering cross section given by the integral in Eq. (13) with $\mathbf{n}'=\mathbf{n}_i$. The generalized theorem, Eq. (13), gives an integral condition on $A(\mathbf{n}',\mathbf{n}_i)$ for arbitrary scattering direction $\mathbf{n}'$ and incident direction $\mathbf{n}_i$. The integral involves scattering amplitudes in all directions where the incident directions become both $\mathbf{n}_i$ and $\mathbf{n}'$. A corollary to Eq. (13) is that for the broad class of symmetric targets considered, the integral must be a real-valued function of $\mathbf{n}'$ and $\mathbf{n}_i$. Consequently, within the integral $A(\mathbf{n},\mathbf{n}_i)A^*(\mathbf{n},\mathbf{n}')$ may be replaced by $\mathrm{Re}[A(\mathbf{n},\mathbf{n}_i)A^*(\mathbf{n},\mathbf{n}')]$. For the particular case of *backscattering*, $\mathbf{n}'=-\mathbf{n}_i$.

IV. GENERALIZED THEOREM APPLIED TO SPHERES

For the special case of scattering by spheres, $A(\mathbf{n},\mathbf{n}_i)=(a/2)f(\mathbf{n}\cdot\mathbf{n}_i)$ and the generalized theorem gives

$$\mathrm{Im}[f(\mathbf{n}'\cdot\mathbf{n}_i)]=\left(\frac{ka}{8\pi}\right)\int_0^{\pi}d\theta\, f(\cos\theta)\sin\theta\int_0^{2\pi}d\varphi\, f^*(\cos\gamma), \tag{14}$$

where θ and φ denote, respectively, the polar and azimuthal angles of $\mathbf{n}$ and $\cos\gamma=\mathbf{n}\cdot\mathbf{n}'$. The polar axis is taken along $\mathbf{n}_i$ so that $\theta=0$ denotes forward scattering. The explicit dependence of γ on θ and φ follows from the addition theorem, Eq. (6), which gives[30] (with $n=1$)

$$\cos\gamma=\cos\theta\cos\theta'+\sin\theta\sin\theta'\cos(\varphi-\varphi'). \tag{15}$$

With the aforementioned polar axis, $\mathbf{n}'\cdot\mathbf{n}_i=\cos\theta'$ so that any dependence on φ' is lost because of the integration. Equation (14) may also be derived directly from Eqs. (5)–(8) and the unitarity condition $|s_n|=1$.

In the case of backscattering by spheres, $\theta'=\pi$ so that γ no longer depends on $\varphi-\varphi'$. The form function for backscattering $f(-1)$ will be denoted by f_B. The generalized optical theorem gives

$$\mathrm{Im}[f_B]=(ka/4)\int_0^{\pi}\mathrm{Re}[f(\cos\theta)f^*(-\cos\theta)]\sin\theta\,d\theta. \tag{16}$$

Let $\mu=\cos\theta$ and $f(\mu)=\alpha(\mu)+i\beta(\mu)$ where α and β are real functions. It follows that

$$\mathrm{Re}[f(\mu)f^*(-\mu)]=\alpha(\mu)\alpha(-\mu)+\beta(\mu)\beta(-\mu), \tag{17}$$

is an even function of μ so that

$$\mathrm{Im}[f_B]=(ka/2)\int_0^{\pi/2}\mathrm{Re}[f(\cos\theta)f^*(-\cos\theta)]\sin\theta\,d\theta. \tag{18}$$

The results in Eqs. (16) and (18) were confirmed numerically for the case of a perfectly soft sphere. In that case f is given by Eq. (5) with[29,31] $s_n=-h_n^{(2)}(ka)/h_n^{(1)}(ka)$, where $h_n^{(1,2)}$ denote spherical Hankel functions. The partial-wave series was terminated at a value of n sufficiently larger than ka to ensure convergence. For the high-frequency example shown in Fig. 2, Eq. (18) was used to reduce the time for computation. For large ka, $\mathrm{Im}[f_B]$ is approximately periodic in ka with a period of π. This is because the scattering is largely the specular contribution, which is[31] $f_B\approx-\exp(-i2ka)$. The cause of the oscillations of $\mathrm{Im}[f_B]$ follows from elementary geometrical considerations.[21]

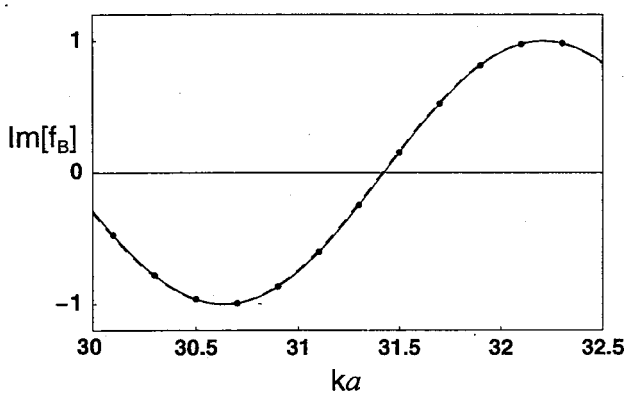


FIG. 2. The solid curve gives $\mathrm{Im}[f_B]$ for backscattering by a perfectly soft sphere as computed directly from the partial-wave series. The points give $\mathrm{Im}[f_B]$ from the generalized optical theorem where the integral in Eq. (18) is numerically evaluated. The dashed curve is given from the geometric approximation to the specular reflection: $\mathrm{Im}[-\exp(-i2ka)]$.

To understand why the integral in Eq. (18) is approximately periodic in ka when ka is large, it is instructive to plot the integrand $F(\theta,ka)=\mathrm{Re}[f(\cos\theta)f^*(-\cos\theta)]\sin\theta$ as shown in Fig. 3. The ka values were chosen as follows: (a) $ka=31.423\,858...$, which is close to a null in $\mathrm{Im}[f_B]$ such that the partial-wave series gives 1×10^{-6}; (b) $ka=30.64$ where $\mathrm{Im}[f_B]=-1.0007$; and (c) $ka=32.21$ where $\mathrm{Im}[f_B]=1.0006$. For case (a), the numerical evaluation of the integral nearly vanishes and the right side of Eq. (18) is found to be less than 10^{-5}. For cases (b) and (c), the behavior of $F(\theta,ka)$ changes at small angles and the sign of $\mathrm{Im}[f_B]$ corresponds to the sign of the first oscillation of $F(\theta,ka)$ on the left. The sign of this oscillation is governed largely by the ka dependence of the phase of the backscattering. That is because when θ is small, $f(\cos\theta)$ is dominated by the forward diffraction contribution[31] and the phase of that contribution depends only weakly on ka. The significance of the first oscillation of $F(\theta,ka)$ was also confirmed by truncating the integration in Eq. (18) below $\pi/2$.

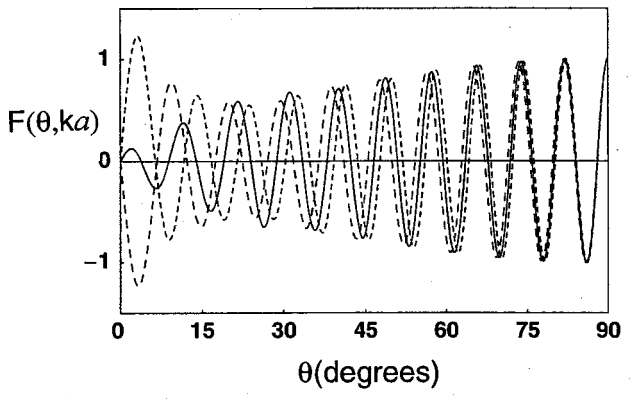


FIG. 3. The integrand $F(\theta,ka)$ for the integral in Eq. (18) evaluated for a perfectly soft sphere. (a) Solid curve: $ka=31.423\,858...$; (b) long dashes: $ka=30.64$; and (c) short dashes: $ka=32.21$. These correspond to the null and the negative and positive extreme of $\mathrm{Im}[f_B]$ in Fig. 2.

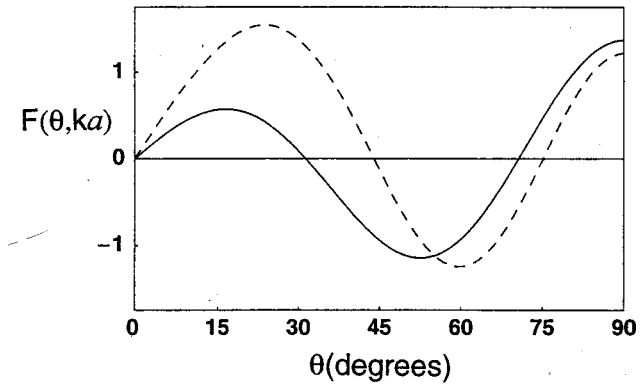


FIG. 4. As in Fig. 3, except $ka=3.207\,75$ (solid curve) and $ka=3.98$ (dashed). These correspond to a null and a maximum of $\mathrm{Im}[f_B]$.

When ka is not large, $F(\theta,ka)$ oscillates more slowly, as shown in Fig. 4. The ka values were selected to lie close to a null and a maximum of $\mathrm{Im}[f_B]$ corresponding to (a) $ka=3.20775$ and (b) $ka=3.98$. The corresponding values of $\mathrm{Im}[f_B]$ computed directly from the partial-wave series as well as from the integral in Eq. (18) were 8.2109×10^{-6} and 1.030 11, respectively.

V. DISCUSSION

As noted in the Introduction, one application of the generalized optical theorem is the determination of a component of the complex scattering amplitude from measurements of the complex amplitude at *other* scattering angles. Inspection of Fig. 3 suggests, however, that this would be impractical at high frequencies because of the angular oscillations of the integrand even for the case as simple as backscattering by a soft target. When ka is not large, however, Fig. 4 indicates that the complex $f(\theta)$ may be sampled at relatively coarse intervals in order to infer $\mathrm{Im}[f_B]$. This inversion procedure may be repeated at different ka to be able to infer $|f_B|$ and the differential cross section for backscattering in cases where the ka dependence of $|f_B|$ is weak.

For the case of scattering by a bounded array of scatterers, the amplitude $A(\mathbf{n},\mathbf{n}_i)$ for the array must include the effects of *multiple scattering* (see, e.g., Ref. 32). The generalized optical theorem, Eq. (13), is applicable provided the collection of scatterers has inversion symmetry and energy is not dissipated. It is widely known (see, e.g., Ref. 33) that the ordinary optical theorem is applicable to an array of scatterers. Figure 5 illustrates an example where the individual members of the two-element array do not have inversion symmetry though the collection does and the theorem applies. The number of array elements offset from the origin must be even. The array may be three-dimensional as if, for example, the members are placed at the vertices of imaginary inversion-symmetric polyhedra.

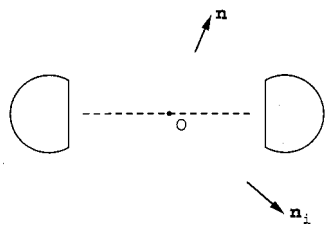


FIG. 5. An example of an array of two targets having inversion symmetry about the origin. In this example, the array also has rotational symmetry about the dashed axis. The targets are truncated spheres.

For the special case of backscattering by a nonspherical target, Eq. (13) may be reduced to Eq. (18) provided two supplemental conditions are met: (a) the target (or array of targets as in Fig. 5) has both inversion symmetry and *axial symmetry*; (b) the backscattering is to be evaluated on the symmetry axis. This may be demonstrated by noting that if $\mathbf{n}$ or $\mathbf{n}_i$ lies along the symmetry axis, a dimensionless form function $f(\mathbf{n}\cdot\mathbf{n}_i)$ exists such that $A(\mathbf{n},\mathbf{n}_i)=(a/2)f$, where a has dimensions of length and f is given by Eq. (5) with $|s_n|=1$. The amplitude $A(\mathbf{n},\mathbf{n}_i)$ has lost its azimuthal dependence.

ACKNOWLEDGMENT

This research was supported by the Office of Naval Research.

[1] A. T. De Hoop, ''On the plane-wave extinction cross-section of an obstacle,'' Appl. Sci. Res., Sect. B **7**, 463–469 (1958).
[2] H. C. van de Hulst, ''On the attenuation of plane waves by obstacles of arbitrary size and form,'' Physica (Amsterdam) **15**, 740–746 (1949); reprinted in *Geometrical Aspects of Scattering*, edited by P. L. Marston (SPIE, Bellingham, WA, 1994), pp. 144–147.
[3] V. Varatharajulu, ''Reciprocity relations and forward amplitude theorems for elastic waves,'' J. Math. Phys. **18**, 537–543 (1977).
[4] S. G. Kargl and P. L. Marston, ''Ray synthesis of Lamb wave contributions to the total scattering cross section for an elastic spherical shell,'' J. Acoust. Soc. Am. **88**, 1103–1113 (1990).
[5] H. Levine and J. Schwinger, ''On the theory of diffraction by an aperture in an infinite plane screen. I,'' Phys. Rev. **74**, 958–974 (1948).
[6] H. Levine, ''Variational principles in acoustic diffraction theory,'' J. Acoust. Soc. Am. **22**, 48–55 (1950).
[7] D. Burnett and R. L. Holford, ''Prolate and oblate spheroidal acoustic infinite elements,'' Comput. Methods Appl. Mech. Eng. **158**, 117–141 (1998).
[8] P. A. Chinnery, J. Zhang, and V. F. Humphrey, ''Acoustic scattering by nonmetallic and metallic cubes in the elastic resonance regime: Experimental measurements and combined finite element/boundary element modeling,'' J. Acoust. Soc. Am. **102**, 60–66 (1997).
[9] E. Gerjuoy and D. S. Saxon, ''Variational principles for the acoustic field,'' Phys. Rev. **94**, 1445 (1954).
[10] P. C. Waterman, ''New formulation of acoustic scattering,'' J. Acoust. Soc. Am. **45**, 1417–1429 (1969); V. V. Varadan, A. Lakhtakia, and V. K. Varadan, Comments on ''Recent criticism of the *T*-matrix method'' [J. Acoust. Soc. Am. **84**, 2280–2284 (1988)].
[11] W. Heisenberg, Z. Phys. **120**, 513–539 (1943); **120**, 673–703 (1943).
[12] R. G. Newton, ''Optical theorem and beyond,'' Am. J. Phys. **44**, 639–642 (1976).
[13] R. G. Newton, ''Determination of the amplitude from the differential cross section by unitarity,'' J. Math. Phys. **9**, 2050–2055 (1968).
[14] K. Chadan and P. C. Sabatier, ''Scattering amplitudes from elastic cross sections,'' in *Inverse Problems in Quantum Scattering Theory*, 2nd ed. (Springer, New York, 1989), Chap. 10, pp. 155–181.
[15] J. H. Crichton, ''Phase-shift ambiguities for spin-independent scattering,'' Nuovo Cimento A **45**, 256–258 (1966).
[16] R. Glauber and V. Schomaker, ''The theory of electron diffraction,'' Phys. Rev. **89**, 667–671 (1953).
[17] G. Dassios, ''Second order low-frequency scattering by the soft ellipsoid,'' SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **38**, 373–381 (1980).
[18] G. Dassios and R. Kleinman, *Low Frequency Scattering* (Oxford University Press, Oxford, 2000), pp. 50, 84–90.
[19] R. H. Hackman, ''Acoustic scattering from elastic solids,'' Phys. Acoust. **22**, 1–194 (1993).

[20] H. Überall, ''Interference and steady-state scattering of sound waves,'' in *Encyclopedia of Acoustics*, edited by M. J. Crocker (Wiley, New York, 1997), Chap. 5, pp. 55–67.
[21] P. L. Marston, ''Quantitative ray methods for scattering'' in *Encyclopedia of Acoustics*, edited by M. J. Crocker (Wiley, New York, 1997), Chap. 43, pp. 483–492.
[22] P. D. Thorne, S. Sun, J. Zhang, I. Bjorno, and T. Mazoyer, ''Measurements and analysis of acoustic backscattering by elastic cubes and irregular polyhedra,'' J. Acoust. Soc. Am. **102**, 2705–2713 (1997).
[23] K. Gipson and P. L. Marston, ''Backscattering enhancements due to retroreflection of ultrasonic leaky Rayleigh waves at corners of solid elastic cubes in water,'' J. Acoust. Soc. Am. **105**, 700–710 (1999).
[24] K. Gipson and P. L. Marston, ''Backscattering enhancements due to reflection of meridional leaky Rayleigh waves at the blunt truncation of a tilted solid cylinder in water: Observations and theory,'' J. Acoust. Soc. Am. **106**, 1673–1680 (1999).
[25] H. S. M. Coxeter, *Regular Polytopes* (Dover, New York, 1963).
[26] B. T. Hefner and P. L. Marston, ''Backscattering enhancements associated with the excitation of symmetric Lamb waves on a circular plate: Direct and holographic observations,'' ARLO **2**, 55–60 (2001).
[27] M. L. Rummerman, ''Contribution of membrane wave reradiation to scattering from finite cylindrical steel shells in water,'' J. Acoust. Soc. Am. **93**, 55–65 (1993).
[28] S. F. Morse, P. L. Marston, and G. Kaduchak, ''High frequency backscattering enhancements by thick finite cylindrical shells in water at oblique incidence: Experiments, interpretation and calculations,'' J. Acoust. Soc. Am. **103**, 785–794 (1998).
[29] L. Flax, L. R. Dragonette, and H. Überall, ''Theory of elastic resonance excitation by sound scattering,'' J. Acoust. Soc. Am. **63**, 723–731 (1978).
[30] J. D. Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, New York, 1999), Sec. 3.5–3.6, pp. 107–111.
[31] H. M. Nussenzveig, ''Uniform approximation in scattering by spheres,'' J. Phys. A **21**, 81–109 (1988).
[32] R. Lim and R. H. Hackman, ''A formulation of multiple scattering by many bounded obstacles using a multicentered, *T* supermatrix,'' J. Acoust. Soc. Am. **91**, 613–638 (1992).
[33] V. Twersky, ''Multiple scattering of waves and optical phenomena,'' J. Opt. Soc. Am. **52**, 145–171 (1962).

On the modeling of narrow gaps using the standard boundary element method

Vicente Cutanda[a)]
Brüel & Kjær Sound & Vibration, Skodsborgvej 307, DK-2850 Nærum, Denmark

Peter Møller Juhl[b)]
ITF Institute of Applied Physics, Odense University/The Engineering College of Odense, Niels Bohrs Allé 1, DK-5230 Odense M, Denmark

Finn Jacobsen[c)]
Department of Acoustic Technology, Technical University of Denmark, Ørsteds Plads, Building 352, DK-2800 Kgs. Lyngby, Denmark

(Received 6 October 1999; revised 9 September 2000; accepted 21 December 2000)

Numerical methods based on the Helmholtz integral equation are well suited for solving acoustic scattering and diffraction problems at relatively low frequencies. However, it is well known that the standard method becomes degenerate if the objects that disturb the sound field are very thin. This paper makes use of a standard axisymmetric Helmholtz integral equation formulation and its boundary element method (BEM) implementation to study the behavior of the method on two test cases: a thin rigid disk of variable thickness and two rigid cylinders separated by a gap of variable width. Both problems give rise to the same kind of degeneracy in the method, and modified formulations have been proposed to overcome this difficulty. However, such techniques are better suited for the so-called thin-body problem than for the reciprocal narrow-gap problem, and only the first is usually dealt with in the literature. A simple integration technique that can extend the range of thicknesses/widths tractable by the otherwise unmodified standard formulation is presented and tested. This technique is valid for both cases. The modeling of acoustic transducers like sound intensity probes and condenser microphones has motivated this work, although the proposed technique has a wider range of applications.
[DOI: 10.1121/1.1350399]

PACS numbers: 43.20.Fn [ANN]

I. INTRODUCTION

Phenomena where the physical setup contains very close domain boundaries are not uncommon in practice. Typical cases are fins or appendages from bulky structures, thin plates, and shells. All of them have parts that are very thin compared to the overall dimensions. It is well known that the conventional boundary element method has difficulties in dealing with such tasks. Considerable effort has been put into this so-called thin-body problem in recent years in order to make it tractable with BEM, and different formulations have been proposed that can alleviate or remove such difficulties.[1–4]

There is a second family of cases that shares many features with thin bodies. This may be named the narrow-gap problem. Examples are coatings, lubricant layers, cracks, and some acoustic transducers. The domain of interest in these cases is situated *between* the close boundaries, and very often a two-dimensional simplification cannot be used if the transversal variations are of physical relevance.

BEM formulations that can deal effectively with thin bodies become of no use in narrow gaps, and only a few techniques remain that can provide results in practical engineering work. One of these techniques is described in this paper, which has been motivated by numerical modeling of sound intensity probes and condenser microphones.[5] The behavior of sound fields inside narrow gaps plays a central role in these devices. Therefore, the performance of BEM formulations in such situations must be understood in order to obtain meaningful results.

The conventional BEM has two difficulties whenever two parts of the surface are very close.[4] First, the coefficient matrix becomes ill-conditioned as the distance gets smaller, and second, the integrals are near singular and difficult to solve numerically. The methods proposed to get around these difficulties in the thin-body variant fall into two groups: multidomain methods and normal-derivative equation methods. In multidomain methods an imaginary surface is constructed so as to replace the original problem of a thin body in a single domain with bulky bodies in two domains or more, coupled at an imaginary surface.[2] Such methods are advantageous for certain problems such as muffler analysis, but for other cases the imaginary surface may be quite large and will therefore give rise to considerable computational work. The alternative methods involve the normal derivative of the Helmholtz integral equation (HIE). A popular implementation is a combination of the HIE with its normal derivative, since this combination also can be shown to overcome the nonuniqueness problem of the standard HIE.[6] A strategy often chosen is to apply the combined equation on the midsurface of the thin body.[1,3] Another procedure, which does not

[a)]Electronic mail: vcutanda@bksv.com
[b)]Electronic mail: pmjuhl@itf.sdu.dk
[c)]Electronic mail: fjac@oersted.dtu.dk

assume the approximation of an infinitely thin body, applies the HIE on one side of the thin body and its normal derivative on the other.[4]

For a plane narrow gap a multidomain strategy could be used in which the gap could be modeled as a two-dimensional problem coupled to an exterior three-dimensional problem. However, this approach is only approximate for any finite gap width, and not suitable for problems where the sound field details inside the gap are important. As to the combined formulation in Ref. 4, the regularizing effect seems to rely on a medium being present outside, as in the case of inclusions.

Since complete removal of the ill-conditioning in the gap case is problematic, it is interesting to examine whether the conventional formulation can still provide correct results for practical cases under such circumstances. As pointed out in Ref. 4, adequate treatment of the near-singular integrals can prevent a breakdown due to poor integration, but it leaves the ill-conditioning. We will show in this paper how a simple numerical integration strategy can extend the range of aspect ratios (smaller dimension/overall dimension) by several orders of magnitude despite ill-conditioning, thus placing most practical gap problems within reach of the standard BEM. Besides, the proposed numerical integration only requires a very small increase of computing resources.

The nature of the problem of close boundaries is reviewed in the next section, with emphasis on the features of the gap case. An explanation of the numerical integration technique developed follows. The remainder of the paper is dedicated to thin-body and narrow-gap test cases. An axisymmetrical formulation is used, which is outlined along with details of the study methodology. The behavior of the method's convergence towards the solution is analyzed on the test cases, as well as the ill-conditioning of the coefficient matrix. The influence of frequency, mesh density, and aspect ratio is investigated.

II. THE PROBLEM OF CLOSE DOMAIN BOUNDARIES

A. The standard Helmholtz integral equation (HIE)

The BEM approach to acoustic radiation and scattering problems is based on the Helmholtz integral equation that relates the pressure $p(Q)$ and normal velocity $\nu(Q)$ on the surface of a body of any shape (see Fig. 1) with the pressure at any point $p(P)$ and the pressure of an incoming wave $p^I(P)$.[7] The harmonic time dependence $e^{i\omega t}$ is omitted, giving

$$C(P)p(P)=\int_S\left(\frac{\partial G}{\partial n}p(Q)+ikz_0\nu(Q)G\right)dS+4\pi p^I(P), \tag{1}$$

where S is the surface of the body, Q a point on that surface, and P any exterior or interior point. The normal vector n is directed into the computational domain. The Green's function for 3D free space is

$$G(R)=\frac{e^{-ikR}}{R},\quad R=|P-Q|. \tag{2}$$

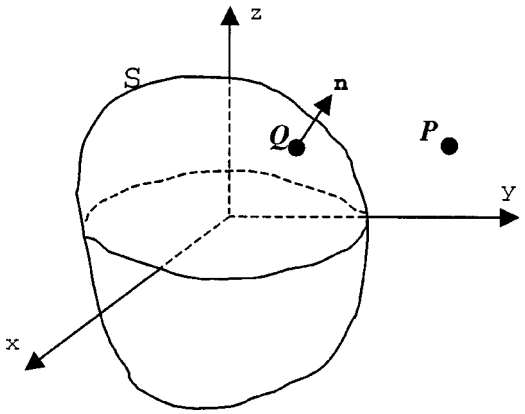


FIG. 1. Generic integration domain and boundary surface.

The factor $C(P)$ is the geometrical constant and represents the exterior solid angle at P. It is calculated by

$$C(P)=4\pi+\int_S\frac{\partial}{\partial n}\left(\frac{1}{R}\right)dS. \tag{3}$$

The expression (3) is valid for exterior problems; for interior problems 4π should be subtracted. In the present study only scattering by rigid bodies is considered; thus, $p^I(P)$ is the excitation and $\nu(Q)=0$, making Eq. (1) simpler. However, the coefficient matrix obtained from the BEM numerical implementation of (1) will be the same as for radiation problems, and therefore the conclusions discussed below will still be general. The standard collocation formulation will be used.

B. The thin-body problem

The case of a thin planar body has been discussed extensively in the literature, and hence the results are merely listed. For a thin disk using constant elements and unmodified numerical integration, the coefficient matrix of the BEM becomes

$$2\pi\begin{pmatrix}\mathbf{I} & \sim\mathbf{0}\\ \sim\mathbf{0} & \mathbf{I}\end{pmatrix}, \tag{4}$$

where $\mathbf{I}$ is the identity matrix and $\sim\mathbf{0}$ is approximately a matrix of zeros. This matrix is perfectly conditioned, but does not lead to the correct solution of the problem.[8] The problem is grounded in the nonhandling of the near singularity of the integral as the integration point Q to the surface S_+ passes the collocation point P on the near surface S (see Fig. 2). With proper treatment of the near singularity, one finds a correct matrix representation for the problem[4,8]

$$2\pi\begin{pmatrix}\mathbf{I} & \sim\mathbf{I}\\ \sim\mathbf{I} & \mathbf{I}\end{pmatrix}. \tag{5}$$

Here, the near degeneracy is reflected in an ill-conditioned matrix, which potentially may lead to incorrect solutions. As mentioned in the Introduction, two solutions have been proposed to deal with thin bodies, the multidomain method and the use of the normal derivative. Both of them reformulate the problem in a way that makes the ill-conditioning disappear, although in the case of some combined formulations

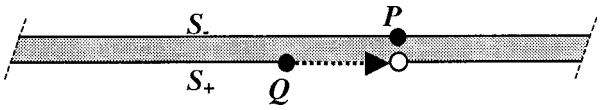


FIG. 2. Thin-body setup.

using the normal derivative, near-singular integrals are still present and must be taken care of.[4,9]

C. The narrow-gap problem

Consider a rigid bulky object with a narrow gap, as represented in Fig. 3. The surface on the exterior of the object is denoted by S, and the surfaces in the gap by S_- and S_+, respectively. Since there is no pressure jump across the gap in the limit of a gap of zero thickness, the normal derivative methods do not pose any new information for this kind of problems. The standard HIE for a scattering problem with such an object is

$$C(P)p(P)=\int_S p(Q)\frac{\partial G}{\partial n}dS+\int_{S_-} p_-(Q)\frac{\partial G}{\partial n}dS + \int_{S_+} p_+(Q)\frac{\partial G}{\partial n}dS+4\pi p^I(P). \tag{6}$$

Let P be on S. In the limit of an infinitely narrow gap, continuity of the pressure requires that $p_-=p_+$ so the two integrals for the two gap surfaces cancel out. Hence, in the limit of narrow gaps the sound field outside the gap is not influenced by the sound field inside the gap. This is in agreement with what one would expect physically. Therefore, the solution strategy for a problem involving a very narrow planar gap with rigid surfaces could be to solve an exterior problem neglecting the gap, and then to solve the gap problem as a two-dimensional interior problem with the pressures obtained for the exterior problem as the boundary conditions. However, if the gap width cannot be neglected, if parts of the gap surface are of finite impedance, and/or if viscosity is to be taken into account, the need for a simultaneous analysis of the gap and the exterior field is envisaged. If the near singularity is properly taken into account, the block of identity matrices of Eq. (5) is found in the matrix equivalent of Eq. (6)

$$\begin{pmatrix} \mathbf{A} & \mathbf{D} & -\mathbf{D} \\ \mathbf{C} & 2\pi\mathbf{I} & \sim 2\pi\mathbf{I} \\ \sim\mathbf{C} & \sim 2\pi\mathbf{I} & 2\pi\mathbf{I} \end{pmatrix}\begin{pmatrix} p \\ p_- \\ p_+ \end{pmatrix}=4\pi p^I. \tag{7}$$

The first block of rows in Eq. (7) refers to collocation points on the exterior surface S, and the second and third block of rows refer to collocation points on S_- and S_+, respectively. Again, the near degeneracy due to the two close surfaces is reflected in ill-conditioning of the coefficient matrix. If the near singularity is not dealt with properly, the lower right corner of the coefficient matrix is to be replaced with Eq. (4)

$$\begin{pmatrix} \mathbf{A} & \mathbf{D} & -\mathbf{D} \\ \mathbf{C} & 2\pi\mathbf{I} & \sim\mathbf{0} \\ \sim\mathbf{C} & \sim\mathbf{0} & 2\pi\mathbf{I} \end{pmatrix}\begin{pmatrix} p \\ p_- \\ p_+ \end{pmatrix}=4\pi p^I, \tag{8}$$

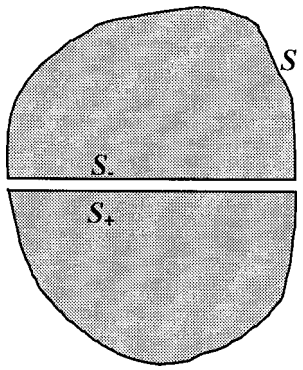


FIG. 3. Narrow-gap setup.

and the resulting (false) system of equations will be well-conditioned. In the latter case, which represents a standard numerical implementation, it is easy to show that the solution will tend towards zero inside the gap as the gap width tends towards zero. The solution outside the gap will still tend towards the solution of the equivalent exterior problem with the gap removed.

III. IMPROVEMENT OF THE NUMERICAL INTEGRATION

To achieve the goal of obtaining valid solutions for narrow-gap problems, the strategy chosen is the improvement of the numerical integration technique. A system of equations of the form in (7) should be guaranteed for a range of aspect ratios of practical importance. As mentioned in the last section, there is still ill conditioning, and therefore the accuracy of the elements in the coefficient matrix, which are obtained by numerical integration, imposes a limit to the aspect ratio that can be calculated. Nevertheless, if this limit is high enough, the cases left out will be too narrow to have practical use or even physical meaning.

Several authors propose ways to handle near-singular integrals. For example, in Refs. 4 and 9 an analytical removal of the near singularity that splits the integral is performed. The resulting terms are treated with variable changes and the Stokes theorem in order to reduce or eliminate their difficulty. However, the approach taken in this paper will be numerical, not analytical.

The behavior of a near-singular integrand differs in many ways from a genuine singularity. If a collocation point is in the neighborhood of an element to be integrated, it produces a perturbation on the integrand around its projection on the element, more localized and more acute the shorter the distance. This effect can be observed in Fig. 4, which is a near-singular integrand along a one-dimensional element. It has been obtained from the disk test case to be presented in the following section. Note that integrand values are represented in a logarithmic scale and that only a part of the element is shown. It becomes clear from Fig. 4 why standard numerical integration methods, like Gauss–Legendre, miss this troublesome area for a given relative distance and lead to results as shown in Eqs. (4) and (8). This situation has often been described as the ''breakdown'' of the BEM standard formulation.

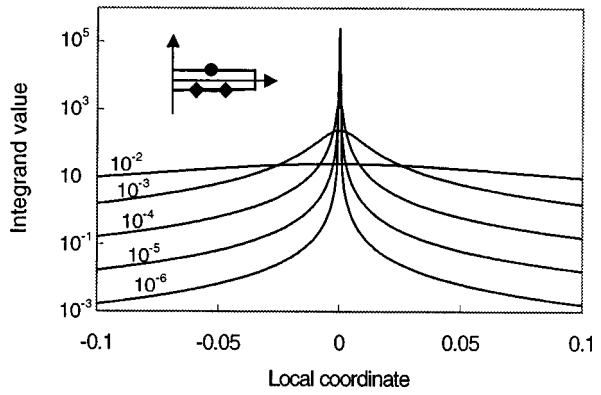


FIG. 4. Near-singular behavior of an integrand in the thin disk calculation. Abscissas are local coordinates along the element $[-1,1]$; ordinates are values of the integrand on a logarithmic scale. The curves are integrands with a collocation point at 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, and 10^{-6} units' distance from the local coordinate 0. The disk radius is 1 unit. Five elements per disk side are used.

Only very limited improvement is obtained by merely increasing the order of the numerical integration formula. The integration does not concentrate the effort around the near singularity, where the integrand has a peak. The use of numerical formulas designed for singular integrals has also been considered, but they do not perform well due to the different nature of the near-singular integrand.

Another possible choice is an adaptive numerical integration routine. Such routines have been used for many years on all kinds of problems and are implemented in most mathematical software packages. Basically, the algorithm tries to decrease the integration error down to a given value. To do this, the error is estimated during every iterative step, and more integration points are used on the difficult areas. When applied to near-singular integrals in BEM, there is a clear improvement. However, these methods often fail in the limit of very close collocation points. The floating-point precision of the machine is reached during the internal calculation of integration error estimates, with unpredictable results. This lack of reliability and control on the routine behavior seems to rule them out.

An optimal numerical integration scheme for this particular problem should be simple enough to minimize errors but, on the other hand, it should concentrate the effort around the difficult area. If the information about the near singularity and its strength can be used in order to perform each integration in the most appropriate way, the computer load can also be reduced.

In view of this, the approach finally adopted employs an exponential interval division according to the expected near-singular behavior, as shown in Fig. 5. In this study, one-dimensional quadratic elements are used, but the technique can be translated to other implementations. The distance from each subinterval boundary to the collocation point projection is expressed by $c \cdot b^n \cdot d$, where b and c are constants, d is the distance from collocation point to the element, relative to the element size, and n is the subinterval number. In practical calculations, it was adjusted to $b=2$ and $c=2^{-5}$. Each subinterval can be numerically integrated using low-order standard techniques. In this way the information available about the integrand is used to perform the numerical integration, and no complex adaptive integration strategy is needed.

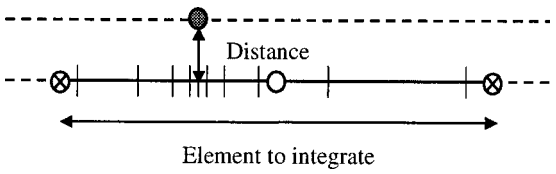


FIG. 5. Exponential interval division on a one-dimensional quadratic element. Relative distance and local coordinate of the collocation point are used to calculate the number and sizes of the subintervals.

TABLE I. Processing time results on a DEC Alpha 433 MHz for the two-cylinder test case (see Sec. IV B) with 60 elements. Gauss–Legendre numerical integration of order 20 and the interval division technique are compared.

Gap width	Integration	Time (s)	Result
10^{-2}	Gauss	82.7	Good
	Division	91.4	Good
10^{-3}	Gauss	83.0	Fail
	Division	93.2	Good
10^{-4}	Gauss	83.2	Fail
	Division	93.9	Good
10^{-5}	Gauss	82.7	Fail
	Division	94.4	Good
10^{-6}	Gauss	82.6	Fail
	Division	96.7	Fail

The improved integration does not involve a significant increase in computer load. Computational effort is only applied where it is necessary; that is, only in those cases where the collocation point is very close to the element and, within that element, around the near singularity and proportional to its strength. Table I presents some run times for the narrow-gap test case and a 60-element fixed mesh. The interval division technique adds around 10%–20% to the normal processing time, with a slight increase for narrower gaps. This is logical if we consider that the technique is only used to calculate an order of N elements in an $N \times N$ matrix.

IV. TEST CASES

A. BEM axisymmetrical formulation

If an axisymmetric body or bodies on the same axis are considered, it is possible to simplify the standard BEM from a surface integral to an integral along the generator and another over the angle of revolution, in a cylindrical coordinate system. The use of a cosine expansion of $p(Q)$ and $\nu(Q)$ in orthogonal terms allows the isolation of the singularities contained in the revolution integrals so that only the generator has to be discretized, saving computing time and storage capacity. Although the excitations described in this paper are also axisymmetrical, the cosine expansions permit nonaxisymmetrical boundary conditions. This implementation retains most features of the full 3D version, which means that the problem of close boundaries can be studied more easily and over a larger number of test cases.[10]

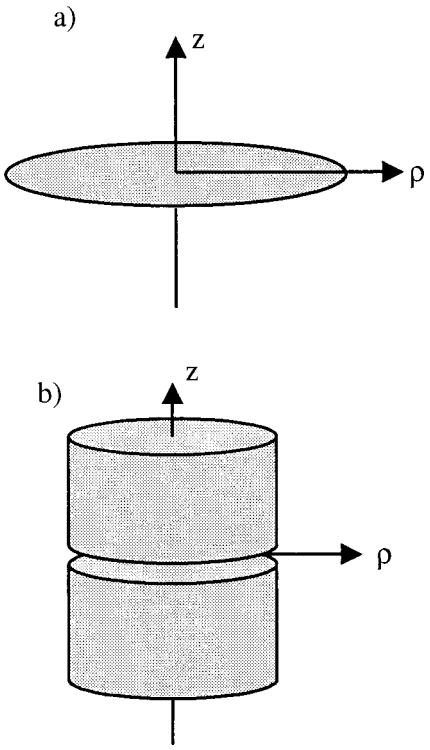


FIG. 6. Test cases. (a) thin disc; (b) narrow gap.

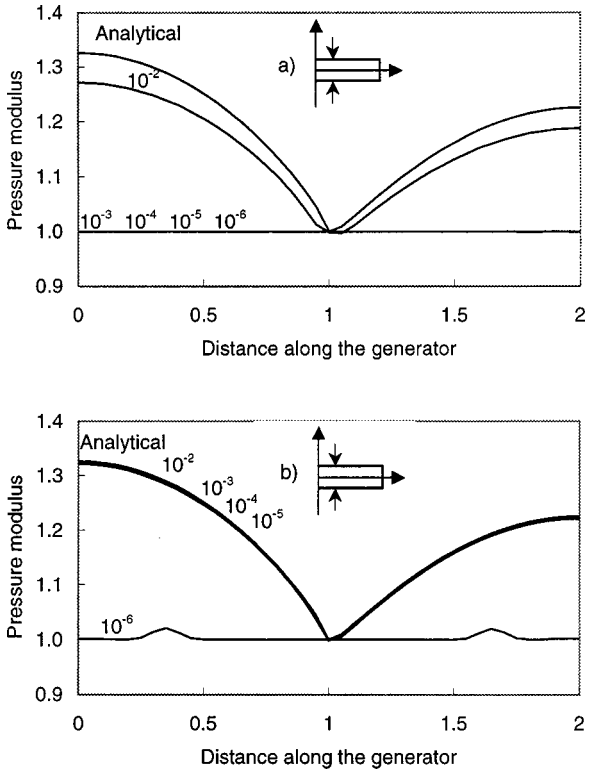


FIG. 7. Calculated sound-pressure modulus on the generator of a disk [see Fig. 6(a)] of variable thickness using a mesh of 10 elements per unit, 20 elements in total. The thicknesses are 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, and 10^{-6}. The analytical solution for a infinitely thin disk is also plotted. An axial plane wave of $ka=1$ and unit amplitude is scattered by the disk. Calculations using (a) Gauss–Legendre numerical integration of order 20; (b) with interval division as explained in Sec. III.

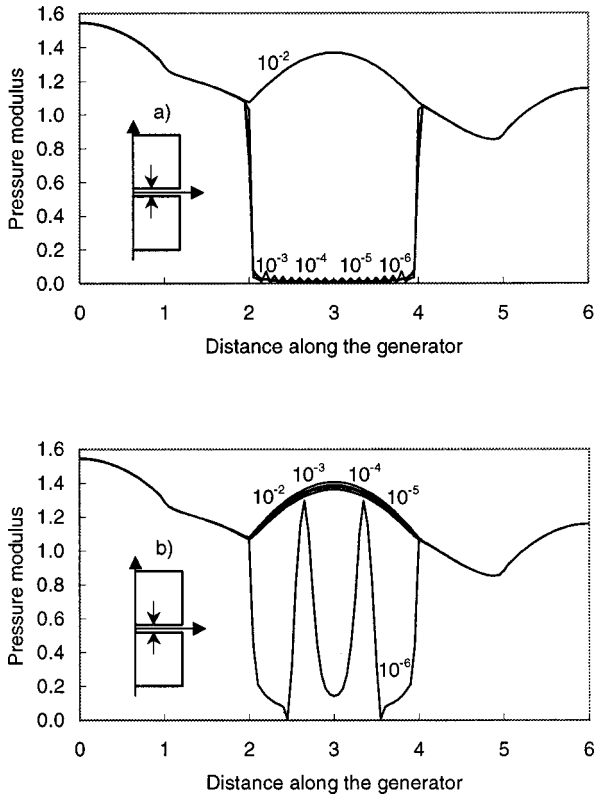


FIG. 8. Calculated sound-pressure modulus on the generator of two coaxial rigid cylinders [see Fig. 6(b)] separated by a variable narrow gap calculated using a mesh of 10 elements per unit, 60 elements in total. The gap widths are 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, and 10^{-6}. An axial plane wave of $ka=1$ and unit amplitude is scattered by the setup. Calculations using (a) Gauss–Legendre numerical integration of order 20; (b) with interval division as explained in Sec. III.

B. Test cases

Two test cases have been chosen to represent the thin-body problem and the narrow-gap problem. These are, respectively, a disk of variable thickness and two cylinders with a common axis that form a narrow gap. They are represented in Fig. 6.

These cases have rotational symmetry along the cylindrical coordinate θ and therefore the axisymmetrical BEM formulation can be applied. The bodies are supposed to be rigid (infinite impedance) and excited by a plane wave coming from the z^+-axis direction. Several values of ka (wave number times radius) have been used in the calculations, but only the results for $ka=1$ are presented, since no relevant variation with the frequency has been observed. The radius a of disk and cylinders is normalized to 1, as well as cylinder lengths.

The numerical implementation is made by dividing the generators into line elements. Both pressure and geometry are modeled using quadratic shape functions.

C. Sound-pressure results

The sound-pressure modulus along the generator of the objects in Fig. 6 is shown in Figs. 7 (disk) and 8 (gap) for a variety of thicknesses/widths. Standard Gauss–Legendre integration [Figs. 7(a)/8(a)] succeeds only if the disks or gaps are thicker than 10^{-2} units, while the proposed interval di-

vision can give good results down to 10^{-5} units [Figs. 7(b)/8(b)]. The two breakdowns are different. An integration failure causes the pressure to drop to zero on the close surfaces, while the improved integration avoids this down to a point where ill-conditioning is too serious for the limited precision of the computer to cope with, and unpredictable results appear. Other calculations not shown here have demonstrated that neither mesh density nor frequency has any influence on the ill-conditioning breakdown. It is only dependent on the aspect ratio.

Figure 7 also includes the corresponding analytical solution for a disk of zero thickness, which can be obtained as a series of oblate spheroidal functions. This solution has been calculated in order to validate the results and study the convergence. A brief summary is given in the Appendix of this paper.

The pressure increase in the gap observed in Fig. 8 can be explained in the limiting case of a vanishing gap. Ideally, the external sound field would not be influenced by the presence of a very narrow gap. Therefore, considering the gap as two-dimensional, the sound pressure along the radius (generator) inside the gap has the form of a Bessel function.[11] Its boundary condition is the sound pressure on the external surface connected to the gap (abscissas 2 and 4 in Fig. 8). For $ka=1$, there are no zeros of the Bessel function within the gap; therefore, only a pressure increase is observed.

D. Convergence of the improved solution

Using the calculated complex sound pressure on the generator nodes and the corresponding analytical values for an infinitely thin disk, it is possible to study the convergence of the thin-disk calculation as a function of mesh density and disk thickness.

The thicknesses examined are small enough to make them physically very similar to a disk with no thickness. The error is calculated as the length of the residual vector relative to the analytical solution

$$\text{Relative error}=\frac{\sqrt{\Sigma_{j=1}^{M}[(\mathrm{Re}(P_{j\ \text{analytical}})-\mathrm{Re}(P_{j\ \text{calculated}}))^2+(\mathrm{Im}(P_{j\ \text{analytical}})-\mathrm{Im}(P_{j\ \text{calculated}}))^2]}}{\sqrt{\Sigma_{j=1}^{M}[\mathrm{Re}(P_{j\ \text{analytical}})^2+\mathrm{Im}(P_{j\ \text{analytical}})^2]}}, \tag{9}$$

where $P_{j\ \text{analytical}}$ is the analytical solution at node j and $P_{j\ \text{calculated}}$ is the calculated complex pressure at the same node. M is the total number of nodes of every solution. See Fig. 9.

Unfortunately, the narrow-gap case does not have a suitable analytical solution. The two-dimensional analogy mentioned in the last section is not precise enough to study convergence, since it is dependent on the boundary conditions at the rim that are themselves subject to calculation errors, which are aggravated because of the sharp edge singularity.[12] The convergence can nevertheless be studied by using a solution with a very fine mesh as a reference. This is what Fig. 10 represents.

The convergence is clear in the disk case, and independent of the thickness. Only the 10^{-6} case cannot converge due to the ill-conditioning of the coefficient matrix, but thicker disks do not show any strange behavior. The gap case is physically more complicated because what happens inside the gap is strongly influenced by the phenomena on the gap rim; therefore, the convergence pattern differs. Nevertheless, the solutions clearly converge again except for the 10^{-6} case. The use of a calculated solution as a reference produces also a shift in the error scale.

E. Ill-conditioning

The condition number of the coefficient matrix has been calculated for a number of mesh densities and disk/gap thicknesses. It was shown in Sec. III that a good numerical integration in a problem of close surfaces produces an ill-

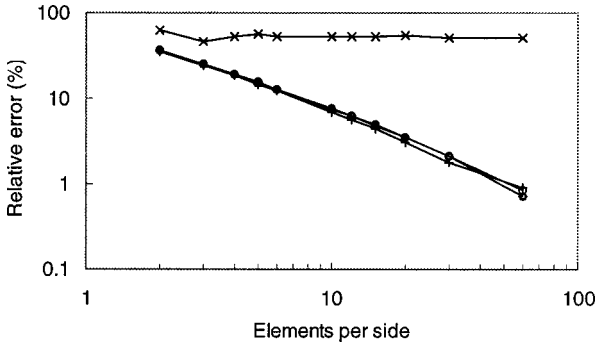


FIG. 9. Convergence towards the analytical solution of a thin disk [see Fig. 6(a)] of variable thickness, using the proposed numerical integration. The thicknesses are 10^{-2} (+), 10^{-3} ($\diamond$), 10^{-4} ($\square$), 10^{-5} ($\bigcirc$), and 10^{-6} ($\times$). The relative error is represented as a function of the mesh density. An axial plane wave of $ka=1$ and unit amplitude is scattered by the disk.

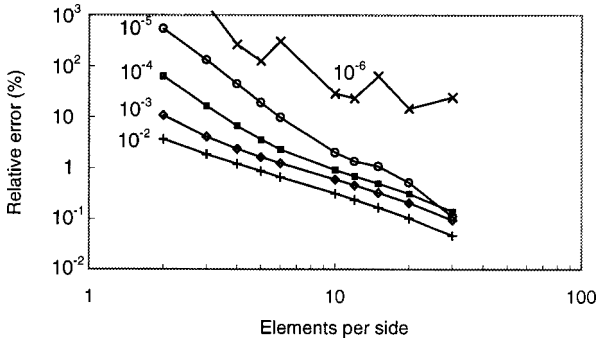


FIG. 10. Convergence towards the ''true'' solution of a narrow gap [see Fig. 6(b)] of variable width, using the proposed numerical integration. The widths are 10^{-2} (+), 10^{-3} ($\diamond$), 10^{-4} ($\square$), 10^{-5} ($\bigcirc$), and 10^{-6} ($\times$). The relative error in the gap nodes is represented as a function of the mesh density. An axial plane wave of $ka=1$ and unit amplitude is scattered by the cylinders.

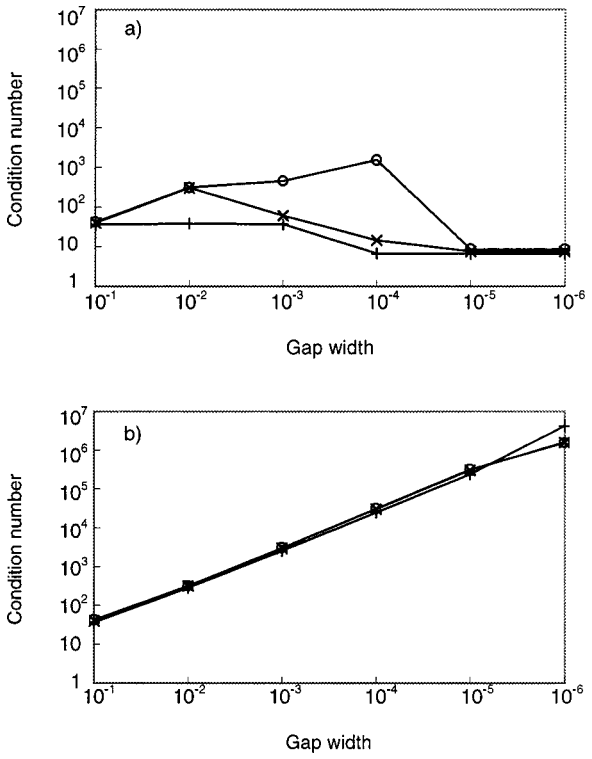


FIG. 11. Condition number of the coefficient matrix for the narrow-gap test case, as a function of gap width. Three mesh densities are plotted: 2 (+), 10 (×), and 60 (○) elements per unit. Calculations using (a) Gauss–Legendre numerical integration of order 20; (b) with interval division as explained in Sec. III. Calculation made with $ka=1$.

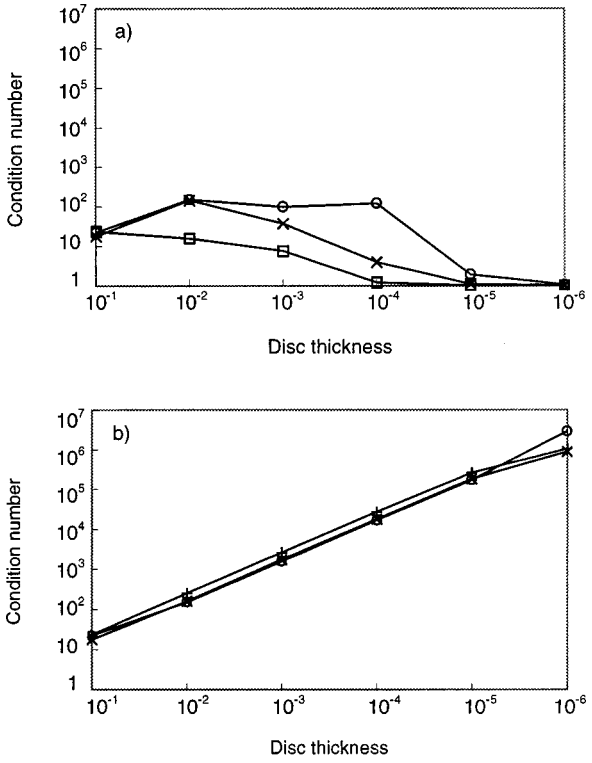


FIG. 12. Condition number of the coefficient matrix for the thin-disk test case, as a function of thickness. Three mesh densities are plotted: 2 (+), 10 (×), and 60 (○) elements per side. Calculations using (a) Gauss–Legendre numerical integration of order 20; (b) with interval division as explained in Sec. III. Calculation made with $ka=1$.

conditioned system of equations, and therefore a high condition number of the coefficient matrix. Conversely, a failure of the integration generates low condition numbers, but with erroneous solutions.

Figures 11 and 12 show that the ill-conditioning behaves similarly in the two test cases. If the near-singular integrals are not dealt with, the condition number does not grow very much, but the calculation gives erroneous results. When the interval division technique is used the condition number grows exponentially as the surfaces get closer, no matter which mesh density is used. The frequency also has very little influence, as other calculations not presented here have shown. Hence, the aspect ratio indeed seems to determine the condition number, independently of whether we deal with narrow gaps or thin bodies.

V. CONCLUSIONS

It has been shown that a case of practical importance in engineering like the narrow gap can be modeled using conventional BEM despite the ill-conditioning of the coefficient matrix and with no need of denser meshes. A convergence study has also revealed that the solution improves normally for gradually denser meshes. A simple strategy of interval division with low computational cost is enough to extend the range of tractable aspect ratios by three orders of magnitude, from 10^{-2} to 10^{-5}. This makes it possible to deal with problems like condenser microphones and sound intensity probes, which have motivated this study. For example, the air layer behind the diaphragm of a 1/2-in. condenser microphone is about 20 μm, giving an aspect ratio of about 10^{-3}.

APPENDIX: ANALYTICAL SOLUTION FOR SCATTERING BY A FLAT DISK

The infinitely thin disk is the limiting case of the oblate spheroid, which is better described in the oblate spheroidal coordinate system (ξ, η, ϕ). Its relation with rectangular Cartesian coordinates is

$$x=\tfrac{1}{2}d\sqrt{(\xi^2+1)(1-\eta^2)}\cos\phi,$$
$$y=\tfrac{1}{2}d\sqrt{(\xi^2+1)(1-\eta^2)}\sin\phi, \quad z=\tfrac{1}{2}d\xi\mu. \tag{A1}$$

The wave equation is separable in this coordinate system into spheroidal wave functions, thus providing analytical solutions for a range of cases. The particular solution used in this paper corresponds to an infinitely thin hard disk ($\xi=0$ and $d=$diameter) excited by a plane wave coming from the positive z axis. The sound pressure (incident and scattered) on its surface can be expressed by the series[13]

$$P^i+P^s=\frac{2}{ka}\sum_{n=0}^{\infty}\frac{i^n}{N_{0n}(-ika)}\frac{S_{0n}(-ika,-1)S_{0n}(-ika,\eta)}{R_{0n}^{(3)\prime}(-ika,i0)}, \tag{A2}$$

where S_{mn} are the oblate spheroidal angular functions, $R_{mn}^{(3)}$ are the derivatives of the oblate spheroidal radial functions of the third kind, and N_{nm} are the normalization factors.[14] These functions are also expressed as infinite series, and computer algorithms have been used to calculate them with sufficient accuracy.[15]

[1] R. Martinez, ''The thin-shape breakdown (TSB) of the Helmholtz integral equation,'' J. Acoust. Soc. Am. **90**, 2728–2738 (1991).
[2] C. Y. R. Cheng, A. F. Seybert, and T. W. Wu, ''A multidomain boundary element solution for silencer and muffler performance prediction,'' J. Sound Vib. **151**, 119–129 (1991).
[3] T. W. Wu and G. C. Wan, ''Numerical modeling of acoustic radiation and scattering from thin bodies using a Cauchy principal integral equation,'' J. Acoust. Soc. Am. **92**, 2900–2906 (1992).
[4] G. Krishnasamy, F. J. Rizzo, and Y. Liu, ''Boundary integral equations for thin bodies,'' Int. J. Numer. Methods Eng. **37**, 107–121 (1994).
[5] 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).
[6] A. J. Burton and G. F. Miller, ''The application of integral equation methods to the numerical solutions of some exterior boundary value problems,'' Proc. R. Soc. London, Ser. A **323**, 201–210 (1971).
[7] 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).
[8] T. W. Wu, ''A direct boundary element method for acoustic radiation and scattering from mixed regular and thin bodies,'' J. Acoust. Soc. Am. **97**, 84–91 (1995).
[9] Y. Liu and F. J. Rizzo, ''Scattering of elastic waves from thin shapes in three dimensions using the composite boundary integral equation formulation,'' J. Acoust. Soc. Am. **102**, 926–932 (1997).
[10] P. M. Juhl, ''An axisymmetric integral equation formulation for free space nonaxisymmetric radiation and scattering of a known incident wave,'' J. Sound Vib. **163**, 397–406 (1993).
[11] P. M. Morse, *Vibration and Sound*, 2nd ed. (The Acoustical Society of America, New York, 1983).
[12] P. M. Juhl, ''A note on the convergence of the direct collocation boundary element method,'' J. Sound Vib. **212**, 703–719 (1998).
[13] J. J. Bowman, T. B. A. Senior, and P. L. E. Uslenghi, *Electromagnetic and Acoustic Scattering by Simple Shapes* (North-Holland, Amsterdam, 1969).
[14] C. Flammer, *Spheroidal Wave Functions* (Stanford University Press, Stanford, 1957).
[15] S. Zhang and J. Jin, *Computation of Special Functions* (Wiley, New York, 1996).

Wave reflections from duct terminations

A. Selamet and Z. L. Ji
Department of Mechanical Engineering and The Center for Automotive Research, The Ohio State University, Columbus, Ohio 43210-1107

R. A. Kach
Powertrain Operations, Ford Motor Company, Dearborn, Michigan 48121

(Received 12 October 2000; accepted for publication 18 December 2000)

The reflection coefficients and inertial end corrections of several duct terminations, including finite length duct extensions perpendicular to an infinite wall, as well as at a number of angles, curved interface surfaces, and annular cavities, are determined and analyzed in the absence of flow by employing the boundary element method. Predictions for the classical unflanged and flanged circular ducts show good agreement with analytical and computational results available in the literature. The predictions for curved interface surfaces (bellmouth or horn) are also consistent with the available experimental data. In view of its high reflection coefficient, the duct termination with an annular cavity may be suggested for the suppression of noise radiation in a specific frequency band or for an effective wave reflection from the termination. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1348298]

PACS numbers: 43.20.Mv [ANN]

I. INTRODUCTION

Part of the acoustic wave reaching a duct exit is reflected back to the source and the remainder radiates into the surroundings, presenting a coupled phenomenon between the sound propagation in the duct and external sound radiation at this boundary. The reflection and radiation characteristics depend strongly on the geometry of duct termination. Thus, the reflection coefficient amplitude and the inertial end correction (or the complex reflection coefficient) of duct termination are the important parameters for the complete acoustic analysis of duct system.

For an infinite circular duct without flange, Levine and Schwinger[1] presented an elaborate analytical derivation for the reflection coefficient and end correction in the absence of mean flow. While the numerical results were given as a function of Helmholtz number, $H \equiv ka$, the limiting value of the end correction δ was determined to be $\delta/a = 0.6133$ as $H \to 0$, with k being the wave number and a the duct radius. Davies *et al.*[2] provided empirical fits to end correction of Ref. 1 as $\delta/a = 0.6133 - 0.1168H^2$ for $H < 0.5$ and $\delta/a = 0.6393 - 0.1104H$ for $0.5 < H < 2$.

For a circular duct with an infinite rigid flange, Rayleigh[3] obtained an approximate value of the end correction as $\delta/a = 0.824\,22$, while Daniell[4] provided the bounds as $0.821\,41 < \delta/a < 0.821\,68$ at zero frequency. This estimate was later recalculated by Nomura *et al.*[5] as $\delta/a = 0.8217$, by Kergomard and Garcia[6] as $\delta/a = 0.261\,53\pi = 0.821\,62$, and by Norris and Sheng[7] as $\delta/a = 0.821\,59$. Nomura *et al.*[5] used Weber–Schafheitlin integrals and Jacobi polynomials to derive a coupled system of two infinite sets of linear equations for the unknowns, which are then solved numerically. An approximate expression, $\delta/a = 0.8217 - 0.367H^2$, was provided by Peters *et al.*[8] for $H < 0.5$. Norris and Sheng[7] also presented an analytical approach for this configuration. While the method also involves numerical computation, only a single set of equations needs to be solved for the modal amplitudes in the duct. They implemented a rational function approximation to get the expressions for the reflection coefficient and end correction of the flanged and unflanged ducts. Ando[9] and Peters *et al.*[8] included the effect of duct wall thickness on the reflection coefficient and end correction. Their studies have demonstrated that, as the thickness of the duct wall is increased, these two quantities vary between the limits of the unflanged duct and infinite flanged duct at lower Helmholtz numbers. Ando's work has been reconsidered later by Bernard and Denardo.[10] Peters *et al.*[8] also experimentally investigated the reflection coefficient and end correction of a duct terminated by a horn with a radius of curvature twice the duct diameter. Their measurements revealed $\delta/a \cong 2.3$ for low H, and an increasing δ/a with increasing H for $H < 0.3$.

While the foregoing studies have provided information for some basic configurations, there remains a significant need to determine the reflection coefficient and the end correction for a number of other duct terminations that are frequently encountered in practice. The objective of the present study is then to calculate and analyze these two quantities for a chosen set of duct terminations, including finite-length duct extensions (perpendicular as well as at an oblique angle to the wall), bellmouthed interfaces, and interfaces with annular cavities.

Following the Introduction, Sec. II provides the formulation for the calculation of reflection coefficient amplitude and inertial end correction of duct termination, and Sec. III develops a numerical approach based on the subdomain boundary element method. The results from the predictions are discussed in Sec. IV. The study is concluded with final remarks in Sec. V.

II. FORMULATION

The duct terminations considered in this study include (1) a finite duct, (2) an extended duct from an infinite rigid

wall, (3) a duct with bellmouth, (4) a duct with annular uniform cavity, and (5) a duct with annular step cavity. Assuming plane wave propagation in the duct, the acoustic load impedance of duct may be expressed as

$$Z_l = \frac{p_l}{\rho c \nu_l} = \frac{Z_r \cos(kl) + j \sin(kl)}{jZ_r \sin(kl) + \cos(kl)}, \tag{1}$$

where p_l and ν_l are the acoustic pressure and velocity, ρ is the density, c is the speed of sound, Z_r is the equivalent radiation impedance at the exit, k is the wave number, l is the length of the duct, and j is the imaginary unit. Equation (1) may be rearranged to solve for Z_r as

$$Z_r = \frac{Z_l \cos(kl) - j \sin(kl)}{\cos(kl) - jZ_l \sin(kl)}. \tag{2}$$

The complex reflection coefficient R at the exit of duct can then be expressed in terms of Z_r as

$$R = -|R| \exp(-j2k\delta) = \frac{Z_r - 1}{Z_r + 1}, \tag{3}$$

where δ is the inertial end correction of the duct. Thus, once Z_l is determined, Eq. (2) gives Z_r and Eq. (3) the reflection coefficient amplitude $|R|$ and the inertial end correction δ. To evaluate the acoustic load impedance Z_l that accounts for the multidimensional waves, particularly in the neighborhood of duct exit, a subdomain boundary element approach is developed, as described next.

III. NUMERICAL APPROACH

To employ the boundary element method in determining the acoustic load impedance of the duct, the sound field is first divided into the internal and the external acoustic domains separated by the opening surface.

A. Internal acoustic domain

The internal domain may also be divided into several subdomains. For the internal domain or each subdomain, the boundary integral expression is represented[11] as

$$C(X)P(X) = \int_\Gamma \left[G(X,Y) \frac{\partial P}{\partial n}(Y) - P(Y) \frac{\partial G}{\partial n}(X,Y) \right] d\Gamma(Y), \tag{4}$$

where Γ is the boundary surface of the acoustic domain, n is the unit normal vector on Γ directed away from the domain, the function $G(X,Y) = \exp(-jkR)/4\pi R$ is Green's function of free space, R being the distance between any two points X and Y in the domain or on the surface, and $C(X)$ is a coefficient which depends on the position of point X and may be evaluated[11] by

$$C(X) = \int_\Gamma \frac{\partial}{\partial n} \left(\frac{1}{4\pi R} \right) d\Gamma(Y). \tag{5}$$

A numerical solution of the boundary integral Eq. (4) can be achieved by discretizing the boundary surface of the domain into a number of elements. By using discretization and numerical integration, the following algebraic system of equations is obtained[11]

$$[F]\{P\} = \rho c [G]\{V\}, \tag{6}$$

where $[F]$ and $[G]$ are the coefficient matrices, and $\{P\}$ and $\{V\}$ are the vectors of sound pressure and outward normal particle velocity at boundary nodes, respectively, ρc is the characteristic impedance of the medium. The boundaries are grouped into the inlet, opening and wall, represented by the subscripts i, o, and w, respectively. Equation (6) combined with the rigid wall boundary condition yields[12]

$$\begin{Bmatrix} P_i^d \\ P_o^d \end{Bmatrix} = \rho c [T^d] \begin{Bmatrix} V_i^d \\ V_o^d \end{Bmatrix}, \quad \text{(for the duct)}, \tag{7}$$

$$\{P_o^c\} = \rho c [T^c]\{V_o^c\}, \quad \text{(for the cavity)}. \tag{8}$$

Combining duct and cavity gives

$$\begin{Bmatrix} P_i^d \\ P_o^i \end{Bmatrix} = \rho c [T^i] \begin{Bmatrix} V_i^d \\ V_o^i \end{Bmatrix}, \tag{9}$$

where

$$\{P_o^i\} = \begin{Bmatrix} P_o^d \\ P_o^c \end{Bmatrix}, \quad \{V_o^i\} = \begin{Bmatrix} V_o^d \\ V_o^c \end{Bmatrix},$$

and

$$[T^i] = \begin{bmatrix} T^d & 0 \\ 0 & T^c \end{bmatrix}$$

is the transfer impedance matrix between the inlet and opening for the internal acoustic domain.

B. External acoustic domain

For the full-space external sound field, the boundary integral expression is represented[11] as

$$C^e(X)P^e(X) = \int_\Gamma \left[G(X,Y) \frac{\partial P^e}{\partial n}(Y) - P^e(Y) \frac{\partial G}{\partial n}(X,Y) \right] d\Gamma(Y), \tag{10}$$

$$C^e(X) = 1 - \int_\Gamma \frac{\partial}{\partial n} \left(\frac{1}{4\pi R} \right) d\Gamma(Y). \tag{11}$$

Similar to the formulation of internal acoustics, discretization and numerical integration of Eq. (10) lead to the following algebraic system of equations:

$$[F^e]\{P^e\} = \rho c [G^e]\{V^e\}. \tag{12}$$

The boundaries are grouped into the opening and wall, represented by subscripts o and w, respectively. Equation (12) combined with the rigid wall boundary condition yields

$$\{P_o^e\} = \rho c [T^e]\{V_o^e\}, \tag{13}$$

$[T^e]$ being the radiation impedance matrix for the external acoustic domain.

For the external acoustic domain with duct extension from an infinite rigid wall, the boundary integral equation can be written[11] as

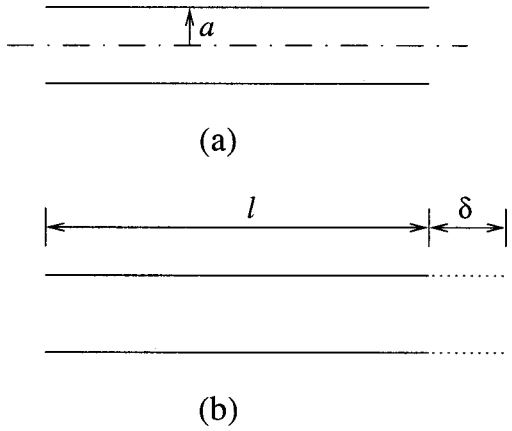


FIG. 1. Unflanged finite duct.

$$C^e(X)P^e(X)=\int_{\Gamma_w+\Gamma_o}\left[G_H(X,Y)\frac{\partial P^e}{\partial n}(Y)-P^e(Y)\frac{\partial G_H}{\partial n}(X,Y)\right]d\Gamma(Y), \tag{14}$$

where Γ_w is the wall of extended duct from the infinite wall, Γ_o is the opening,

$$G_H=\frac{\exp(-jkR)}{4\pi R}+\frac{\exp(-jkR_1)}{4\pi R_1}$$

is the half-space Green's function, R_1 being the distance between Y and the image point of X with respect to the flange; and

$$C^e(X)=1-\int_\Gamma\frac{\partial}{\partial n}\left(\frac{1}{4\pi R}\right)d\Gamma(Y), \tag{15}$$

Γ being the total boundary of the duct extended from the infinite wall. Similar to the full-space problem, the acoustic radiation impedance matrix $[T^e]$ for this case can be obtained by using discretization and numerical integration of Eq. (14). For the limiting case of no-extension (a duct with infinite rigid flange), the sound pressure at the opening may be expressed as

$$P^e(X)=\int_{\Gamma_o}\left[\frac{\exp(-jkR)}{2\pi R}\frac{\partial P^e}{\partial n}(Y)\right]d\Gamma(Y). \tag{16}$$

Similarly, the acoustic radiation impedance matrix $[T^e]$ for the semi-infinite external acoustic domain can also be obtained by using discretization and numerical integration of Eq. (16).

C. Coupling

At the opening, the solution should satisfy the continuity conditions of sound pressure and particle velocity:

$$\{P_o^e\}=\{P_o^i\}, \tag{17}$$

$$\{V_o^e\}=-\{V_o^i\}. \tag{18}$$

Combining Eqs. (9) and (13) with the continuity conditions (17) and (18) yields the acoustic load impedance as

$$Z_l=[T_{11}^i-T_{12}^i(T_{22}^i+T^e)^{-1}T_{21}^i]. \tag{19}$$

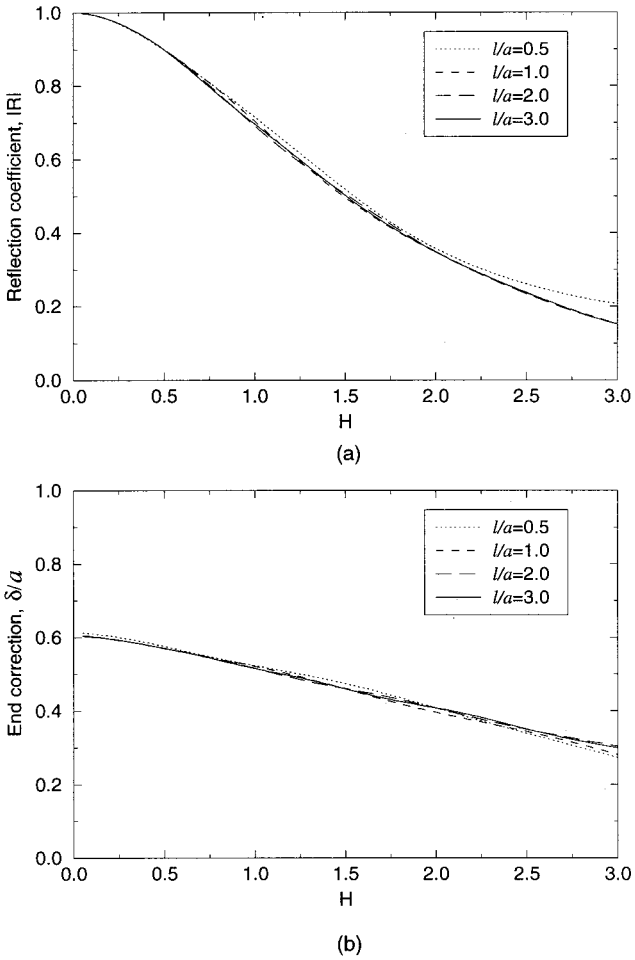


FIG. 2. Unflanged finite duct: (a) reflection coefficient and (b) end correction.

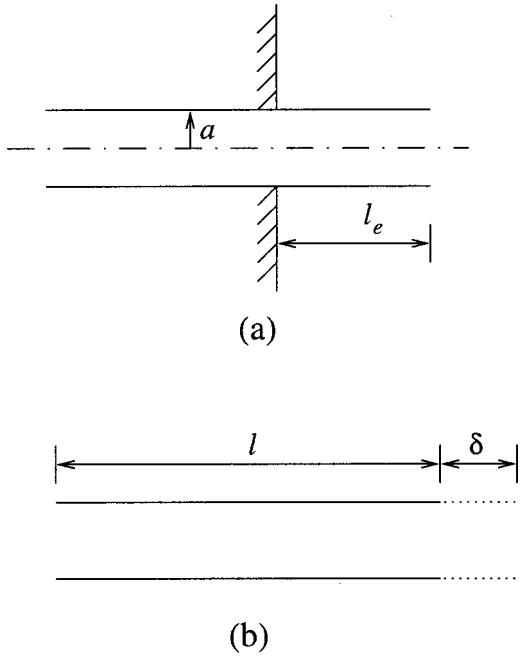


FIG. 3. Duct extended from an infinite rigid flange.

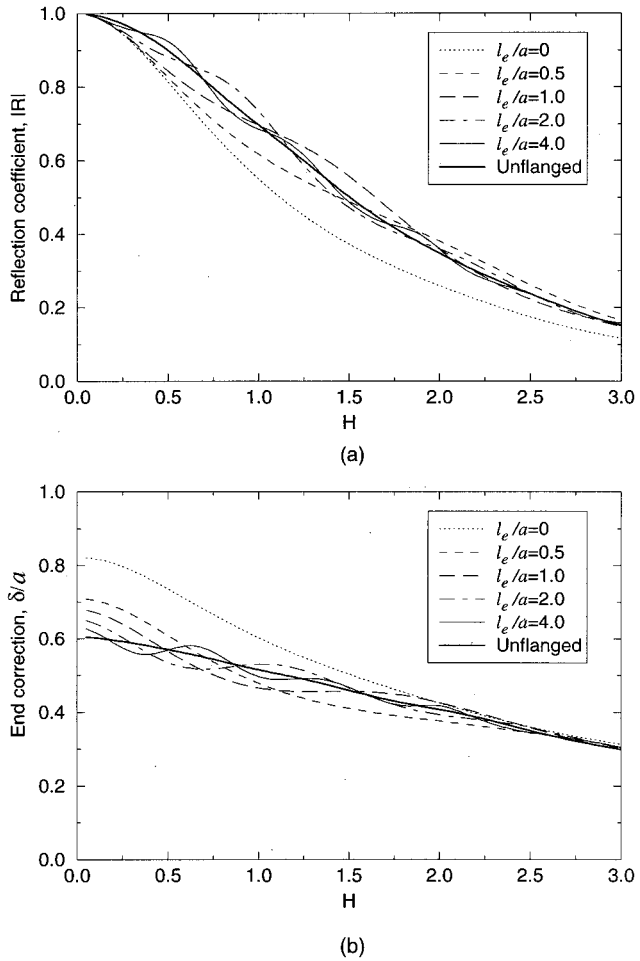


FIG. 4. Duct extended from an infinite rigid flange: (a) reflection coefficient and (b) end correction.

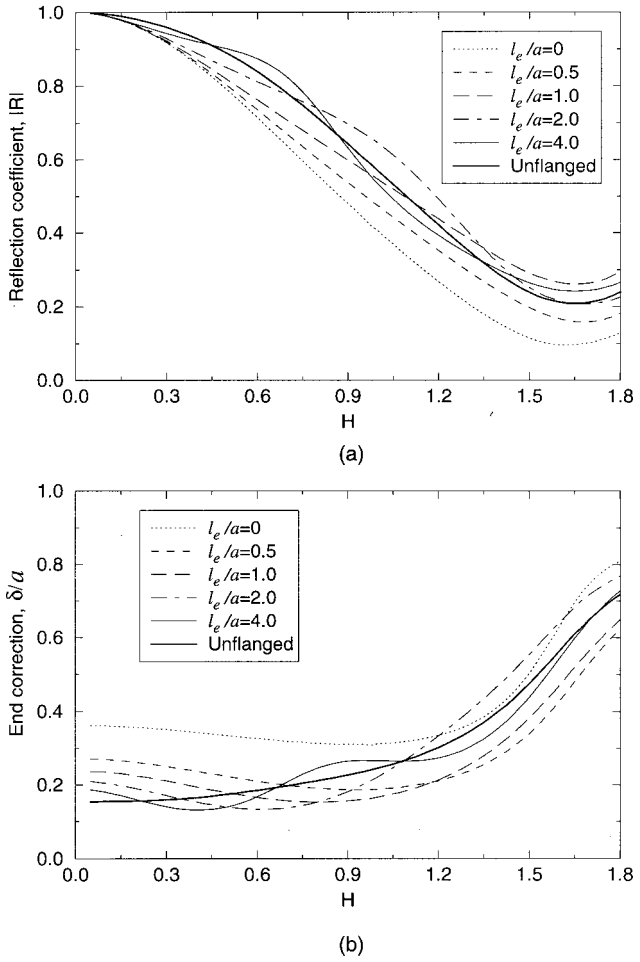


FIG. 6. Duct at an angle $\theta=45°$ extended from an infinite rigid flange: (a) reflection coefficient and (b) end correction.

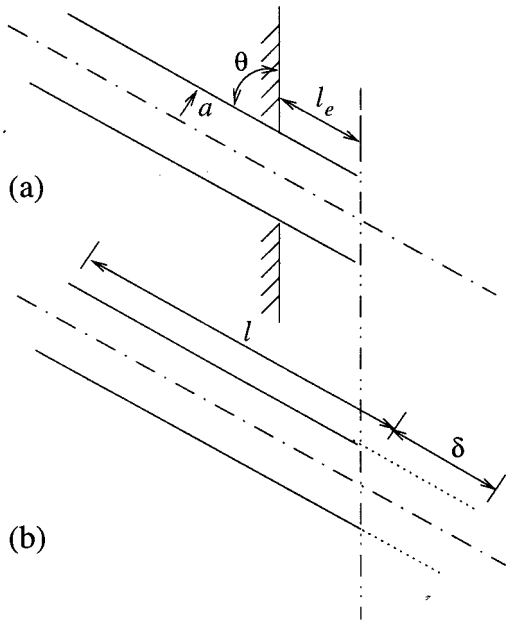


FIG. 5. Duct at an angle θ extended from an infinite rigid flange.

IV. RESULTS AND DISCUSSION

First, the effect of length of an unflanged finite duct (Fig. 1) on the reflection coefficient and end correction is examined. Figure 2 shows the numerical predictions of reflection coefficient and end correction for four different duct lengths ($l/a=0.5, 1.0, 2.0, 3.0$). For a duct with $l/a>2$, the results illustrate that the effect of duct length on these two quantities becomes negligible in the region of $H<3$, and the present predictions agree with the analytical results of Levine and Schwinger[1] for the classical configuration of the infinite duct without flange.

The effect of duct extension l_e from the infinite rigid wall (Fig. 3) on the reflection coefficient and end correction is shown in Fig. 4. The finite extension introduces an oscillatory behavior on both quantities as a function of H due to the presence of an infinite rigid wall. The period of oscillations in H may be crudely (via a one-dimensional argument) related to two subsequent quarter wave resonances between the duct opening and the infinite rigid wall leading to $\Delta H \cong \pi/[(l_e+\delta)/a]$. As expected, the oscillatory behavior diminishes as l_e/a becomes large. For the limiting case of no-extension ($l_e/a=0$, flanged duct), the present predictions of the reflection coefficient and end correction are in good agreement with the available literature (Rayleigh,[3] Daniell,[4]

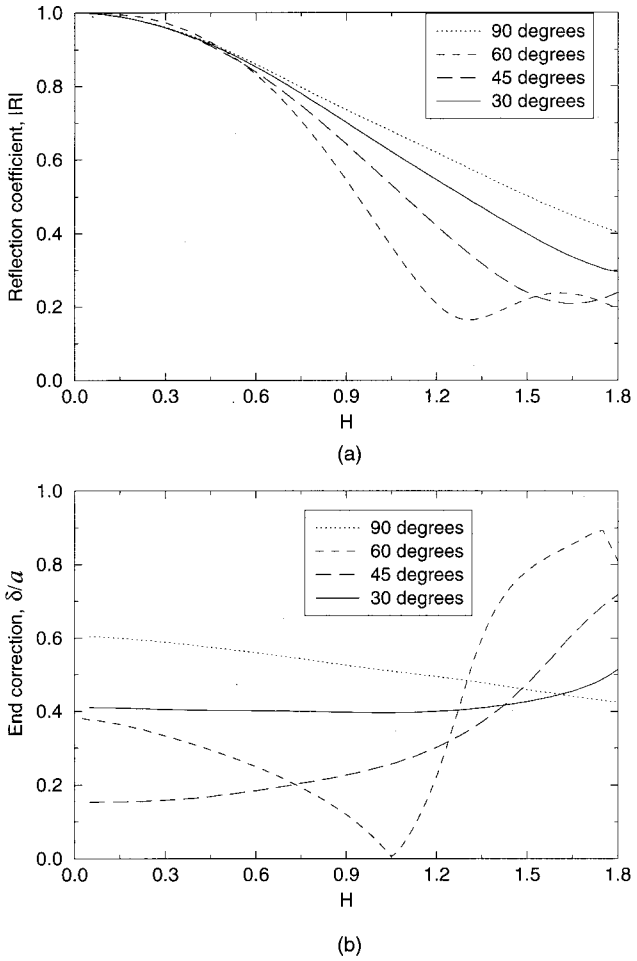


FIG. 7. Unflanged ducts with θ=30°,45°,60°,90°: Comparison of (a) reflection coefficient and (b) end correction.

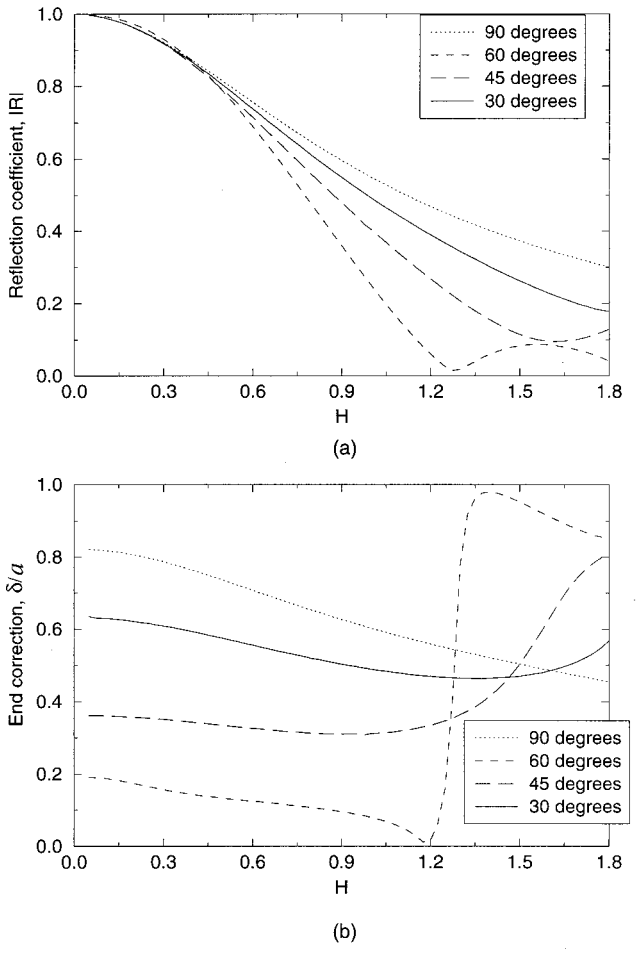


FIG. 8. Flanged ducts with $l_e=0$ and θ=30°,45°,60°,90°: Comparison of (a) reflection coefficient and (b) end correction.

Nomura *et al.*,[5] Kergomard and Garcia,[6] and Norris and Sheng[7]). The present predictions in Fig. 4(b) for $l_e/a=0$ agree very well with Fig. 7 of Nomura *et al.*[5] to the degree that the two sets of results cannot be distinguished from each other over $0<H<3.0$, including, for example, a specific value of $\delta/a=0.821$ for $H=0.05$.

The predictions for the reflection coefficient and end correction for a duct extension from an infinite rigid wall with an oblique angle of θ (Fig. 5) are shown in Fig. 6 for θ=45°. Similar to the duct extension with 90° angle, the oscillatory behavior is observed again. Figures 7 and 8 compare the reflection coefficients and end corrections for, respectively, the unflanged and flanged ducts with $l_e=0$ and θ=30°,45°,60°,90°. At higher Helmholtz numbers ducts with larger θ provide, in general, higher reflection coefficients. In view of the definition of δ in Fig. 5, a reduction on the order of radius may be anticipated for θ=45° (relative to θ=90°) at low Helmholtz numbers as shown in Fig. 7. Note that $H\leqslant 1.8$ in Figs. 6–8 to remain below the first diametral mode (1,0).

The acoustic characteristics of a duct with bellmouth (Fig. 9) are examined next to understand the effect of radius of curvature on the reflection coefficient and end correction. The latter is defined relative to the end of the straight section of the duct, consistent with Peters *et al.*[8] Figures 10 and 11 depict the computed results for these two quantities for five different radii of curvature (r/a=0.5,1.0,2.0,3.0,4.0), and without and with an infinite flange, respectively. With increasing radius of curvature, the reflection coefficient decreases and the end correction increases. For the special case of $r/a=4$, the present predictions agree with the experimental results of Peters *et al.*[8] who determined $\delta/a\cong 2.3$ at low values of H. For this special case at $H=0.025$, for example, the predicted numerical values are $\delta/a=2.322$ and

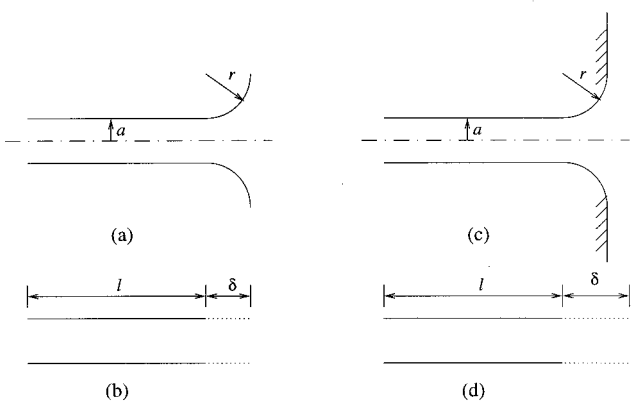


FIG. 9. Ducts with unflanged and flanged bellmouth.

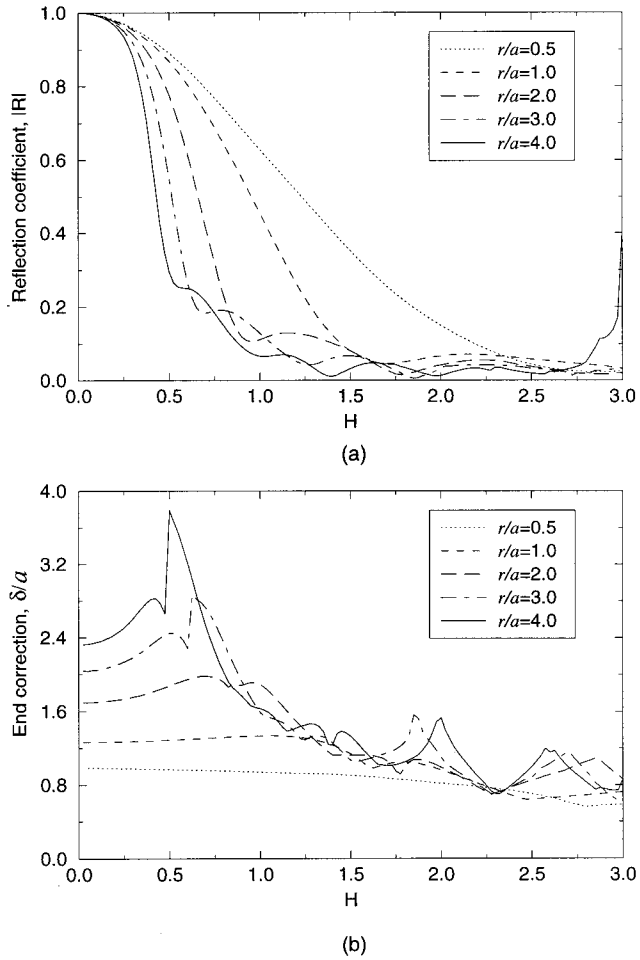


FIG. 10. Duct with unflanged bellmouth: (a) reflection coefficient and (b) end correction.

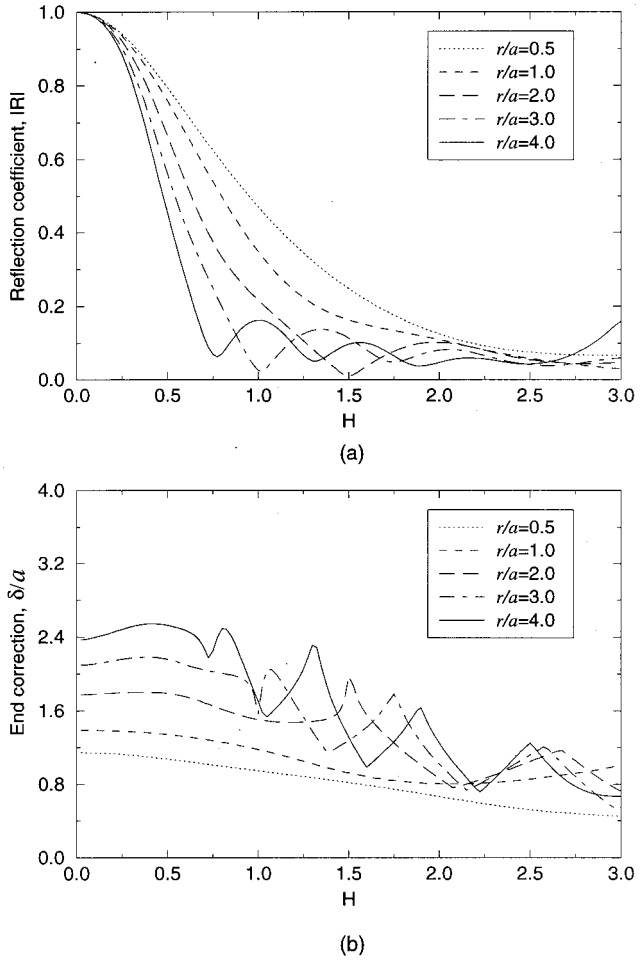


FIG. 11. Duct with flanged bellmouth: (a) reflection coefficient and (b) end correction.

$\delta/a=2.369$ for Fig. 10(b) and Fig. 11(b), respectively. The differences in these configurations beyond the radiused inlet do not have a significant impact on the end correction for low H. In comparison with the straight duct, the duct with bellmouth leads to much lower wave reflection at higher Helmholtz numbers. Thus the bellmouth with large radius of curvature should be avoided, for example, at the exhaust exit in order to promote effective pressure reflections from the termination for noise control. Figure 12 shows the effect of curved interface with the radiused length subtracted from the end correction, transforming the dimensionless end correction to $(\delta-r)/a$. Such manipulation essentially shifts the reference point for end correction to the very end of the duct (end of bellmouth section). Relative to the end of the curved duct, Fig. 12 then suggests that a duct with a large-radius termination, such as $r/a=4$, will appear to be shorter than the actual total length when the end correction is incorporated.

Finally, ducts with annular cavities (unflanged and flanged) are considered as shown in Fig. 13 with a constant cross section and Fig. 14 with a step transition. The computed results for the reflection coefficient and the end correction are compared in Fig. 15 for $l_c/c=4$ and $c/a=1/2$. The results with cavities here may be contrasted with the ducts alone of Figs. 2 and 4. As a result of the interaction of waves reflected from attached cavities with those radiated from the opening, it is important to note that (1) the behavior of reflection coefficient changes dramatically including the additional peaks, therefore the sound radiation may be suppressed effectively in a specific frequency band, and/or (2) the pressure wave reflections from the termination back to the duct would be strengthened. Compared to the uniform cavity, the step cavity moves the first additional peak to a higher Helmholtz number, and the second peak to a lower Helmholtz number. The high reflection coefficient, in a specific frequency band, of duct with annular cavity may be used to compensate the low attenuation in the corresponding frequency region of a silencer, therefore an improved acoustic attenuation performance may be obtained. Similar behavior can possibly be utilized to promote stronger pressure wave reflections from the opening, thereby providing mechanisms to manipulate engine tuning.

V. CONCLUDING REMARKS

This study has shown the effect of duct termination geometry on the reflection coefficient and the inertial end correction by employing the subdomain boundary element approach. These two quantities were determined and analyzed for (1) an unflanged finite duct, (2) a duct extended from an infinite rigid wall, (3) a duct extended with an angle θ from

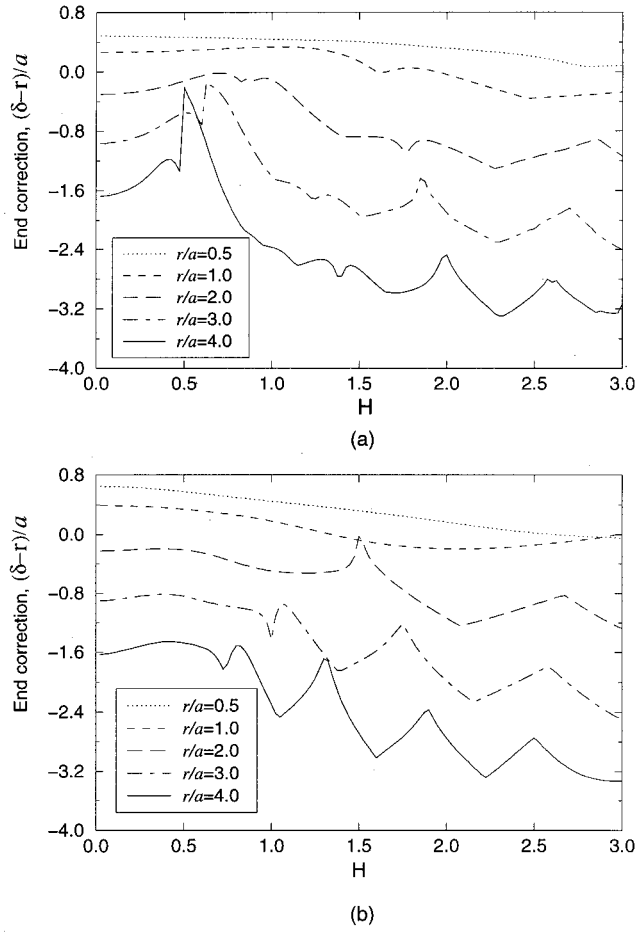


FIG. 12. $(\delta - r)/a$ for ducts with (a) unflanged bellmouth and (b) flanged bellmouth.

an infinite rigid wall, (4) ducts with unflanged and flanged bellmouth, and (5) ducts with unflanged and flanged annular cavities. The effect of flow, which was well established in the literature for certain configurations (see, for example, Peters *et al.*[8]), as a function of Mach and Strouhal numbers was excluded from the study. Instead, the current work has chosen to concentrate on quantifying the impact of interface geometry which has not been investigated in detail with the exception of limiting cases. For unflanged configurations, a finite duct with $l/a > 2$ exhibits similar acoustic characteris-

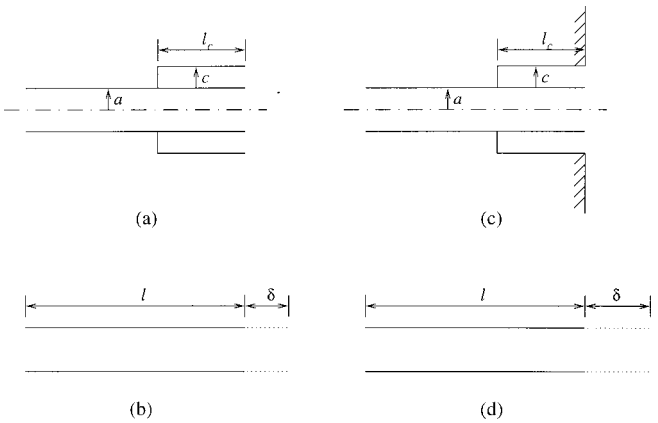


FIG. 13. Ducts with unflanged and flanged annular uniform cavity.

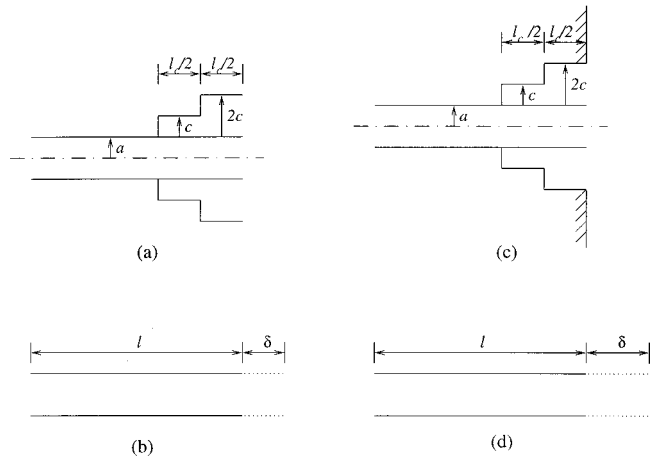


FIG. 14. Ducts with unflanged and flanged annular step cavity.

tics to an infinite duct at low Helmholtz numbers. The duct extension from an infinite rigid wall introduces an oscillatory behavior in the reflection coefficient and end correction near the baseline unflanged duct due to the interaction of waves reflected from the infinite wall and those radiated from the duct opening. Within about a diameter-distance from the

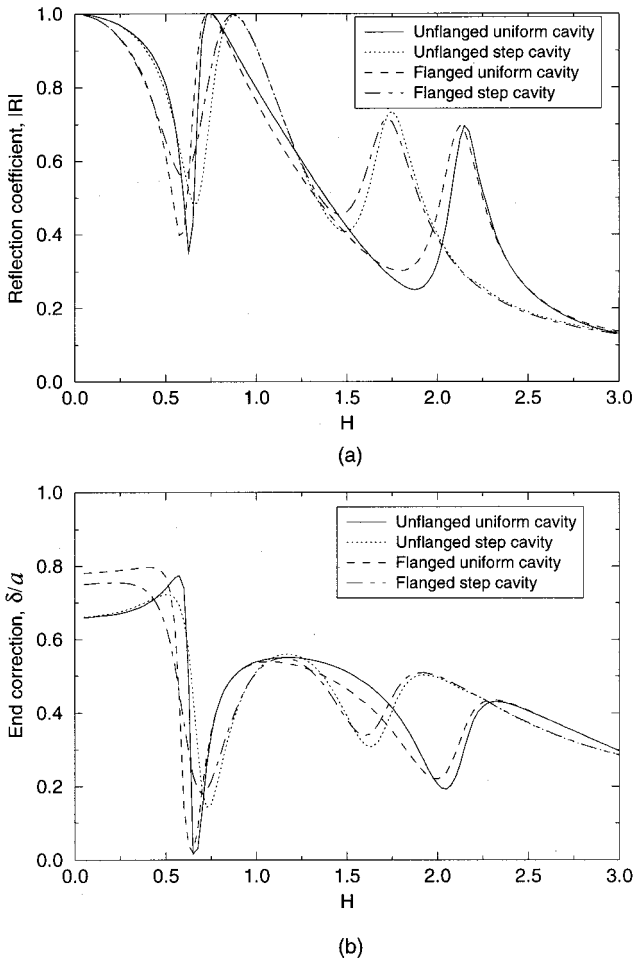


FIG. 15. Ducts with unflanged and flanged annular cavity with $l_c/c = 4$ and $c/a = 1/2$: (a) reflection coefficient and (b) end correction.

wall, the acoustic behavior of the two quantities approaches that of an unflanged wall with the exception of oscillations. The oblique duct extension from an infinite rigid wall, in general, decreases the reflection coefficient of a duct at higher Helmholtz numbers as compared to the perpendicular extension. The bellmouth interface reduces the reflection coefficient significantly at higher Helmholtz numbers. The duct termination with cavity influences end correction and introduces additional peaks in the reflection coefficient and may be used for the suppression of noise radiation, or to promote effective wave reflection from the open end.

[1] H. Levine and J. Schwinger, ''On the radiation of sound from an unflanged circular pipe,'' Phys. Rev. **73**, 383–406 (1948).

[2] P. O. A. L. Davies, J. L. Bento Coelho, and M. Bhattacharya, ''Reflection coefficients for an unflanged pipe with flow,'' J. Sound Vib. **72**, 543–546 (1980).

[3] Lord Rayleigh, *The Theory of Sound* (Dover, New York, 1945).

[4] P. J. Daniell, ''The coefficient of end-correction,'' Philos. Mag. **30**, 137–146, 248–256 (1915).

[5] Y. Nomura, I. Yamamura, and S. Inawashiro, ''On the acoustic radiation from a flanged circular pipe,'' J. Phys. Soc. Jpn. **15**, 510–517 (1960).

[6] J. Kergomard and A. Garcia, ''Simple discontinuities in acoustic waveguides at low frequencies: Critical analysis and formulae,'' J. Sound Vib. **114**, 465–479 (1987).

[7] A. N. Norris and I. Sheng, ''Acoustic radiation from a circular pipe with an infinite flange,'' J. Sound Vib. **135**, 85–93 (1989).

[8] M. C. A. M. Peters, A. Hirschberg, and A. J. Reijnen, ''Damping and reflection coefficient measurements for an open pipe at low Mach and low Helmholtz numbers,'' J. Fluid Mech. **256**, 499–534 (1993).

[9] Y. Ando, ''On the acoustic radiation from semi-infinite circular pipe of certain wall thickness,'' Acustica **22**, 219–225 (1969/1970).

[10] M. Bernard and B. Denardo, ''Re-computation of Ando's approximation of the end correction for a radiating semi-infinite circular pipe,'' Acustica **82**, 670–671 (1996).

[11] R. D. Ciskowski and C. A. Brebbia, *Boundary Element Methods in Acoustics* (Computational Mechanics Publications, Boston, 1991).

[12] Z. L. Ji, Q. Ma, and Z. H. Zhang, ''Application of the boundary element method to predicting acoustic performance of expansion chamber mufflers with mean flow,'' J. Sound Vib. **173**, 57–71 (1994).

Modeling of high-frequency scattering from objects using a hybrid Kirchhoff/diffraction approach

J. A. Fawcett
Defence Research Establishment Atlantic, PO Box 1012, Dartmouth, Nova Scotia B2Y 3Z7, Canada

(Received 8 June 1999; revised 18 September 2000; accepted 21 December 2000)

In this paper a numerical discretization and transformation into the time domain of a hybrid Kirchhoff/diffraction method is presented for the modeling of high-frequency pulse scattering from rigid bodies. A series of benchmark cases, for smooth and rough spheres and cylinders, is presented to establish the accuracy of the method in the time domain. [DOI: 10.1121/1.1350400]

PACS numbers: 43.20.Px, 43.20.Fn [ANN]

I. INTRODUCTION

The Kirchhoff approximation (see, for example, Refs. 1–5) is often used to model high-frequency scattering from rough surfaces[1–3] or objects.[4,5] At a point on the scattering surface, the surface is considered to be locally flat and a tangent–plane approximation is used to relate the field on the surface to the incident field. The Helmholtz integral equation

$$p_{sc}(\mathbf{r}) = \int_S [G_n(\mathbf{r}|\mathbf{r}')p(\mathbf{r}') - G(\mathbf{r}|\mathbf{r}')p_n(\mathbf{r}')]dS \qquad (1)$$

is then used to compute the scattered pressure field at a spatial point $\mathbf{r}$ where G is the free-space Green's function defined below in Eq. (3). In this paper, we only consider the case that the object is rigid; in this case the value of the pressure field on the object is approximated as twice the value of the incident field and Eq. (1) becomes

$$p_{sc}(\mathbf{r}) = 2\int_S G_n(\mathbf{r}|\mathbf{r}')p_{inc}(\mathbf{r}')dS. \qquad (2)$$

For an object or surface that has sharp edges, such as a flat-ended cylinder, the Kirchhoff approximation does not yield the correct scattering behavior from these edges. This shortcoming has been discussed by some authors who have derived the correct diffraction behavior for edges.[6,7] In the case of a smooth body with edges, it is necessary to combine the Kirchhoff (or physical optics) solution with the diffraction solution in such a way that the hybrid solution is valid for both the diffractive and smooth body effects. This issue has been addressed by various authors[8–10] who have derived uniformly valid solutions. In Ref. 5 the authors combine the Kirchhoff approximation for a smooth flat-ended cylinder with Keller's[6] diffraction theory to yield a hybrid solution for scattering from a cylinder with ribs. Sammelman *et al.*[11] have described a hybrid approach to modeling ''smooth,'' diffractive, and point-scattering effects with a facet model.

In Sec. II we summarize the theory due to Tran Van Nhieu[10] which models uniformly the smooth and diffraction scattering from a rigid three-dimensional body. An efficient time-domain numerical implementation of this frequency-domain theory is then presented in Sec. III. A set of numerical examples is presented comparing this method, in the time domain, to the results of analytic and boundary integral equation method (BIEM) benchmark computations. Scattering model comparisons are often presented in the frequency domain (perhaps, for multiple frequencies) in terms of, for example, scattering strength as a function of angle. For broadband pulse scattering measurements, it is, in fact, the time series which is of interest. It is sometimes difficult to deduce how approximation errors in the frequency domain will manifest themselves in the time domain, and so in this paper, we emphasize time-domain computations. In the time domain it is sometimes possible to see at what time and possibly where on the scatterer an approximation begins to fail. Some of the numerical benchmarks presented in this paper are for standard objects, such as a sphere and a flat-ended cylinder. Some of the examples, however, involve significantly corrugated spheres and cylinders, where multiple scattering or shadowing effects could possibly be significant. The benchmark computations for all these cases, when transformed into the time domain, provide an interesting set of test cases for the Kirchhoff/diffraction method. The advantage of the approximate method described in this paper is that it is at least a couple of orders of magnitude faster than the BIEM and, as we shall see, the accuracy is often very good.

II. THEORY

We first briefly describe the concepts of Kirchhoff modeling for a smooth body (i.e., there are no discontinuities in the derivatives of the function describing the surface of the body) and then briefly describe the approach taken in Ref. 10 to uniformly correct for edge. In Sec. III we then describe our numerical discretization of the edge-corrected Kirchhoff formulation and its analytic transformation into the time domain.

The Kirchhoff approximation to scattering is given by Eq. (2). For the Green's function, we use

$$G(\mathbf{r}|\mathbf{r}') = \frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{4\pi R}, \qquad (3)$$

and the normal derivative is approximated by the far-field expression,

$$G_n \equiv ik\hat{r}\cdot\hat{n}\frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{4\pi R}, \tag{4}$$

where $R \equiv |\mathbf{r}-\mathbf{r}'|$.

For some simple shapes, such as a flat-ended cylinder, it is possible to perform the surface integral of Eq. (2) analytically or semianalytically using the far-field approximation for the Green's function.[4] In general, for azimuthally symmetric objects, the azimuthal integral can be expressed in terms of special functions and the remaining ''vertical'' integration can be done numerically. However, in this paper, although most of our examples fall into this class, we consider a numerical discretization of the surface integral, using triangular facets, which is applicable to general surfaces.

The expressions for the field diffracted by an edge have been given by various authors.[5–7] For sufficiently high frequencies and at scattering angles away from specular, these expressions yield an accurate approximation to the scattered field. However, at angles near specular, these diffraction expressions are not accurate. On the other hand, the Kirchhoff approximation does model near-specular scattering accurately, but is inaccurate for the scattering from edges. We use the approach of Tran Van Nhieu[10] to combine these solutions.

In Ref. 10 the field diffracted by an edge, p_d, is expressed as a line integral around the edge. Also, an expression for the edge contributions (which are incorrect) of the Kirchhoff expression p_e^k is derived in terms of a line integral. A uniform expression is then obtained by subtracting p_e^k from the Kirchhoff solution, $p^k(\mathbf{r})$ and adding in the correct edge terms; i.e.,

$$p(\mathbf{r}) = (p^k(\mathbf{r}) - p_e^k(\mathbf{r})) + p_d(\mathbf{r}). \tag{5}$$

Let us consider an incident plane wave with wave vector $\mathbf{k}$ (it is straightforward to generalize this to other incident fields such as that due to a point source). An expression for the diffracted field in terms of a source distribution along the curve, C, is given by

$$p_d = \frac{i}{2\pi}\int_C \xi(\alpha,l)H_e(l)\frac{e^{i\mathbf{k}\cdot\mathbf{r}'(l)+k\rho}}{\rho}dl, \tag{6}$$

where α, ξ, and $H_e(l)$ are defined below. Along the curve of the edge three unit vectors are defined: $\hat{e}_1$ is tangent to the curve edge, $\hat{e}_2$ is perpendicular to $\hat{e}_1$ and is pointed inwards along the top surface, and $\hat{e}_3 = \hat{e}_2 \times \hat{e}_1$. For example, for a cylinder aligned along the z axis the edges at the upper and lower plates are a circle; $\hat{e}_1$ is tangent to the circle, $\hat{e}_2$ points inward along the circle's radial vector, and $\hat{e}_3$ is along the z axis. The vector $\hat{x}_r$ is a unit vector in the direction of the receiver and $\hat{k}$ is unit vector oriented opposite to that of the incident plane wave.

Some necessary definitions of angles are

$$\beta = A\cos(\hat{e}_1\cdot\hat{x}_r), \tag{7}$$

$$\beta_s = A\cos(\hat{e}_1\cdot\hat{k}), \tag{8}$$

$$\alpha = A\cos\left(\frac{\hat{e}_2\cdot\hat{x}_r}{\sin(\beta)}\right), \tag{9}$$

$$\alpha_s = A\cos\left(\frac{\hat{e}_2\cdot\hat{k}}{\sin(\beta_s)}\right). \tag{10}$$

The function ξ is then defined as

$$\xi(\alpha) = 1, \quad 0 < \alpha < 2\pi - \sigma, \tag{11}$$

$$\xi(\alpha) = 0, \quad 2\pi - \sigma < \alpha < 2\pi, \tag{12}$$

and

$$H_e \equiv -\frac{(H(\nu,u,\alpha,-\alpha_s) + H((\nu,u,\alpha,\alpha_s))}{\nu}, \tag{13}$$

and

$$H \equiv \frac{e^{i\pi/\nu} - u^2 e^{-i\pi/\nu}}{e^{i\pi/\nu} + u^2 e^{-i\pi/\nu} - 2u\cos((\alpha+\alpha_s)/\nu)}, \tag{14}$$

where $u \equiv (\sin\beta/\sin\beta_s)^{1/\nu}$; ν is the open angle formed by the two tangent planes at the edge. For the example of a flat-ended cylinder, $\nu = 3/2\pi$. Simple expressions[5–7] for the diffraction field due to the edge can be obtained from a stationary phase analysis of the integral of Eq. (6).

Similarly, for the edge contribution of the Kirchhoff solution, we can write

$$p_e^k = \frac{i}{2\pi}\int_C \xi(l)K_e(l)\frac{e^{i\mathbf{k}\cdot r'(l)+k\rho}}{\rho}, \tag{15}$$

where

$$K_e \equiv -\epsilon(\alpha_s)K(\alpha,\alpha_s) - \epsilon(\nu\pi-\alpha_s)K(\nu\pi-\alpha,\nu\pi-\alpha_s), \tag{16}$$

and

$$t = \frac{\sin\beta}{\sin\beta_s}, \quad K(\alpha,\alpha_s) = -\frac{e^{-i\alpha_s} - t^2 e^{i\alpha_s}}{e^{-i\alpha_s} + t^2 e^{i\alpha_s} + 2t\cos\alpha}. \tag{17}$$

The function ϵ is equal to one or zero depending if that section of the face is included in the surface integral. Using the expressions of Eqs. (15) and (6) in Eq. (5), the hybrid Kirchhoff/diffraction formulation is obtained.

III. NUMERICAL DISCRETIZATION AND TRANSFORMATION TO THE TIME DOMAIN

Above, we discussed the theoretical expression for the field scattered by a smooth, rigid body with edges at a single frequency. To obtain a numerical implementation we will first spatially discretize the surface and edge integrals and then we will analytically perform the Fourier transform of these results to yield the time-domain expressions. This is similar in spirit to the approach taken, for example, in Ref. 2, to the standard Kirchhoff approach to rough surface scattering.

The surface of an object is broken into triangular facets; it is assumed that any edges of the object will be covered by the sides of some of the triangular facets. Let us first consider a single triangular facet with the simple Kirchhoff expression applied to it. From Eq. (2) let us approximate the field scattered from a single facet, P_{sc}, as

$$P_{sc} = -\frac{1}{2\pi R}ik\hat{r}\cdot\hat{n}\exp(ik|\mathbf{r}-\mathbf{r}'|)\exp(i\mathbf{k}\cdot\mathbf{r}'). \tag{18}$$

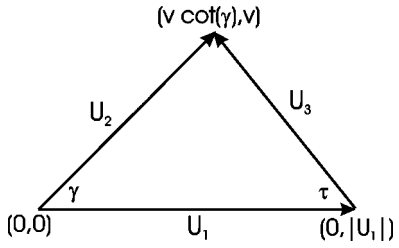


FIG. 1. A schematic of a triangular facet, shown as a two-dimensional projection.

To numerically integrate this expression over the facet, we will consider the amplitude to be a constant, the value at the midpoint of the triangular facet, and allow the phase terms to vary linearly about their midpoint values. Consider the triangle as shown in Fig. 1 with the vectors defined between the three vertices (note that although the triangle is shown on a plane in Fig. 1, it represents a general orientation in three dimensions). We also define two orthogonal unit vectors lying in the plane of the triangle

$$\hat{u} \equiv \hat{u}_1, \tag{19}$$

$$\hat{v} \equiv \frac{\mathbf{u}_2 - (\mathbf{u}_2 \cdot \hat{u})\hat{u}}{|\mathbf{u}_2 - (\mathbf{u}_2 \cdot \hat{u})\hat{u}|} \tag{20}$$

(we use $\hat{u}$ to denote a unit vector), and the two angles

$$\gamma \equiv a \cos(\hat{u}_2 \cdot \hat{u}_1), \tag{21}$$

$$\tau \equiv a \cos(-\hat{u}_3 \cdot \hat{u}_1). \tag{22}$$

Defining

$$a \equiv \hat{u} \cdot (\hat{r} - \hat{k}), \tag{23}$$

$$b \equiv \hat{v} \cdot (\hat{r} - \hat{k}), \tag{24}$$

we can express $\int_S P_{sc}\, dS$ as

$$\frac{1}{2ik\pi}(\hat{r} \cdot \hat{n})[\exp(ik\phi_1)/a_1 + \exp(ik\phi_2)/a_2 + \exp(ik\phi_3)/a_3], \tag{25}$$

where ϕ_1, ϕ_2, ϕ_3 are the phases corresponding to the distances from the three vertices of the triangle to the receiver and the values of the incident phase. The amplitudes are given by

$$a_1 \equiv a(a \cot\gamma + b), \quad a_2 \equiv a(a \cot\tau - b)$$
$$a_3 \equiv -\frac{(a \cot\gamma + b)(b - \cot\tau)}{(\cot\tau + \cot\gamma)}. \tag{26}$$

This is the contribution from a single facet; the total field is computed by adding together all these contributions. It is important to note that Eq. (25) can also be written as

$$\left[\frac{e^{ik\phi_2} - e^{ik\phi_3}}{a_2}\right] - \left[\frac{e^{ik\phi_3} - e^{ik\phi_1}}{a_1}\right]. \tag{27}$$

In this form, it is seen that there is the possibility of small numerators over small denominators. Numerically, we monitor the coefficients of Eq. (26) and if any of them exceed 1000, we use the simpler form for the facet contribution; namely, Eq. (18) evaluated at the midpoint of the facet and multiplied by the area of the facet.

It is straightforward to put Eq. (25) into the time domain; namely, we perform the Fourier transform of the terms $\exp(ik\phi)/ik$ to yield

$$p_{sc}(t) = \frac{1}{2\pi}(\hat{r} \cdot \hat{n})\left[\frac{\overline{S}(t-\Omega_1)}{a_1} + \frac{\overline{S}(t-\Omega_2)}{a_2} + \frac{\overline{S}(t-\Omega_3)}{a_3}\right], \tag{28}$$

where Ω_i is the time delay from point i, S is the incident source pulse, and $\overline{S}$ denotes

$$\overline{S} \equiv \int_{-\infty}^{t} S(\tau)\, d\tau. \tag{29}$$

In other words, $\overline{S}$ is the time integral of the incident source pulse. If the simpler form of numerical integration is used then the derivative of the source pulse, $S'(t-\Omega_i)$ is required.

To account for the diffraction terms we discretize the line integrals of Eq. (6) and Eq. (15) and we will take the line segments to correspond to the side of one of the facet triangles, although this is not necessary. Considering one such line segment, it can be shown that the contribution to the field is given by

$$P_{\text{edge}} = \frac{E}{A}(\exp(ik\phi_1) - \exp(ik\phi_2))/(ik), \tag{30}$$

where E is the midpoint value of $H_e - K_e$, these values given by Eqs. (13) and (16), and $A = \hat{e} \cdot (\hat{k} - \hat{r})$; $\hat{e}$ is the unit tangent vector of the line segment and $\hat{k}$ and $\hat{r}$ are the unit directional vectors of the incident wave and receiver direction, respectively. In the time domain this contribution can be expressed as

$$P_{\text{edge}}(t) = \frac{E}{A}[(\overline{S}(t-\Omega_1) - \overline{S}(t-\Omega_2)], \tag{31}$$

which has the same time-functional form as Eq. (28).

We do not use any sophisticated shadowing corrections to the Kirchhoff approximation in our method. We simply set facet contributions to zero if either $\hat{r} \cdot \hat{n} \leqslant 0$ or $-\hat{k} \cdot \hat{n} \leqslant 0$.

IV. NUMERICAL EXAMPLES

A. A rigid sphere

In this example we consider plane wave Ricker pulses centered at 10 and 40 kHz incident upon a rigid sphere of radius 0.25 m. For the wavelet $\overline{S}$ (the time integral of the source pulse) we use the integral of the Ricker wavelet, yielding a derivative of a Gaussian pulse. The receiver is located 10-m distance from the center of the sphere. For the sphere there are no edges, so this example only verifies the smooth Kirchhoff approximation to the solution. We compare the Kirchhoff solution to an analytic solution obtained from the standard harmonic series solution using Fourier synthesis to obtain the time-domain pulse. For the facet model, 41 points for the 10-kHz case and 101 points for the 40-kHz case are used to discretize the vertical angle ϕ,

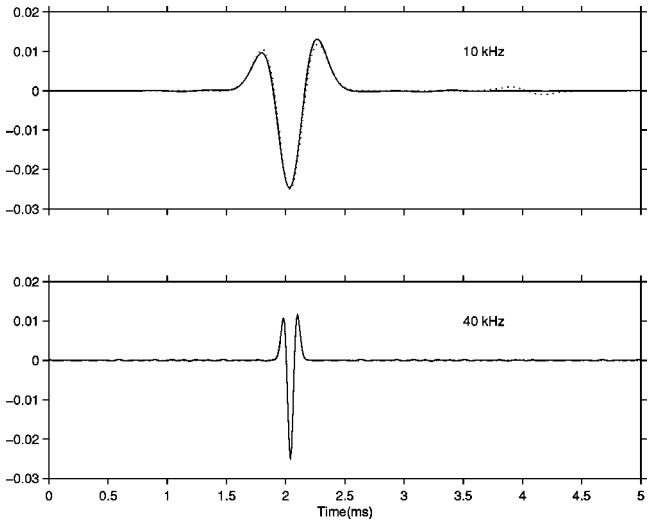


FIG. 2. Comparisons between analytic (solid) and Kirchhoff time series (dashed) for scattering from a rigid sphere for (a) 10 kHz and (b) 40 kHz.

$0 \leqslant \phi \leqslant \pi$ and azimuthal angle θ, $0 \leqslant \theta \leqslant 2\pi$. The wave is incident at an angle of 1° off the vertical [this was done to simply avoid a numerical (but removable) singularity in one of our expressions] and the receiver is placed in the monostatic backscatter direction. Good results were obtained in both cases using fewer discretization points, but slight numerical oscillations could be observed in the time series behind the specular reflection if, for example, half this number of points were used. As can be seen in Fig. 2 the agreement between the Kirchhoff (dotted line) and the analytic result is very good. The only differences between the two are that the Kirchhoff approximation predicts a small arrival corresponding to the hemispherical boundary between the illuminated and shadow faces of the sphere; this is stronger for the 10-kHz case than for the 40-kHz case. This arrival is not evident in the analytic solution.

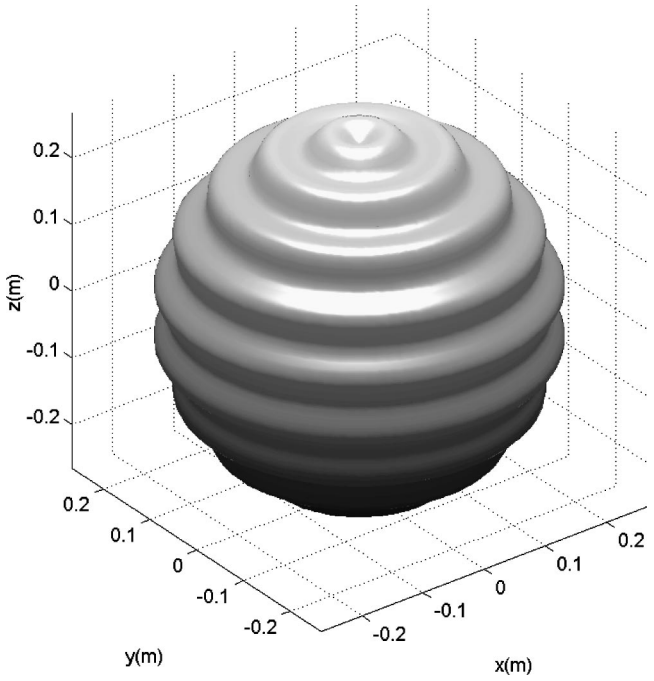


FIG. 3. Three-dimensional rendering of corrugated sphere.

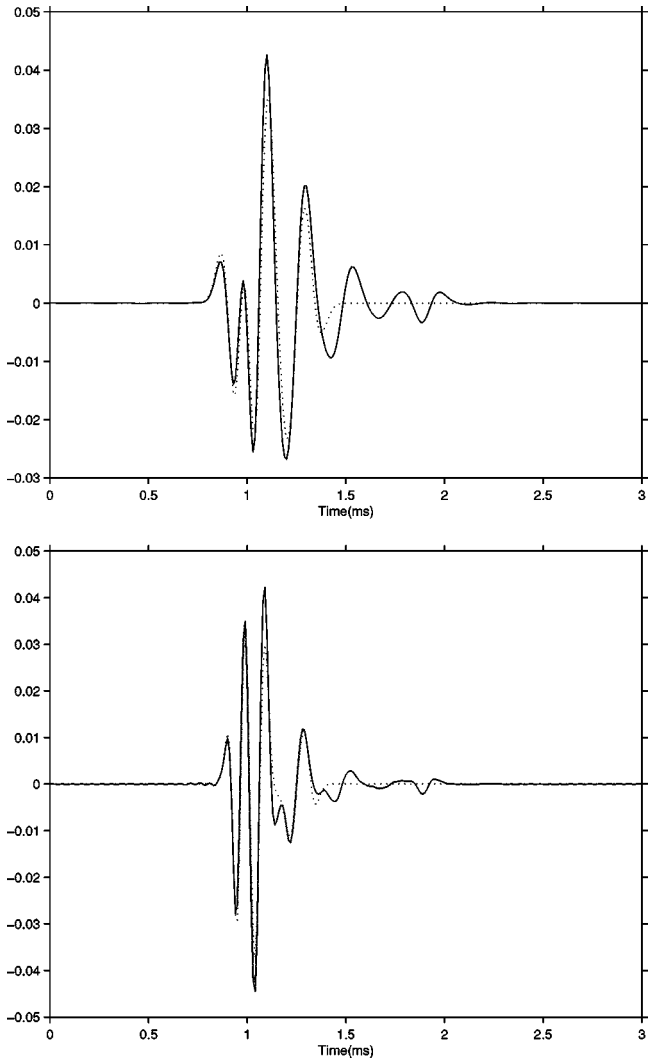


FIG. 4. Comparison between benchmark BIEM computation (solid) and Kirchhoff approximation (dashed) for $\beta = 5$ and $f = 5$ kHz and 8 kHz.

In order to show the flexibility of the method, we now compute the scattering from a corrugated sphere. This sphere has a radius which is defined as a function of the angle off vertical,

$$R(\phi) = 0.25 + 0.02 \sin(\beta\phi)^2. \tag{32}$$

A three-dimensional rendering of this sphere is shown in Fig. 3 for $\beta = 10$.

In Figs. 4 and 5 we show a comparison between the results of a 3D BIEM (for azimuthal symmetry) and the Kirchhoff approximation for the cases of 5- and 8-kHz Ricker pulses and $\beta = 5$ and 10 in Eq. (32). The details of the BIEM can be found in Ref. 12; it is a BIEM formulation for objects which possess azimuthal symmetry, in which case the full three-dimensional scattering problem can be broken down to a series of two-dimensional problems. If the source is placed along the axis of symmetry, then in fact only the scattering solution for the zeroth azimuthal order need be computed and the code can compute the scattered fields efficiently.

Here, we can see that the agreement between the BIEM and the Kirchhoff solutions is very good for the first portion

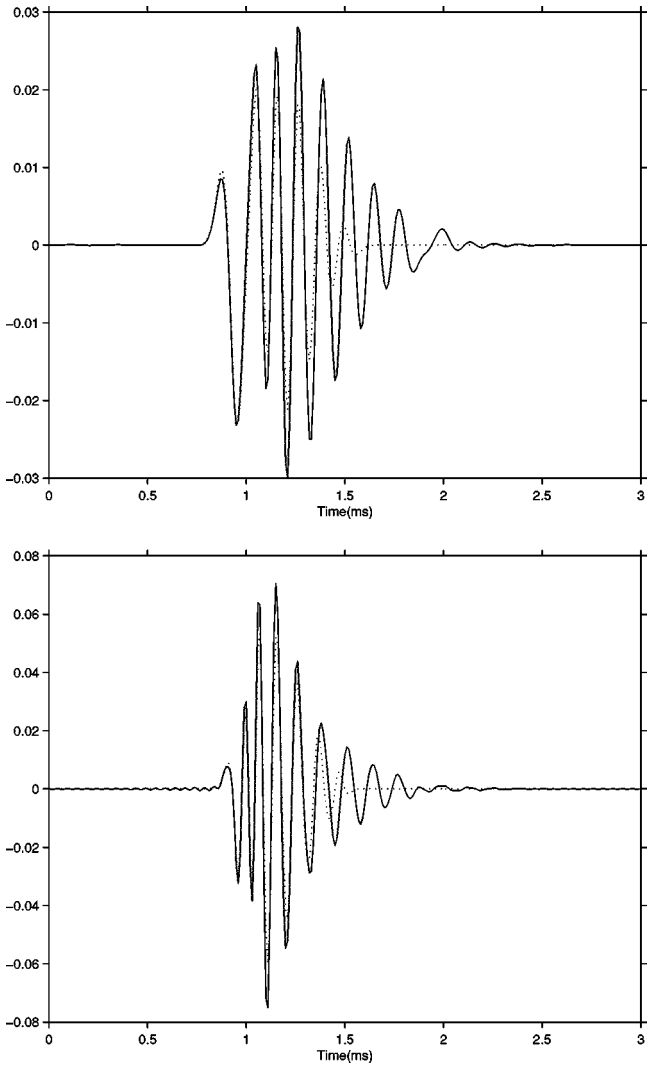


FIG. 5. Comparison between benchmark BIEM computation (solid) and Kirchhoff approximation (dashed) for $\beta=10$ and $f=5$ kHz and 8 kHz.

of the time series until such a time that the Kirchhoff solution is cut off due to the shadowing. It is somewhat surprising that the Kirchhoff approximation does so well as the wavelengths for the center frequency of the pulses (0.29 and 0.18 m, respectively) are larger than the wavelength of the roughnesses which are approximately 0.2 and 0.1 m for the two values of β. The agreement between the Kirchhoff and benchmark solutions is somewhat better for the higher frequency, as would be expected.

Unlike the BIEM code we are using, the Kirchhoff approximation is not limited to objects with azimuthal symmetry; we now also consider a rough sphere, whose functional form is given by

$$R(\phi,\theta)=0.25+0.02\sin(10\phi)^2\sin(5\theta). \tag{33}$$

A three-dimensional rendering of this surface is shown below in Fig. 6.

The resulting backscattered time series for the 40-kHz incident pulse is shown in Fig. 7(a) for the corrugated sphere of Fig. 3 and in Fig. 7(b) is the time series for the rough sphere of Fig. 6. In both cases the amplitude of the return is stronger than for the smooth sphere. The amplitude of the

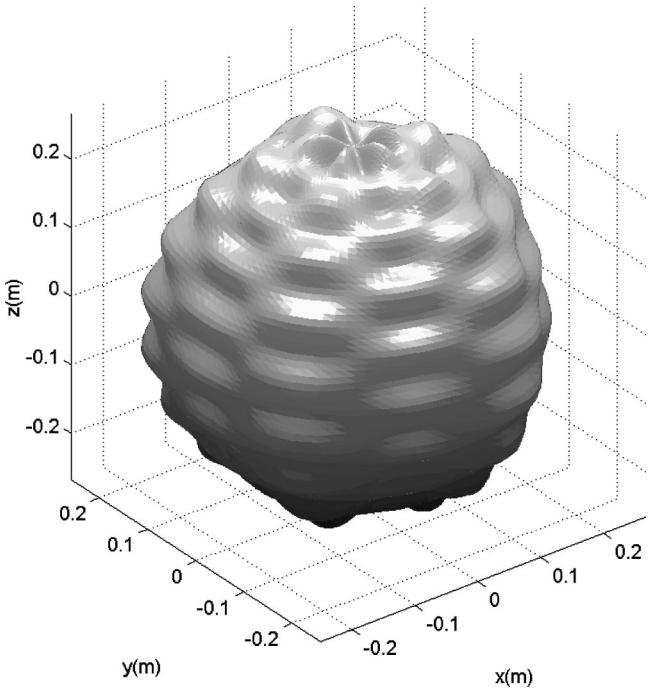


FIG. 6. Three-dimensional rendering of rough sphere described by Eq. (33).

series for the two-dimensional corrugation is somewhat less than that for the one-dimensional corrugation; this is not surprising as in this case the ''ridges'' of Fig. 7(a) are now broken up.

B. A rigid cylinder

We now consider the scattering of a plane wave from a rigid, finite cylinder; the radius of the cylinder is 0.25 m and the length is 2.0 m. A receiver is located 6 m from the center of the cylinder and the monostatic scattering angle is varied between $-90°$ (end-on incidence) to $+90°$. The angle of 0° denotes broadside incidence. The resulting time series (amplitude) as computed by the BIEM for an incident 5-kHz pulse are shown below as a function of the aspect angle in Fig. 8. That is, at each angle in Fig. 8 the backscattered time

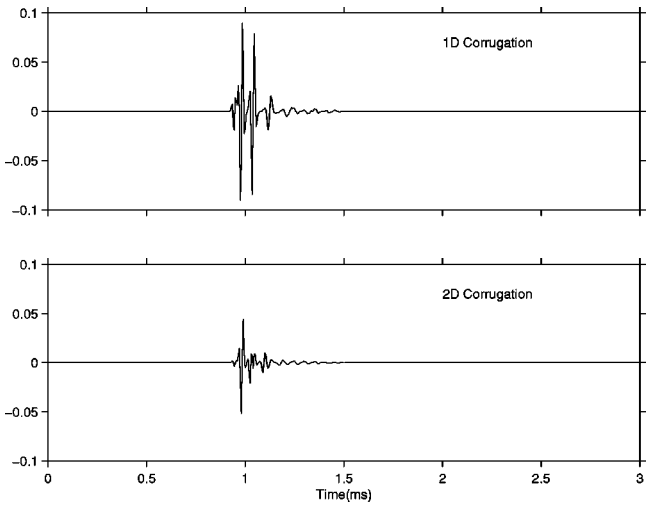


FIG. 7. Kirchhoff scattering of 40-kHz Ricker pulse from (a) rough sphere I and (b) rough sphere II.

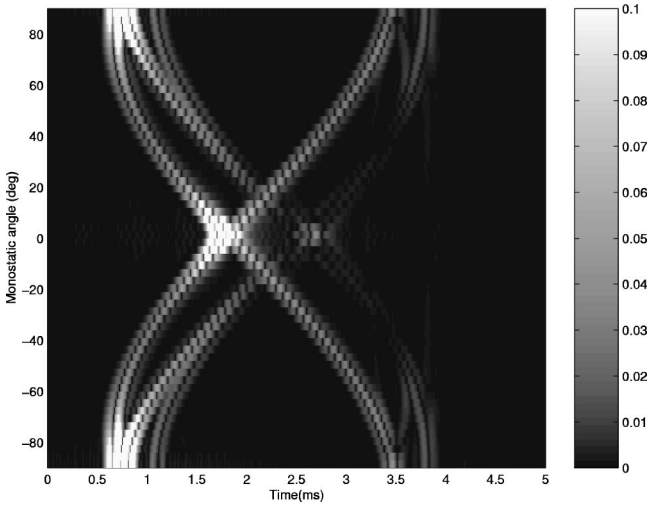


FIG. 8. Time-angle plot of scattering from rigid finite cylinder.

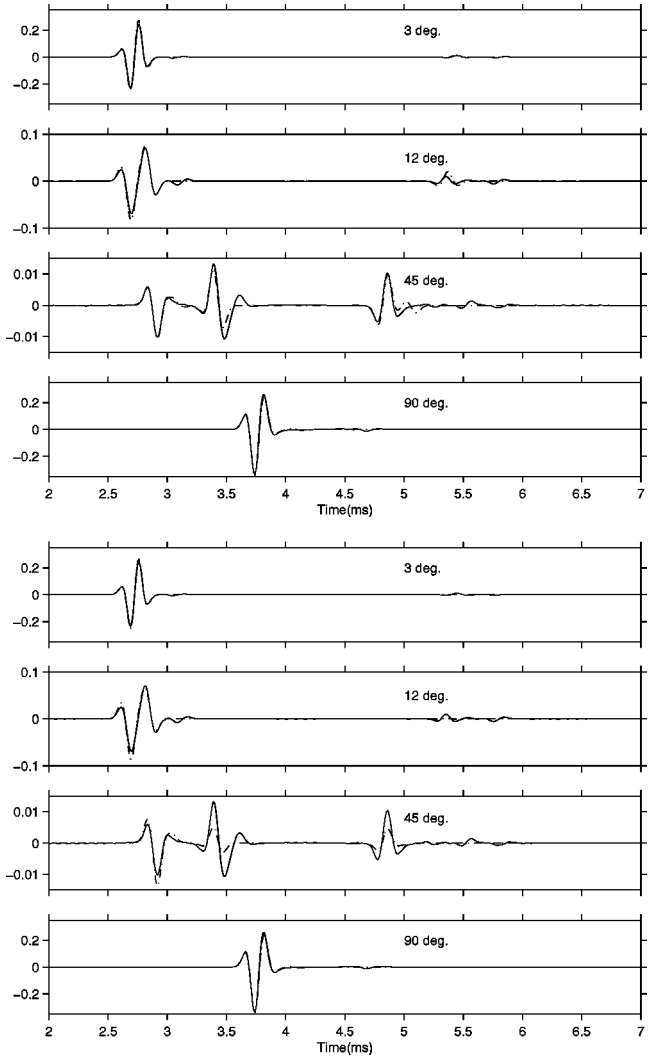


FIG. 9. Comparison of backscattered time series as computed using a BIEM (solid) for monostatic scattering angles of (a) 3°; (b) 12°; (c) 45°; (d) 90°; and for (a) top panel Kirchhoff/Diffraction method (dashed); (b) standard Kirchhoff method (dashed).

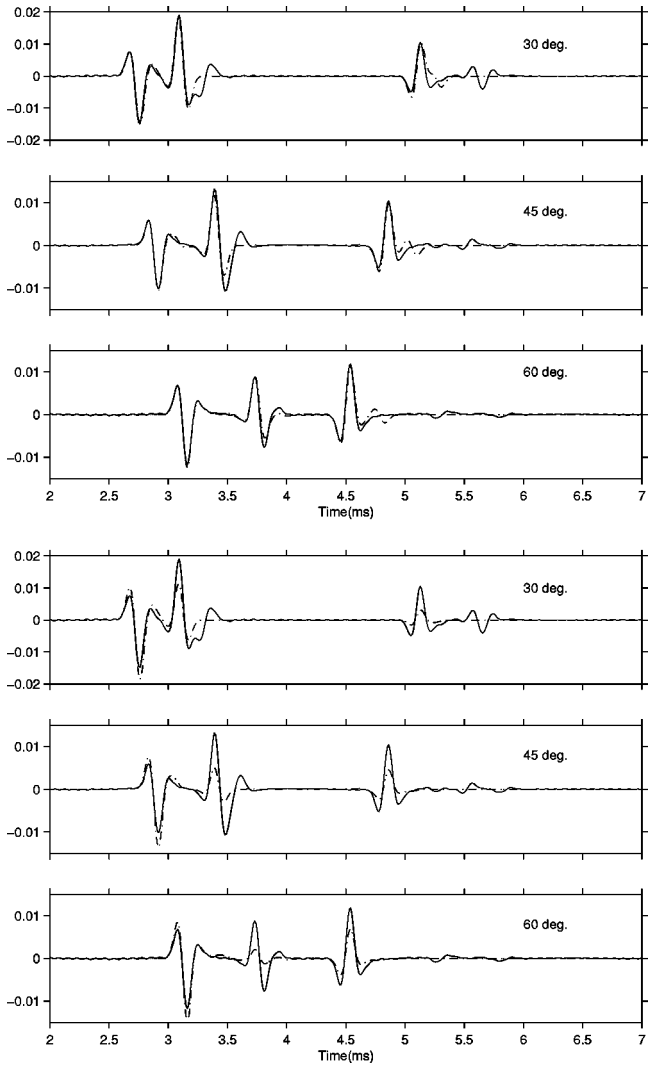


FIG. 10. Comparison of backscattered time series as computed using a BIEM (solid) for monostatic scattering angles of (a) 30°; (b) 45°; (c) 60°; and for (a) top panel Kirchhoff/diffraction method (dashed); (b) standard Kirchhoff method (dashed).

series (amplitude) is shown. The specular and diffraction arrivals are evident as bright curves. There are also weaker events which we believe are multiple diffraction events, and which are not predicted by single diffraction theory. For example, starting at end-on incidence a curve of arrivals behind the first two diffractions from the top face of the cylinder can be seen. This curve eventually coalesces with the second diffraction curve at about 40° off broadside. There are also other weak arrivals which can be seen after the third diffraction curve.

This is a more demanding BIEM computation than that for the rough sphere. In this example, we are now considering off-axis geometries and hence must compute the BIEM solution for several azimuthal orders as well as for the frequencies required for Fourier synthesis of the pulse.

In the top panel of Fig. 9 we show a comparison of the actual time series as computed by the BIEM method (solid line) and the hybrid Kirchhoff/diffraction method (dashed line) for −87° (or 3° off vertical), 12° off vertical, 45° off vertical and broadside (0°). As can be seen, the agreement between the two methods is excellent for all angles consid-

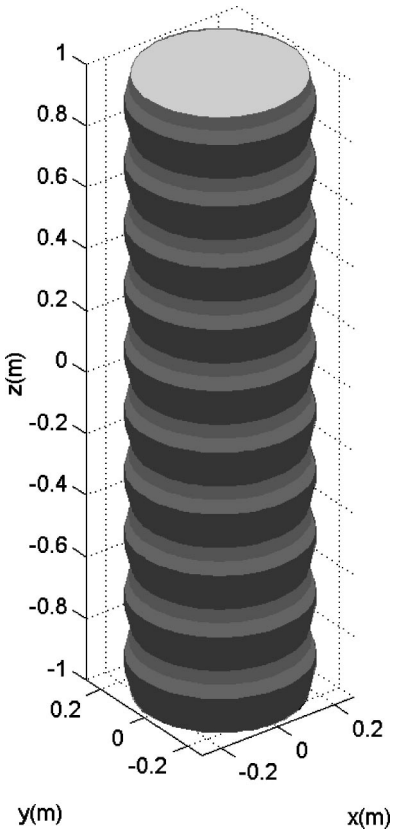


FIG. 11. Three-dimensional rendering of the corrugated cylinder.

ered. In the bottom panel of Fig. 9 we show the corresponding comparison between the BIEM and the time-domain pulses computed using the standard Kirchhoff approximation (i.e., we numerically use only our surface discretizations and ''turn off'' all edge contributions). We can see that, in fact, this approximation also does very well, except at 45°. It is actually where the diffractions dominate over specular scattering that we expect the standard Kirchhoff approximation to deteriorate. To illustrate this, we show in the top panel of Fig. 10 a comparison between the BIEM computations (solid) and the hybrid Kirchhoff/diffraction method for angles 30°, 45°, and 60° off vertical, and in the bottom panel of Fig. 10 we show the corresponding results using the standard Kirchhoff approach. Although the standard approach seems to model the first diffraction well, the second and third diffractions are not modeled as well as by using the hybrid method. Its also interesting to note for the 30° aspect the relatively strong multiple diffraction event arriving after the first three. This is not modeled by either method.

We now consider the same cylinder (see Fig. 11) but with a sinusoidally varying radius down its side

$$r(z) = 0.25 + 0.02 \sin^2(5(z-1)\pi). \tag{34}$$

The benchmark solution is once again computed using the BIEM. The time-angle plot as in Fig. 8 is shown below in Fig. 12. The diffractions are still evident in the time series plots for aspects within about 20° of endfire. However, the scattering from the roughness is also very clear and begins to dominate for aspects past 20° off endfire until within about 20° from broadside (0°), at which point this scattered energy

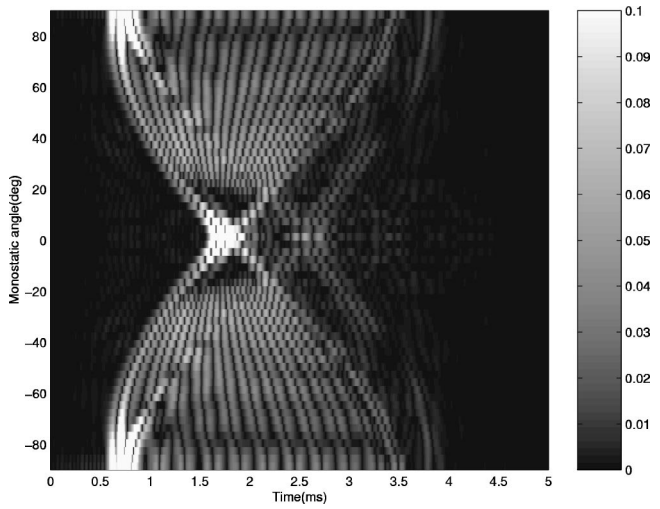


FIG. 12. Time-angle plot of scattering from corrugated, rigid finite cylinder.

is significantly diminished. In Fig. 13 individual time series comparisons (the BIEM solid and the Kirchhoff/diffraction method dashed) are made for selected aspects. As can be seen, the agreement is excellent.

V. SUMMARY

We have implemented a Kirchhoff/diffraction model in the time domain, using triangular facets to describe an object's surface. This method showed excellent agreement with analytic and BIEM methods for smooth and rough spheres and cylinders. Even for situations where the wavelength of the roughness was smaller than the acoustic wavelength, excellent agreement was found. The Kirchhoff/diffraction method described in this paper is applicable to arbitrarily shaped objects where the edges are specified. Although 40 kHz was the largest center frequency considered in this paper, the method can model scattering up to 200 kHz or so in a reasonable amount of time.

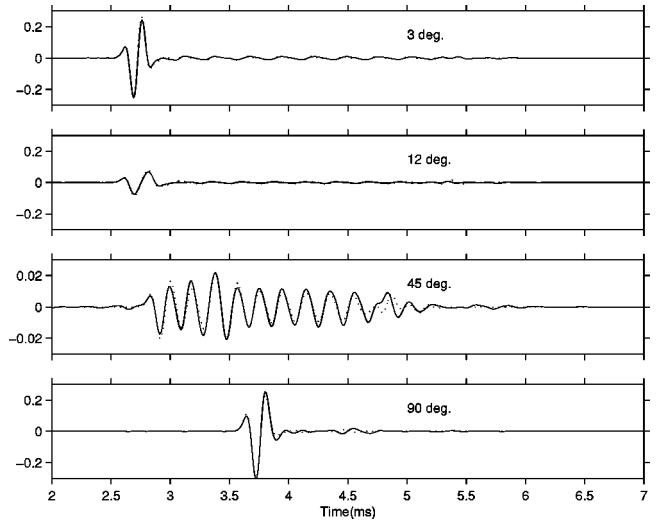


FIG. 13. Comparison of backscattered time series as computed using a BIEM (solid) and the Kirchhoff/diffraction method (dashed) for monostatic scattering angles of (a) 3°; (b) 12°; (c) 45°; and (d) 90°.

[1] L. M. Brekhovskikh and Yu. P. Lysanov, *Fundamentals of Ocean Acoustics*, 2nd ed. (Springer, Berlin, 1991).
[2] E. Pouliquen, O. Bergem, and N. G. Pace, ''Time-evolution modeling of seafloor scatter. I. Concept,'' J. Acoust. Soc. Am. **105**, 3136–3141 (1999).
[3] O. Bergem, E. Pouliquen, G. Canepa, and N. G. Pace, ''Time-evolution modeling of seafloor scatter. II. Numerical and experimental evaluation,'' J. Acoust. Soc. Am. **105**, 3142–3150 (1999).
[4] G. C. Gaunaurd, ''Sonar cross sections of bodies partially insonified by finite sound beams,'' IEEE J. Ocean Eng. **OE-10**, 213–230 (1985).
[5] D. Brill and G. C. Gaunaurd, ''Approximate descriptions of the sound fields scattered by insonified, submerged, ribbed, flat-ended cylindrical structures,'' J. Acoust. Soc. Am. **93**, 71–79 (1993).
[6] J. B. Keller, ''Geometrical theory of diffraction,'' J. Opt. Soc. Am. **52**, 116–130 (1962).
[7] J. B. Keller, ''Asymptotic solution of some diffraction problems,'' Commun. Pure Appl. Math. **9**, 207–265 (1956).
[8] P. Ya. Ufimtsev, ''Theory of acoustical edge waves,'' J. Acoust. Soc. Am. **86**, 463–474 (1989).
[9] D. S. Ahluwalia, ''Uniform asymptotic theory of diffraction by the edge of a three-dimensional body,'' SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **18**, 287–301 (1970).
[10] M. Tran Van Nhieu, ''Diffraction by the edge of a three-dimensional object,'' J. Acoust. Soc. Am. **99**, 79–87 (1996).
[11] G. S. Sammelmann, J. T. Christoff, and J. D. Lathrop, ''Synthetic images of proud targets,'' Proceedings of OCEANS'94, held in Brest, France, 13–16 Sept. 1994, Vol. 2, pp. 266–271.
[12] J. A. Fawcett, ''Finite element modeling of three-dimensional scattering from azimuthally symmetric elastic shells,'' SACLANTCEN Report SR-273, February 1998.

Use of nonsingular boundary integral formulation for reducing errors due to near-field measurements in the boundary element method based near-field acoustic holography

Sung-Chon Kang and Jeong-Guon Ih[a)]
Center for Noise and Vibration Control, Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology, Science Town, Taejon 305-701, Korea

(Received 26 July 1999; accepted for publication 22 December 2000)

In the conformal near-field acoustic holography (NAH) using the boundary element method (BEM), the transfer matrix relating the vibro-acoustic properties of source and field depends solely on the geometrical condition of the problem. This kind of NAH is known to be very powerful in dealing with the sources having irregular shaped boundaries. When the vibro-acoustic source field is reconstructed by using this conformal NAH, one tends to position the sensors as close as possible to the source surface in order to get rich information on the nonpropagating wave components. The conventional acoustic BEM based on the Kirchhoff–Helmholtz integral equation has the singularity problem in the close near field of the source surface. This problem stems from the singular kernel of the Green function of the boundary integral equation (BIE) and the singularity can influence the reconstruction accuracy greatly. In this paper, the nonsingular BIE is introduced to the NAH calculation and the holographic BIE is reformulated. The effectiveness of nonsingular BEM has been investigated for the reduction of reconstruction error. Through interior and exterior examples, it is shown that the resolution of predicted field pressure could be improved in the close near field by employing the nonsingular BIE. Because the BEM-based NAH inevitably requires the field pressure measured in the close proximity to the source surface, the present approach is recommended for improving the resolution of the reconstructed source field. [DOI: 10.1121/1.1350401]

PACS numbers: 43.20.Ye, 43.35.Sx, 43.40.Yq, 43.50.Yw [ANN]

I. INTRODUCTION

The near-field acoustic holography (NAH) is one of the indirect methods for identifying the vibro-acoustic properties of the vibrating sound source. This technique is equivalent to solving an inverse or backward problem recovering the vibro-acoustic properties at the source plane from measured pressure data on the hologram plane. Not only the source data are vital for the effective noise control, but the sound propagation field calculated forward from the restored source data is important as well.

Two major techniques have been suggested in realizing the NAH. In one method, discretely and uniformly measured field pressures on a regular hologram plane are decomposed into the space-wave number domain by the spatial Fourier transform. After compensating the pressure decay during propagation, the pressure on a target plane is reconstructed by the inverse spatial Fourier transform.[1,2] In another method, the sound radiation and transmission between the vibrating source surface and the measurement field are modeled by the vibro-acoustic transfer matrix using the boundary element method (BEM).[3–6] The distribution of surface acoustic parameters of the source can be reconstructed by multiplying the inverse of vibro-acoustic transfer matrix and the measured field pressure vector. Any shape of the near-field plane including the conformal one can be used as the hologram plane and this conformal NAH can deal with the complex-shaped sources that cannot be described with separable coordinates. Compared with this latter method, the former method should be involved only with regular shapes of source surface and hologram plane, otherwise a hypothetical regular plane near the actual source should be considered a target source plane for reconstruction. Due to this advantage, the ''BEM-based NAH'' has been the subject of many previous studies in spite of the fact that it includes difficulties arising from the BEM modeling of source and field, and from dealing with the ill-conditioned system. Successful results were reported by applying this method to several interior or exterior acoustic problems, either experimentally or numerically.[6–12]

When a source field is being restored by this BEM-based NAH, one tends to position sensors as close to the source surface as possible in order to get profound information on the field. When the field points are placed very close to the source surface, it is generally thought that the singularity of transfer matrix will be relatively improved and the ratio of measured field pressure to background noise can be increased. In addition, the measured near-field pressures include the rich information on the nonpropagating wave components. All of these factors seem to be beneficial to the precision of reconstruction. However, the conventional acoustic BEM based on the Kirchhoff–Helmholtz integral equation suffers from the singularity problem in the near field of the source surface. Computational error increases very rapidly approaching from the far field to the source surface. The problem is due to the singular kernel in the

[a)]Electronic mail: ihih@sorak.kaist.ac.kr

fundamental solution of the boundary integral equation (BIE) and it can influence the reconstruction accuracy. In the three-dimensional Kirchhoff–Helmholtz BIE, the monopole and dipole terms expressed by the Green function and its derivative are proportional to $1/R$ or $1/R^2$, respectively, where R is the distance between source and field point. When the field points are located in the very near field, R becomes very small and thus calculation of Green functions diverges in the numerical integration. This can invoke a dilemma that the field pressure should be measured in the near field for the precise reconstruction, while the prediction accuracy is deteriorated in the near field due to the singularity of the Green function and its derivative.

A nonsingular boundary integral equation was formulated to solve this singularity problem of BEM in the near field.[13] In this method, the level of singularity can be lowered by eliminating the propagating plane wave term from the integral identity. All singularities included in the conventional acoustic BIE can be changed to the weak ones or can be entirely removed by utilizing this formulation. As a result, the near-field pressure can be determined precisely.

In this paper, the BIE for holography is reformulated to remove the near-field singularity. Conventional and nonsingular methods are applied to the interior and the exterior examples and the reconstructed results are compared with each other. It is observed that the nonsingular formulation can provide the accurate vibro-acoustic transfer matrix between source and field and thus improve the resolution of the reconstructed source field.

II. THEORETICAL FORMULATION

A. Brief overview of the nonsingular acoustic BIE (Ref. 13)

For a domain V enclosed by the boundary S_o filled with the isothermal, homogeneous, inviscid, compressible, and stationary fluid medium, the well-known Kirchhoff–Helmholtz integral equation is expressed as follows:

$$c(\mathbf{r})p(\mathbf{r})=\int_{S_o}\left[G(\mathbf{r},\mathbf{r}_o)\frac{\partial p(\mathbf{r}_o)}{\partial \mathbf{n}(\mathbf{r}_o)}-\frac{\partial G(\mathbf{r},\mathbf{r}_o)}{\partial \mathbf{n}(\mathbf{r}_o)}p(\mathbf{r}_o)\right]dS(\mathbf{r}_o). \tag{1}$$

Here, $p(\mathbf{r}_o)$ means the surface pressure, $G(\mathbf{r},\mathbf{r}_o)=\exp(-jkR)/R$ is the three-dimensional free-space Green function, $R=|\mathbf{r}-\mathbf{r}_o|$ denotes the distance between field point $\mathbf{r}$ and surface point $\mathbf{r}_o$, $\partial/\partial\mathbf{n}$ means the outward normal derivative on the surface, and the coefficient $c(\mathbf{r})$ represents the solid angle at $\mathbf{r}$.

In the right-hand side of Eq. (1), G and $\partial G/\partial\mathbf{n}$ are inversely proportional to R and R^2, respectively. If a field point is located very close to the boundary surface, R and R^2 become very small. In order to implement Eq. (1) in BEM, each term in Eq. (1) should be integrated numerically. For a function f represented by local coordinates ξ and η, one can employ the discrete Gauss integration technique as

$$\int\int f(\xi,\eta)d\xi\,d\eta=\sum_{i=1}^{q} w_i f(\xi_i,\eta_i), \tag{2}$$

where q is the number of integration points and w_i is the weighting factor for each point. Therefore, the singularities of $1/R$ and $1/R^2$ remain in the numerical calculation and the integral calculations diverge when the field point is located very close to the surface.

In order to remove the singularity existing in the conventional BIE, the propagating plane wave can be deducted from the integral identity of Eq. (1). Introducing an additional propagating plane wave and carrying out several steps of calculations, the nonsingular BIE can be derived as follows:

$$\begin{aligned}
&-j\int_{S_o} G(\mathbf{r},\mathbf{r}_o)[\rho_o\omega V_n(\mathbf{r}_o)+k\mathbf{n}(\mathbf{r}_o)\cdot\{\mathbf{h}(\mathbf{y})\hat{p}(\mathbf{r}_o;\mathbf{y},\mathbf{h}(\mathbf{y}),p(y))-\mathbf{h}(\mathbf{y})\hat{p}(\mathbf{r}_o;\mathbf{y},\mathbf{h}(\mathbf{y}),q)+\mathbf{n}(\mathbf{y})\hat{p}(\mathbf{r}_o;\mathbf{y},\mathbf{n}(\mathbf{y}),q)\}]dS(\mathbf{r}_o)\\
&\quad-\int_{S_o}\frac{\partial G(\mathbf{r},\mathbf{r}_o)}{\partial\mathbf{n}(\mathbf{r}_o)}[p(\mathbf{r}_o)-\hat{p}(\mathbf{r}_o;\mathbf{y},\mathbf{h}(\mathbf{y}),p(\mathbf{y}))-\hat{p}(\mathbf{r}_o;\mathbf{y},\mathbf{h}(\mathbf{y}),q)+\hat{p}(\mathbf{r}_o;\mathbf{y},\mathbf{n}(\mathbf{y}),q)]dS(\mathbf{r}_o)\\
&=\begin{cases}0, & \text{for interior problem } (\mathbf{r}\in S_o,\mathbf{r}\notin V),\\ p(\mathbf{r})-\hat{p}(\mathbf{r};\mathbf{y},\mathbf{h}(\mathbf{y}),p(\mathbf{y}))-\hat{p}(\mathbf{r};\mathbf{y},\mathbf{h}(\mathbf{y}),q)+\hat{p}(\mathbf{r};\mathbf{y},\mathbf{n}(\mathbf{y}),q), & \text{for interior problem } (\mathbf{r}\notin S_o,\mathbf{r}\in V),\\ p(\mathbf{r}), & \text{for exterior problem } (\mathbf{r}\in S_o,\mathbf{r}\notin V,\ \text{or}\ \mathbf{r}\notin S_o,\mathbf{r}\in V).\end{cases}
\end{aligned} \tag{3}$$

Here, $\mathbf{h}(\mathbf{y})$ is the unit tangential vector at $\mathbf{y}$, $\mathbf{n}(\mathbf{x})$ is the unit normal vector at $\mathbf{x}$, $p(\mathbf{y})$ is the sound pressure at $\mathbf{y}$, $q=-\rho_o c V_n(\mathbf{y})$, ρ_0 is the density of medium, c is the speed of sound, and $V_n(\mathbf{y})$ is the normal velocity at $\mathbf{y}$. When $\mathbf{r}$ represents a point on the surface, $\mathbf{y}$ coincides with $\mathbf{r}$, whereas $\mathbf{y}$ represents the closest point to $\mathbf{r}$ on the surface if $\mathbf{r}$ denotes a point in the domain. The propagating plane wave that satisfies both the Helmholtz equation and Kirchhoff–Helmholtz integral equation can be expressed as

$$\hat{p}(\mathbf{x};\mathbf{y},\mathbf{a},\overline{p})=\overline{p}\exp[jk(\mathbf{x}-\mathbf{y})\cdot\mathbf{a}], \tag{4}$$

where $\overline{p}$ is the amplitude of plane wave, $\mathbf{y}$ denotes an arbitrary reference point, $\mathbf{a}$ is the direction vector of the plane wave at $\mathbf{y}$. Physically, Eq. (4) implies the field pressure at

point **x** when a plane wave is propagating with an amplitude of $|\overrightarrow{p}|$ in a direction **a** that is determined at point **y**. Note that a uniform potential field could be combined with the governing equation for solving the singularity problem, but it would involve inefficient calculations. In spite of the fact that Eq. (3) can be used at any position of field point **r**, it is applicable only to the smooth surface where the surface normal vector can be defined uniquely. Although Eq. (3) looks very complicated in comparison with Eq. (1), the same physical meanings with the case of conventional BIE are conserved in the involved terms. The first integrand is connected with the effect of monopole and the second integrand means the effect of dipoles. The bracketed term in each integrand defines the strength of each pole. In Eq. (3), all singularities are eliminated and thus the pressure can be estimated accurately even in the very near field.

B. Reformulation of the holography equation

A discrete form of Eq. (3) can be used in realizing BEM and its matrix-vector form can be written as

$$\{p\}_f=[D]_f\{p\}_s+[M]_f\{v\}_s \quad \text{in the domain,} \tag{5}$$

$$[D]_s\{p\}_s=[M]_s\{v\}_s \quad \text{on the boundary.} \tag{6}$$

Here, $\{p\}_s$, $\{v\}_s$ are the pressure and velocity vector on the surface, respectively, $\{p\}_f$ denotes the field pressure vector in the domain, $[D]_s$, $[M]_s$ mean the dipole and monopole matrices on the surface, and $[D]_f$, $[M]_f$ are those corresponding to field pressures, respectively. From Eq. (5), the following field pressure can be described only by the surface velocity provided $[D]_s^{-1}$ exists:

$$\{p\}_f=([M]_f+[D]_f[D]_s^{-1}[M]_s)\{v\}_s\equiv[G]\{v\}_s. \tag{7}$$

Here, $[G]$ is the vibro-acoustic transfer matrix correlating the surface normal velocity and the field pressure that contains the geometric information of the system as well. If the field pressure at m points is known, the surface velocity at $n(\leqslant m)$ nodes can be uniquely determined. This can be accomplished by utilizing the overdetermined least-squared solutions approach and the singular value decomposition (SVD) technique.[14] The SVD of the transfer matrix $[G]$ is given by

$$[G]=[U][\Lambda][W]^H, \tag{8}$$

where

$$[\Lambda]=\mathrm{diag}(\lambda_1,\lambda_2,\ldots,\lambda_n),$$
$$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n\geqslant 0,\ \{u_i\}^H\{u_j\}=\delta_{ij},$$
$$\{w_i\}^H\{w_j\}=\delta_{ij}. \tag{9}$$

Here, δ_{ij} is the Kronecker delta, the elements of diagonal matrix $[\Lambda]$ are singular values λ_i, and $[U]$, $[W]$ mean the vectors, each of which has orthonormal columns. Physically, $\{u_i\}$ and $\{w_i\}$ mean the wave vectors that decompose the distribution of field pressure on the hologram plane and that of surface velocity on the source plane at a selected frequency. By using Eq. (8), one can derive the following inverse equation:

$$\{v\}_s=[G]^+\{p\}_f=([G]^H[G])^{-1}[G]^H\{p\}_f$$
$$=[W][\Lambda]^{-1}[U]^H\{p\}_f. \tag{10}$$

Here, the superscripts ''+'' and ''H'' signify the pseudo-inverse and the Hermitian operator, respectively. If the transfer matrix $[G]$ is generated by the nonsingular BEM and the field pressure $\{p\}_f$ is measured, the surface velocity $\{v\}_s$ can be determined from Eq. (10).

The transfer matrix $[G]$ is ill-conditioned in general and, therefore, the condition number $(=\lambda_{\max}/\lambda_{\min})$ of this matrix is much larger than unity. In the source reconstruction that involves the inversion of $[\Lambda]$, the high-order components of singular values affect the reconstructed result dominantly and they are correlated with the nonpropagating wave components. If the measured field pressure is contaminated with the noise, the noise can be amplified greatly by the high-order component λ_i and the reconstructed result will be distorted from the actual source field. The problem can be partially sorted out by the regularization of a singular matrix. In the regularization process, the wave-vector filter can be designed optimally by taking advantage of the trade-off relation between bias error and variance error.[7,15] By performing the proper wave-vector filtering, $[\Lambda]^{-1}$ can be regularized as

$$[\Lambda_F]^{-1}=[F][\Lambda]^{-1}$$
$$=\mathrm{diag}(f_1/\lambda_1,f_2/\lambda_2,f_3/\lambda_3,\ldots f_n/\lambda_n), \tag{11}$$

where the diagonal elements of $[F]$ are optimally designed wave-vector filter coefficients. In this study, the iterative regularization method[15] is employed in the calculation.

III. NUMERICAL SIMULATIONS

A. Test of nonsingular BEM for improvement of near-field accuracy

As mentioned before, the nonsingular BEM is the basis of the present study scope: the removal of sensor proximity error in the NAH. In order to test the performance of nonsingular BEM, two simple acoustic systems that permit analytic solutions are dealt with here and they correspond to exterior and interior problems, respectively. All computational details in realizing nonsingular BEM followed those mentioned in Ref. 13.

When the sphere with a radius of a is pulsating with a harmonic surface velocity of $Ue^{j\omega t}$ in the infinite homogeneous medium, the field pressure at distance r from the origin is given by[16]

$$p(r)=\frac{-\rho_0 cU}{r}\frac{jka^2}{(1-jka)}\exp\{jk(r-a)\}. \tag{12}$$

For the BEM calculation, a sphere having a radius of 50 mm was modeled with 98 nodes and 48 quadratic triangular elements. The characteristic length (L_c) of the model was 52.1 mm. Figure 1 shows the predicted error ratio of pressure amplitude obtained by BEM compared to the theoretical value in the radial field. The error ratio is presented as a function of the distance ratio, $(r-a)/L_c$. It is noted that the calculated result of nonsingular BEM (hereafter, NBEM) agrees very well with the theoretical value even in close

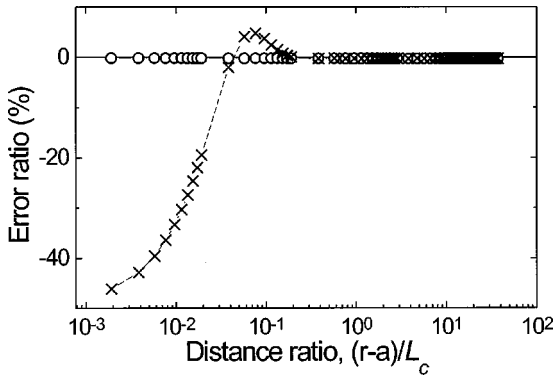


FIG. 1. Predicted error ratio of field pressure amplitude for a pulsating sphere model at $ka=1$: —○—, NBEM; −−×−−, CBEM.

proximity to the surface. However, the result of conventional BEM (hereafter, CBEM) shows large errors in the locations where the distance between the field point and the nearest surface is smaller than about $0.2L_c$. This large amount of error obviously stems from the singularity of the Green function and its derivative in the near field. The estimation error of field pressure is shown as a function of frequency in Fig. 2. For this purpose, a field point located at $0.0575L_c$ from the surface was chosen. For the NBEM, the change of error ratio is less than 1% when the wavelength is longer than $4L_c$. This means that at least eight nodes are required to represent a wavelength in order to expect the accurate results with the error ratio less than 1%. The result of NBEM shows small bias error in the low-frequency range. This bias error is due to the modeling error of BEM that is associated with the difference in surface area or volume velocity from the ideal pulsating sphere and it can be reduced by using the effective radius concept.

A tube-shaped parallelepiped of $300(w)\times300(h)\times1700(l)$ mm in size was considered a test example of the interior problem. The BEM model was made up of 468 linear triangular elements and 236 nodes. The characteristic length of this model was $L_c=141.2$ mm. It was assumed that an end surface at $z=0$ was vibrating as a rigid piston with a harmonic velocity of $Ue^{j\omega t}$ and all other walls were rigid. The theoretical field pressure from the one-dimensional plane wave theory is given by

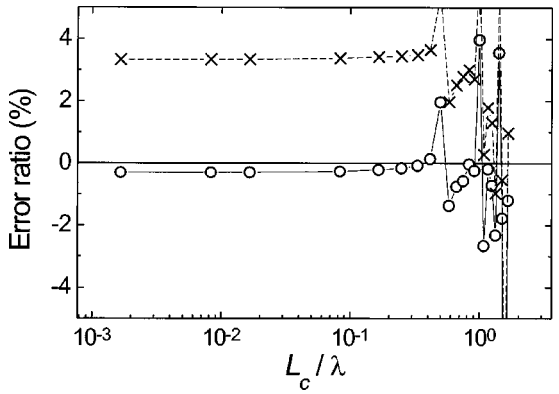


FIG. 2. Predicted error ratio spectrum of field pressure amplitude for a pulsating sphere model at $r=a+0.0575L_c$: —○—, NBEM; −−×−−, CBEM.

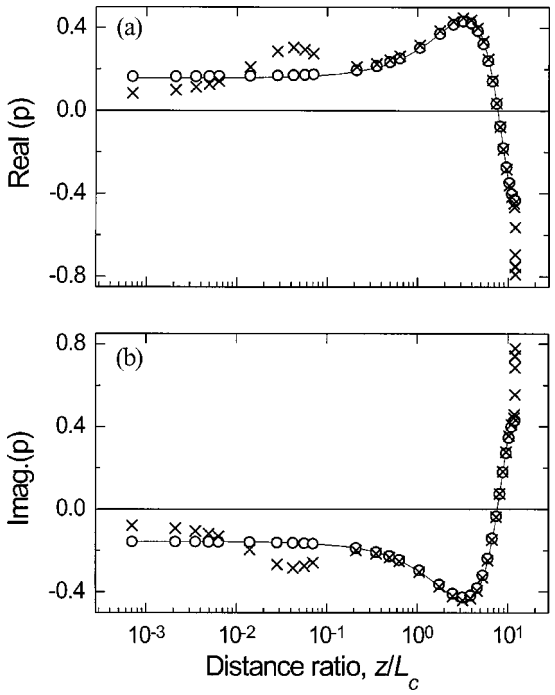


FIG. 3. Predicted field pressure amplitude on the longitudinal axis of the parallelepiped box interior due to a vibrating end face at 140 Hz: ———, exact analysis; ○, NBEM; ×, CBEM.

$$p(z)=\frac{-j\rho_0 cU}{\sin(kl)}\cos[k(l-z)], \tag{13}$$

where l is the length of the parallelepiped and z denotes the longitudinal coordinate. Figure 3 shows the calculated field pressure on the longitudinal axis when $U=(1+j1)e-3$ m/s at 140 Hz. Similar to the previous exterior example, the field pressure predicted by CBEM has large errors when $z<0.2L_c$, whereas NBEM result coincides with the exact theory. In Fig. 4, the predicted pressure response spectrum is shown at $z=0.0425L_c$ on the longitudinal axis. The reso-

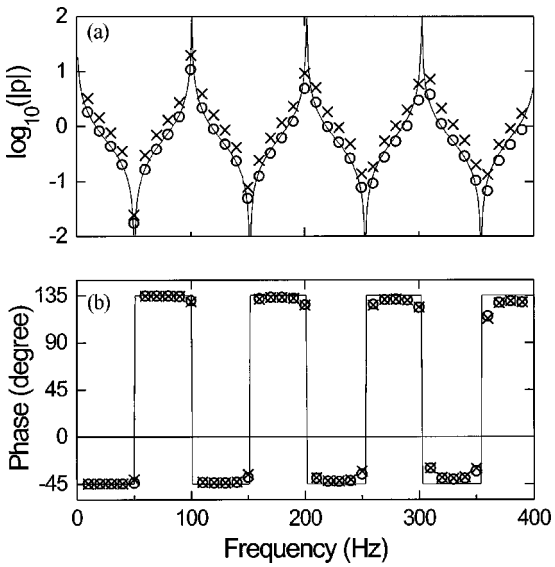


FIG. 4. Predicted pressure response spectrum at $z=0.0425L_c$ in the parallelepiped box interior due to a vibrating end face: ———, exact analysis; ○, NBEM; ×, CBEM.

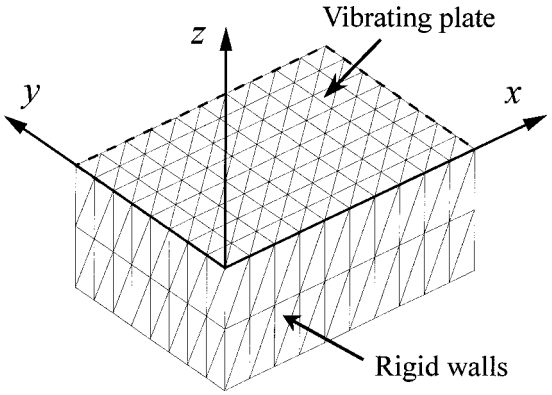


FIG. 5. BEM model of the parallelepiped box (700×500×320 mm) of which the top plate vibrates and radiates sound. Other side walls and bottom plate are rigid.

nance frequencies of longitudinal mode are determined by $kl=m\pi$ (m=integer), i.e., $f_m=100.9,201.8,302.7,...$, in Hz. It is observed that the NBEM result is far better than the CBEM result for all frequency ranges and the solution agrees reasonably with the analytic value, especially when $\lambda \geqslant 6L_c - 8L_c$. At resonance frequencies, the amplitude of theoretical value is infinite and it becomes zero at antiresonance frequencies. Consequently, the prediction accuracy is poor at these frequencies.

B. NAH based on the nonsingular BEM

1. Exterior problem

A parallelepiped radiator with the dimension of 500(w)×320(h)×700(l) mm was chosen as an application example of NAH based on nonsingular BEM. The boundary element model consisted of 234 nodes and 464 linear triangular elements as depicted in Fig. 5. The characteristic length of this model was $L_c=93.8$ mm. It was assumed that only the top plane made of 1-mm-thick steel plate would vibrate and other faces were rigid.

When the top plate is vibrating with (5,1) mode at 163 Hz, the pressure distribution on the vertical field plane ($y=250$ mm) above the box can be calculated as in Fig. 6. One can find that the field pressure estimated by CBEM is highly distorted in the close neighborhood of the source surface where the distance from the surface is less than about 20 mm, *viz.*, $0.213L_c$. In Fig. 7, the predicted error ratio of field pressures calculated by CBEM and NBEM is shown at 163 Hz and the calculation is done for the evenly spaced 609 field points varying the distance z from the source surface. Statistical curves including the error mean (e_{mean}) and the standard deviation (σ_p) are illustrated and the shaded area corresponds to 95% possibility interval of the prediction error. From this figure, one can clearly find that the prediction error in the near-field pressure by CBEM increases very much when the distance of field point to source surface becomes less than about $0.2L_c$. One may recall that the same result was already found in the pulsating sphere model. Here, the prediction inaccuracy reveals that the vibro-acoustic transfer matrix in the close near field does not correctly represent the relation between the surface velocity and the field pressure. The CBEM overestimates or underestimates the field pressure at certain points in the close near field and these points correspond to Gauss integration points. In Fig. 8, the pressure field at $z=1$ mm plane is compared with the surface pressure field and the foregoing remarks can be confirmed easily. The field pressure pattern calculated by NBEM is very similar to the surface pressure distribution, whereas the pattern by CBEM is quite different from the surface pressure pattern, especially near Gauss integration points.

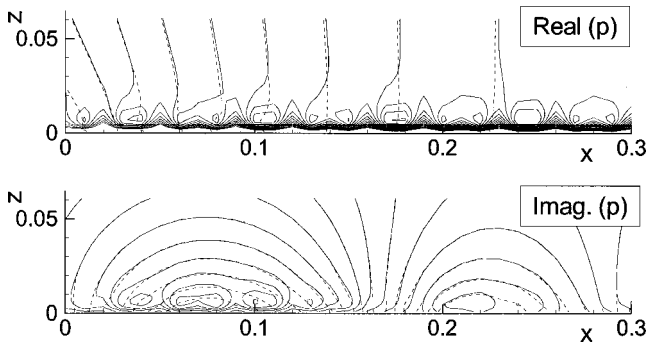


FIG. 6. Distribution of predicted field pressure on the vertical plane at $y=250$ mm (163 Hz):---, NBEM; ———, CBEM.

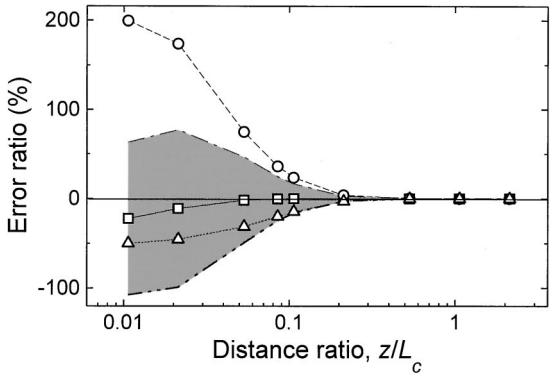


FIG. 7. Predicted error ratio of the field pressure calculated by CBEM and NBEM at 163 Hz, (5,1) mode. The shaded area shows 95% confidence interval of the prediction error with varying z for 609 field points: —□—, mean value; ––○––, maximum; ---△---, minimum; —·—, (mean$+2\sigma_p$); —··—, (mean$-2\sigma_p$).

The holographic reconstruction was carried out for (5,1) vibrating mode of top plate at 163 Hz. The field pressure was estimated for the evenly distributed 117 points in a 50×50-mm pattern and the distance of these points from the surface was within the range of 1–200 mm or, equivalently, $0.0107L_c-2.13L_c$. It was assumed that the measured field pressure was contaminated by the unbiased Gaussian random noise with a noise variance ($=\Sigma|n_i|^2$; n_i=noise included in field pressure at the ith node) of 50. All results were averaged ones for 100 attempts in the simulation.

The singularity of transfer matrix is compared in terms of the singularity factor ($=\Sigma\lambda_i^{-2}$) and the condition number in Fig. 9. The singularity factor is proportional to the square of velocity reconstruction error.[7,15] The condition number represents the upper bound of the excessive error that is expected in the inversion of the transfer matrix. In the far field, the transfer matrix generated by both BEM is nearly the same and highly singular. In the very close near field, the singularity of transfer matrix generated by CBEM is nearly

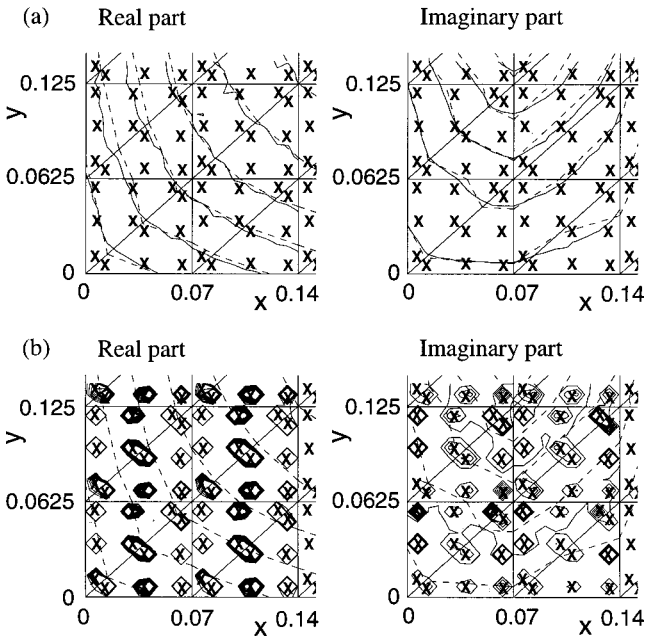


FIG. 8. Distributions of predicted field pressure at $z=1$ mm plane due to the excitation of (5,1) plate mode at 163 Hz, where the points represented by ''×'' denote Gauss integration points: ---, surface pressure; ———, field pressure. (a) NBEM; (b) CBEM.

constant for the distance variation, but the singularity due to NBEM decreases steadily with shortening the distance. This is because smaller high-order components of singular values are calculated by CBEM than those done by the nonsingular method. For an assumed noise variance, the S/N ratio $(=\Sigma|p_i|^2/\Sigma|n_i|^2)$ is compared in Fig. 10. The S/N ratio worsens in the far field since the amplitude of pressure decreases when the field point grows distant from the source surface.

In order to design a proper wave-vector filter for the regularization, the noise variance should be estimated since the true noise variance cannot be known *a priori* in the actual reconstruction process. The estimated noise variance can be defined as

$$\hat{\sigma}^2=\frac{1}{m-n}\|\{\tilde{p}\}_f-[G]\{\hat{v}\}_s\|_F^2, \tag{14}$$

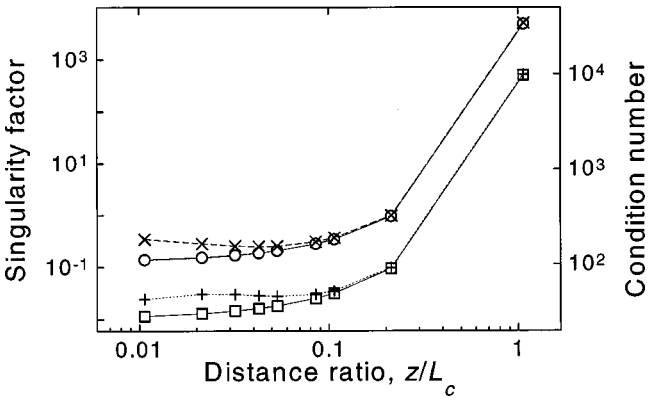


FIG. 9. Calculated singularity factor and condition number of the vibro-acoustic transfer matrix for the exterior problem in Fig. 5. Singularity factor: —○—, NBEM; ––×––, CBEM. Condition number: —□—, NBEM; ---+---, CBEM.

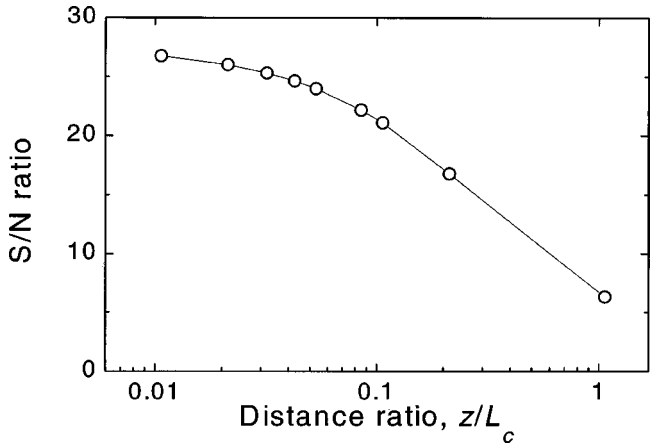


FIG. 10. S/N ratio of the noise-contaminated field pressure.

where $\|\ \|_F^2$ denotes the Frobenious norm and $\{\hat{v}\}_s$ is the estimated velocity vector. In the actual reconstruction based on the measurement, $\{\tilde{p}\}_f$ is the measured field pressure. However, in the simulation, $\{\tilde{p}\}_f$ is the field pressure obtained by using NBEM and is subject to the additive noise having a certain variance. In the simulation, the assumed noise variance of $\{\tilde{p}\}_f$ should be identical with $\hat{\sigma}^2$ in the ideal case. When $[G]$ is determined by CBEM, $[G]$ is distorted in the near field and, consequently, $\hat{\sigma}^2$ is much different from the preset value. Figure 11 demonstrates this fact. The estimated noise variance being determined by NBEM is the same everywhere with the target value. However, the noise variance estimated by CBEM is quite deviant from the target value in the near field. This can lead to the design mistake of the wave-vector filter and can then yield large reconstruction errors.

The error in reconstructing surface velocity before and after the iterative regularization is presented in Fig. 12, where the velocity reconstruction error is given by

$$\varepsilon=\frac{\|\{v\}_s-\{\hat{v}\}_s\|^2}{\|\{v\}_s\|^2}\times 100\ \ (\%). \tag{15}$$

Before applying the optimal wave-vector filter, the reconstruction error has a similar trend with that of singularity of the transfer matrix because small singular values affect the reconstructed results dominantly. The higher-order singular

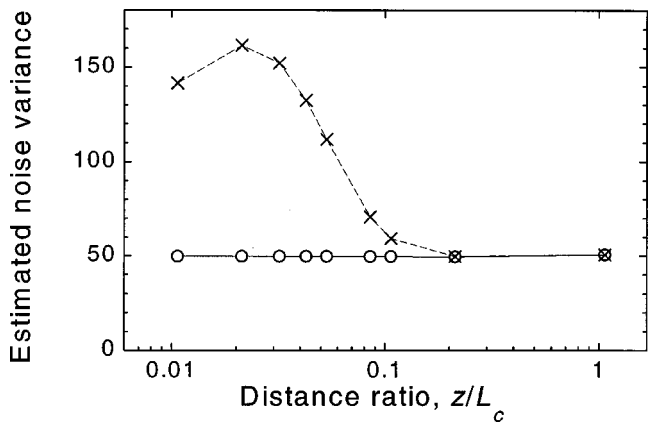


FIG. 11. Estimated noise variance being used for the vibro-acoustic reconstruction: —○—, NBEM; ––×––, CBEM.

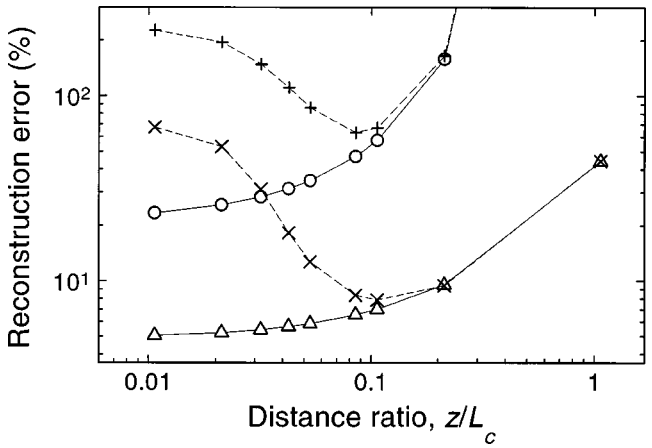


FIG. 12. The error in reconstructing surface velocity for the exterior problem. NBEM: —○—, before regularization; —△—, after regularization. CBEM: –+–, before regularization; ––×––, after regularization.

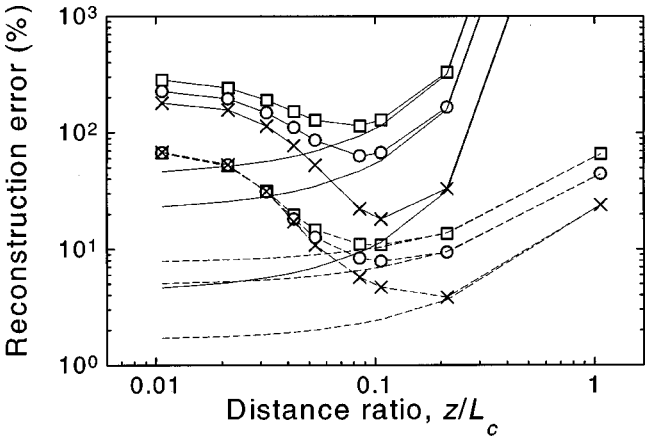


FIG. 13. Effect of target noise variance in the reconstruction error for surface velocity before and after applying the iterative regularization. The curves with symbols are for using CBEM and those without symbols are for NBEM. Before regularization: —□—, $\sigma^2=100$; —○—, $\sigma^2=50$; —×—, $\sigma^2=10$. After regularization: ––□––, $\sigma^2=100$; ––○––, $\sigma^2=50$; ––×––, $\sigma^2=10$.

values corresponding to the nonpropagating wave components becomes smaller as the field point grows distant from the source surface. These small singular values excessively amplify the noise included in the measured field pressure. For the field plane located farther than $0.2L_c$, the results of NBEM and CBEM calculations are nearly the same. When the field plane is located at or far from $1.07L_c$, the velocity reconstruction error is more than $O(10^5)\%$ for both NBEM and CBEM calculations. In the near-field plane located closer than $0.2L_c$, the resultant error of NBEM is reduced steadily closer to the source surface. This fact reveals that the precise information on the nonpropagating wave components is very important to enhance the accuracy of the final reconstruction of the source. When using the CBEM in NAH, the reconstruction error is minimal at certain distance near $r=0.1L_c$ and it increases abruptly closer to the source surface. The reconstruction error minimum is 23.3% at $z=0.010L_c$ when the NBEM is used in NAH and, in contrast, it is 63.2% at $z=0.107L_c$ when the CBEM is used. After applying the regularization, these errors are reduced to 1/4–1/5 of those before the application. The reconstruction error is 5.1% for NBEM and is nearly 70% for CBEM at $z=0.0107L_c$. The minimum reconstruction error in using CBEM can be found at $z=0.107L_c$; that is, 7.9%. By applying the regularization, the relative ratio of errors determined by both BEMs is reduced. From the foregoing facts, one can conclude that the regularization is required for the essential improvement of the reconstruction results.

It is not easy to use CBEM in the close near field for obtaining a fine reconstruction result, notwithstanding the fact that the difference in minimal error is small between the two methods. This is simply because it is very hard to know where the error is minimized. Figure 13 shows the change of reconstruction error with three different levels of noise variances (σ^2). As can be easily expected, the reconstruction error decreases with reducing the variance of added (or measured) noise. From this figure, one can find that the position of minimal error in using CBEM changes with varying the noise variance. If CBEM is utilized in an actual reconstruction process when the field point is located in $z<0.2L_c$, it is very hard to estimate the noise variance accurately. Furthermore, the correct distance for minimal error cannot be found easily and this means that the restoration accuracy cannot be assured. One can also observe from Fig. 13 that the change of noise variance in the signal does not invoke too much variation of the reconstruction accuracy in the close near field. This means that the type of BEM and the regularization are most crucial factors in obtaining the adequate precision for reconstructed results. The actual and reconstructed velocity fields by two boundary element methods are compared in Fig. 14 and the foregoing statement can be easily confirmed with this figure.

2. Interior problem

As an example of the interior problem, a parallelepiped cavity with the dimensions of 500(w)×500(h)×1500(l) mm was chosen and a similar analysis as in the preceding exterior problem was given. The boundary element model of the box consisted of 382 nodes and 761 linear triangular elements as in Fig. 15. The characteristic length of this model was $L_c=88.4$ mm. It was assumed that only an end plate at $z=0$ was vibrating in (2,2) mode and the other walls were rigid.

The field pressure estimated by CBEM is highly distorted in the close proximity to the source surface, where the distance from the nearest surface is less than about $0.2L_c$. This trend is the same with the case of the exterior problem. In the interior problem, the singularity at the rigid walls becomes relatively important compared with the exterior problem. This is because many field points can be located near the boundary surface. Consequently, the field pressure predicted by CBEM diverges near the Gauss integration points on the stationary walls even if the field point is located far from the source surface. The reconstruction error by using CBEM does not become equal to that by using NBEM, even in the far field.

In the simulation, the holographic reconstruction works were carried out for (2,2) mode at 200 Hz. The field pressure was calculated at the evenly arranged 81 points in a 50 ×50-mm pattern and the distance of these points from the

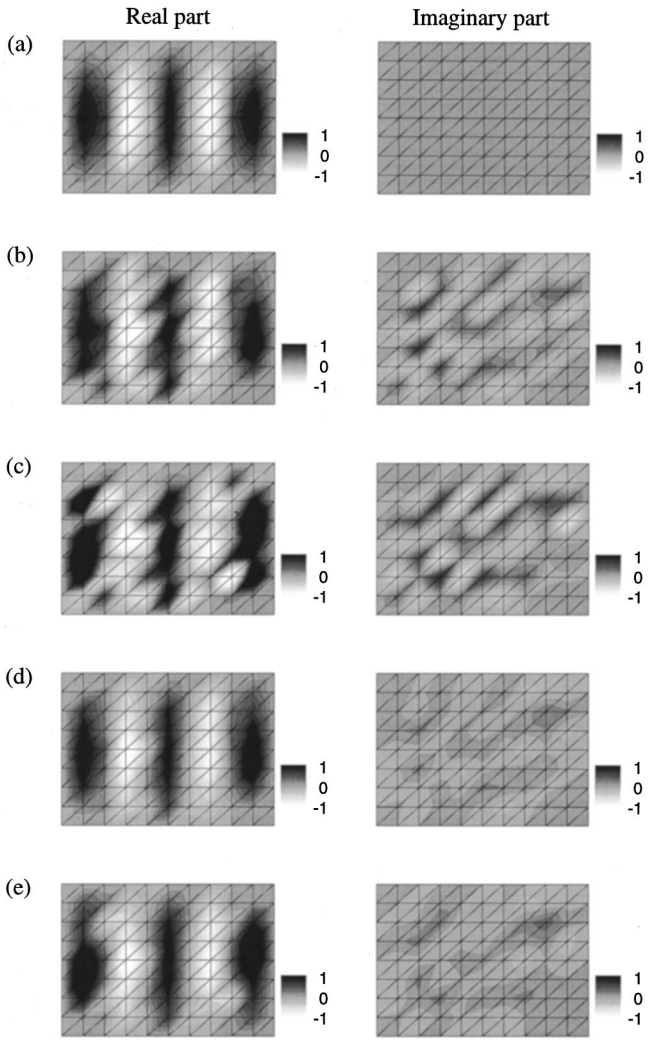


FIG. 14. Reconstructed surface velocity field from evenly distributed 117 field points at $z=4$ mm plane. (a) actual; (b) by NBEM and nonregularized ($e=31.4\%$); (c) by CBEM and nonregularized ($e=111\%$); (d) by NBEM and regularized ($e=5.65\%$); (e) by CBEM and regularized ($e=18.2\%$).

surface were within the range of 1–400 mm or, equivalently, $0.0113L_c-4.526L_c$. The noise variance is assumed as 50.

The reconstruction errors of surface velocity are compared in Fig. 16. Before applying the wave-vector filtering properly, the reconstruction error is 1066 and 60 012% for

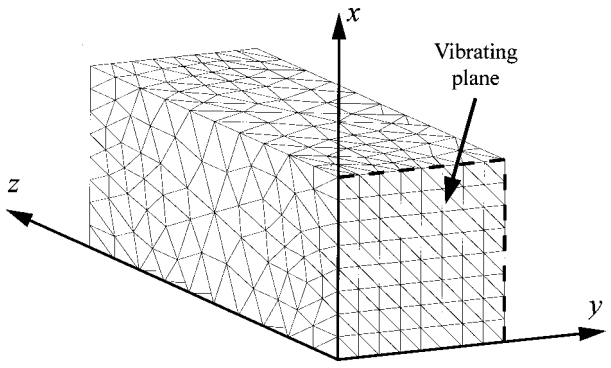


FIG. 15. BEM model of the parallelepiped cavity ($500\times500\times1500$ mm) for the simulation of the interior problem. Only an end plate at $z=0$ vibrates in (2,2) mode and the other walls are rigid.

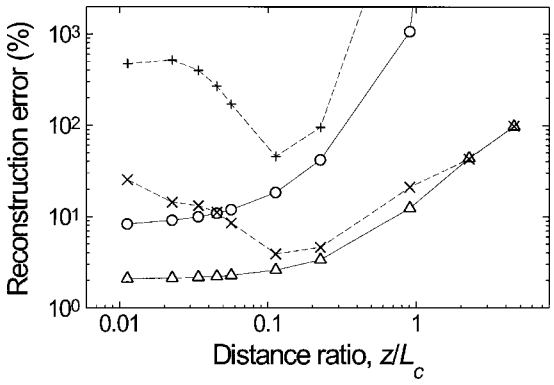


FIG. 16. The error in reconstructing surface velocity for the interior problem. NBEM: —○—, before regularization; —△—, after regularization. CBEM: –+–, before regularization; ––×––, after regularization.

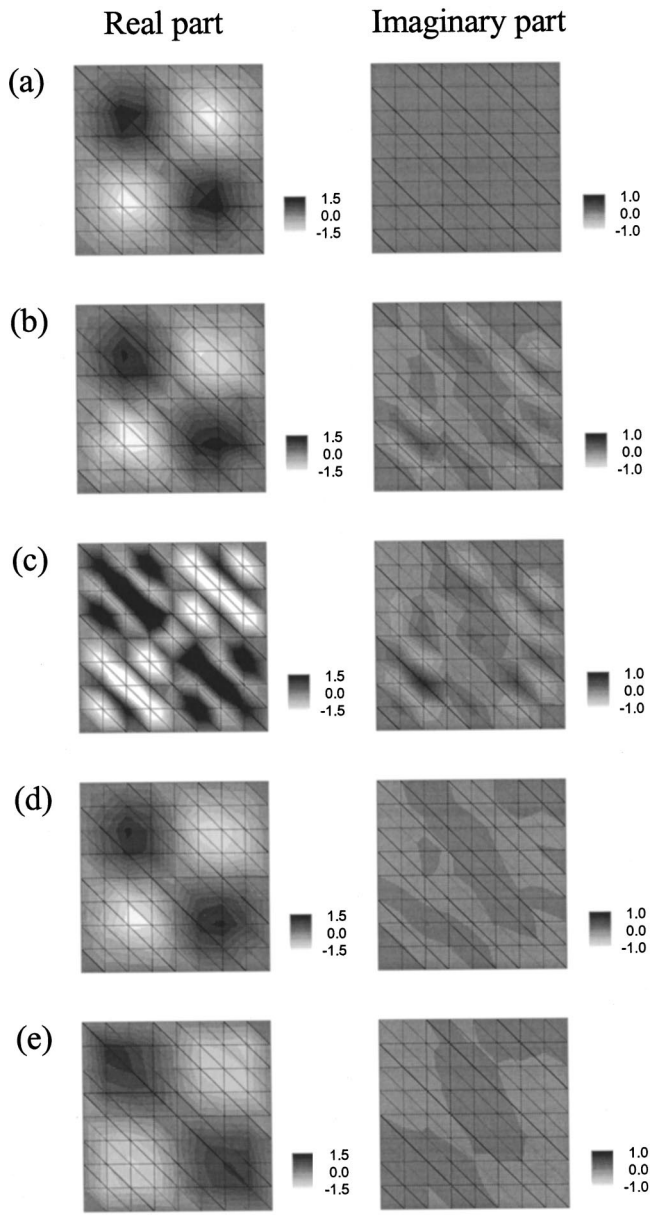


FIG. 17. Reconstructed surface velocity field from evenly distributed 81 field points at $z=3$ mm plane. (a) actual; (b) by NBEM and nonregularized ($e=10.0\%$); (c) by CBEM and nonregularized ($e=399\%$); (d) by NBEM and regularized ($e=2.17\%$); (e) by CBEM and regularized ($e=13.2\%$).

NBEM and CBEM, respectively, when $z=0.905L_c$. Even if the distance of the field plane from source surface is more than $0.2L_c$, the results of the two methods never get close. In the close near field, the reconstruction error in using NBEM is reduced steadily with shortening the distance, but the error increases in using CBEM. The latter trend is similar to the result of the exterior problem. The error minimum is 8.32% at $z=0.0113L_c$ for NBEM and it is 45.5% at $z=0.113L_c$ for CBEM. After applying the iterative regularization, these errors are reduced. When NBEM is adopted, the error minimum becomes 2.10% at $z=0.0113L_c$. When CBEM is utilized, the error minimum becomes 3.39% at $z=0.113L_c$. The actual and reconstructed velocity fields by two boundary element methods are compared in Fig. 17 when 81 evenly spaced field points at a $z=3$ mm plane are used.

IV. CONCLUDING REMARKS

The conventional boundary integral equation has the singularity problem that is originated by the singularity of the Green function near the Gauss integration points. This singularity problem induces inaccurate prediction of field properties and the problem can be solved by utilizing the nonsingular boundary integral equation. Through examples of exterior and interior problems, the strong points of nonsingular BEM in NAH are demonstrated. The effectiveness of nonsingular BEM has been investigated for the reduction of reconstruction error caused by the singularity problem in the near field. It is shown that the resolution of predicted field pressure is improved in the close near field by employing the nonsingular BIE. The BEM-based NAH inevitably requires the field pressure measured in close proximity to the source surface. Therefore, the nonsingular approach in BIE is recommended for yielding an accurate transfer matrix and thus improving the resolution of the reconstructed source field.

In order to generate the transfer matrix by using NBEM, extra computation time is required compared to the case using CBEM due to the complexity of BIE. However, in the actual NAH process, most of the time is spent measuring field pressures and regularizing the results. Once the transfer matrix is calculated for the model at the frequency of interest, it can be used repeatedly for later calculations.

[1] E. G. Williams and J. D. Maynard, ''Holographic imaging without the wavelength resolution limit,'' Phys. Rev. Lett. **45**, 554–557 (1980).
[2] J. D. Maynard, E. G. Williams, and Y. Lee, ''Near-field acoustic holography. I. Theory of generalized holography and the development of NAH,'' J. Acoust. Soc. Am. **78**, 1395–1413 (1985).
[3] H.-S. Kwon and Y.-H. Kim, ''Moving frame technique for planar acoustic holograph,'' J. Acoust. Soc. Am. **103**, 1734–1741 (1998).
[4] K. Gardner and R. J. Bernhard, ''A noise source identification technique using an inverse Helmholtz integral equation method,'' Trans. ASME, J. Vib., Acoust., Stress, Reliab. Des. **110**, 84–90 (1988).
[5] W. A. Veronesi and J. D. Maynard, ''Digital holographic reconstruction of source with arbitrarily shaped surfaces,'' J. Acoust. Soc. Am. **85**, 588–598 (1989).
[6] 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).
[7] 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).
[8] S.-C. Kang and J.-G. Ih, ''The use of partially measured source data in the near-field acoustic holography based on the BEM,'' J. Acoust. Soc. Am. **107**, 2472–2479 (2000).
[9] B.-K. Kim and J.-G. Ih, ''On the use of the BEM-based NAH for the vibro-acoustic source imaging on the nonregular exterior surfaces,'' Proc. Noise-Con **98**, 665–670, Ypsilanti, MI (1998).
[10] B. K. Gardner, J. S. Bolton, and D. L. Hallman, ''Nearfield acoustical holography in arbitrary geometries by the use of the boundary element method,'' Proc. Noise-Con **98**, 671–676, Ypsilanti, MI (1998).
[11] J.-G. Ih, S.-C. Kang, S.-J. Kim, and K.-S. Kang, ''Reconstruction of the vibro-acoustic field on the surface of the refrigerator compressor by using the BEM-based acoustic holography,'' Proc. 1998 Int. Compressor Eng. Conference, 525–529, Purdue University, West Lafayette, IN (1998).
[12] E. G. Williams, ''An overview of the inverse problem for sound reconstruction in interior spaces,'' J. Acoust. Soc. Am. **105**, 968(A) (1999).
[13] B.-U. Koo, B.-C. Lee, and J.-G. Ih, ''A non-singular boundary integral equation for acoustic problem,'' J. Sound Vib. **192**, 263–279 (1996).
[14] D. M. Photiadis, ''The relationship of singular value decomposition to wave-vector filtering in sound radiation problems,'' J. Acoust. Soc. Am. **88**, 1152–1159 (1990).
[15] B.-K. Kim and J.-G. Ih, ''Design of an optimal wave-vector filter for enhancing the resolution of reconstructed source field by near-field holography (NAH),'' J. Acoust. Soc. Am. **107**, 3289–3297 (2000).
[16] A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (McGraw-Hill, New York, 1981), Chap. 4.

A generalized Westervelt equation for nonlinear medical ultrasound

G. Taraldsen[a)]
SINTEF Telecom and Informatics, Acoustics N-7425 Trondheim, Norway

(Received 26 August 2000; accepted for publication 28 November 2000)

A model equation is derived for nonlinear medical ultrasound. Unlike the existing models, which use spatial coordinates, material coordinates are used and hence a model for a heterogeneous medium is able to be derived. The equation is a generalization of the Westervelt equation, and includes the nonlinearity, relaxation, and heterogeneity of soft tissue. The validity of the generalized Westervelt equation as a model equation for a Piola–Kirchoff acoustic pressure and as an equation for the acoustic pressure is discussed. In the second case it turns out that the model follows from two geometric approximations which are valid when the radius of curvature of the phase fronts is much larger than the particle displacements. The model is exact for plane waves and includes arbitrary nonlinearity in the stress–strain relation.
[DOI: 10.1121/1.1344157]

PACS numbers: 43.25.Cb, 43.25.Dc [MAB]

I. INTRODUCTION

It is well-known that nonlinear effects are important in medical ultrasound. This is clearly the case for lithotripsy[1] and hyperthermia[2] applications, but it has been recognized since 1980[3] that nonlinearity plays a role also at diagnostic levels. Nonlinearity at diagnostic levels introduces some problems, for instance when it comes to safety regulations based on linear theory,[4] but it may also be used to obtain a marked improvement in image quality in technically difficult patients by tissue harmonic imaging.[5] The principal ideas behind tissue harmonic imaging were published by Ward *et al.* in 1995 at the World Congress on Ultrasonics in Berlin, and similar results were published by various authors in 1997.[6–8] The reason for the current success in applying nonlinear effects for image improvement is largely due to recent improvements in transducer technology, including in particular fabrication of transducers with sufficient sensitivity over a wide frequency range. This implies that it is now possible to utilize finer properties of the received ultrasound echos, and motivates the investigation of detailed models for nonlinear ultrasound in tissue. Considerable attention is currently being devoted to the modeling and experimental testing of propagation and focusing of intense sound beams. This attention is due in large part to applications in medicine. Despite the fact that tissue is heterogeneous, the main focus has been on models from underwater acoustics which are only valid for a single material, such as water.

In 1971 Kuznetsov stated (Ref. 9, p. 467): *Unlike linear theory, which uses a wave equation, the theory of nonlinear acoustics is beset with added difficulties in connection with the absence of a nonlinear wave equation*. We feel that this is still true, but now in the heterogeneous case, and in particular for nonlinear ultrasound in tissue. The main content of the present article is a derivation of a simple model equation in material coordinates which includes heterogeneity. The resulting generalized Westervelt equation may be transformed to the more traditional spatial description, but the resulting equation is much more complex even for plane waves. When using material coordinates it is customary to derive a wave equation for the displacement field,[10–12] but it seems to be a novelty here to derive an equation for the pressure. We will not consider numerical solution of the suggested generalized Westervelt equation, but many of the existing algorithms in the literature may be adapted to either a solution by propagation along the beam direction[13–16] or by propagation in time.[12,17–20] Anisotropy may be a significant factor in the modeling of ultrasound through the muscle layers in the body wall,[21–23] but this will be ignored in the following. To put our derivation in perspective, we start with a brief review of some existing models. More thorough introductions to nonlinear acoustics may be found in recent books on this subject.[10,11,24]

II. EXISTING MODELS FOR ULTRASOUND IN TISSUE

The majority of the existing models for ultrasound in tissue (Ref. 10, Biomedical applications, pp. 421–447) are modifications of well-known models from underwater acoustics to include relaxation loss. The most popular equation is the KZK equation,[9,25] which may be viewed as a parabolic approximation to the Westervelt equation,[26] which may be viewed as a simplification of the Kuznetsov equation.[9] The lossless Kuznetsov equation is a second-order approximation of the following wave equation (Ref. 27, p. 713):

$$-\ddot{\phi}+c^2\Delta\phi=(\partial_t+\tfrac{1}{2}\nabla\phi\cdot\nabla)(\nabla\phi)^2. \tag{1}$$

In the lossless case the above wave equation is an exact model equation for a potential velocity field $u=\nabla\phi$. An arbitrary nonlinear equation of state gives the speed of sound c as a function of $\dot{\phi}+u^2/2$, and the above wave equation is then a closed equation for the velocity potential ϕ.[28] The

a)Electronic mail: gunnar.taraldsen@informatics.sintef.no

effect of heat conduction and viscosity, and more general relaxation as is needed in medical ultrasound, can be included to first order by standard methods. This gives a starting point for numerical and experimental investigations of nonlinear effects in medical ultrasonics. Investigations by numerical solutions of the KZK equation and experiments in a water tank were initiated in 1980[3] and have continued up to the present time,[29] with good agreement between theory and experiment.[6,10,11] Some have also given numerical solutions of the Westervelt equation combined with the bioheat equation in order to investigate nonlinear heating effects.[30]

However, these models are correct only for homogeneous media, such as water, but tissue is heterogeneous. In the linear case a conventional model which includes heterogeneity is given by $\nabla p = -\rho_0 \dot{u}$ and $\nabla \cdot u = -\kappa_0 *_t \dot{p}$. Instead of a single unknown field ϕ as in Eq. (1), we have here four unknown fields given by the velocity field u and the acoustic pressure p. Relaxation is modeled by a time-dependent compressibility κ_0, and the time convolution $*_t$ can be reformulated in terms of extra relaxation equations.[20] Heterogeneity may be modeled by letting the ambient mass density ρ_0 and compressibility κ_0 be functions of position given by realistic tissue cross sections.[18,31] A related model, which includes the second-order parameter B/A (Ref. 10, Chap. 2) of nonlinearity is given by $\nabla p = -\rho_0 \dot{u}$ and $p = -(1/\kappa_0)(\nabla \cdot \psi + (B/2A) \times (\nabla \cdot \psi)^2) - (\mu_B + (4/3)\mu)\nabla \cdot u$, where ψ is the displacement field.[17,32,33] The model includes viscosity μ and bulk viscosity μ_B. A simpler, but still interesting case is given by just introducing inhomogeneity through variations in the ambient state.[34,35] Considering only second-order effects for a thermoviscous inhomogeneous fluid, one arrives at a generalized Westervelt equation [Ref. 36, Eq. (255)][37]

$$\ddot{p} - \kappa_0^{-1} \nabla \cdot (\rho_0^{-1} \nabla p) = (\beta_n \kappa_0 p^2 + \nu \kappa_0 \dot{p})^{\cdot\cdot}. \tag{2}$$

Second-order nonlinearity is included through $\beta_n = 1 + B/(2A)$, and ν includes viscosity and heat conduction to first order. This is a generalized Westervelt equation compared to the original[26,10] in that the material parameters ρ_0, κ_0, β_n, ν may depend on position as a result of variations in the ambient state. This is adequate in underwater acoustics, where gravity and temperature variations give a variation in the ambient density and pressure of the seawater, but seems questionable in tissue where different materials are present.

However, it is difficult to find references to models based on first principles which includes the nonlinearity and heterogeneity of soft tissue. The reason for this may be that the use of spatial coordinates quickly leads to major problems in the heterogeneous case. The problem is already manifest in the simplest lossless case, where the density ρ is a function of the pressure P. This function describes the material, and as such we have one function for each material point x. Let $\rho = \tilde{\rho}(P,x)$. In the derivation of the equations of fluid mechanics it is assumed that there is a one–one correspondence between material points x and spatial points x' at each time t. This means that we may consider the fields either as functions of (x,t) (material description) or as functions of (x',t) (spatial description). The displacement field ψ is defined by $x' = x + \psi$. In linear theory we may use the approximation $\rho = \tilde{\rho}(P,x) \approx \tilde{\rho}(P,x')$, but not in nonlinear theory. In nonlinear theory we are forced to use $\rho(x',t) = \tilde{\rho}(P(x',t), x' - \psi(x',t))$ when we use spatial coordinates (x',t). This gives an awkward equation for ψ when inserted into the conservation equations for mass and momentum. It is well-known that this awkwardness is removed if we use material coordinates, and this is also demonstrated in the following.

III. PIOLA–KIRCHOFF PRESSURE WAVES

We consider sound in a medium where conservation of momentum (3) and elastic properties (4) are given in material coordinates by

$$\rho_0 \ddot{\psi} = -\nabla p, \tag{3}$$

$$-\nabla \cdot \psi = D(p). \tag{4}$$

It is well-known that the conservation equation for momentum becomes linear in material coordinates by the introduction of the Piola–Kirchoff pressure tensor (Ref. 38, p. 106, Ref. 10, p. 265, Ref. 39, p. 220), but here we have simplified further by restriction to deformations where the Piola–Kirchoff pressure tensor can be replaced by a scalar acoustic Piola–Kirchoff pressure p. The pressure p is force per unit undeformed area and includes viscous forces and possibly more general contributions due to relaxation. The other quantities in Eq. (3) are the ambient mass density ρ_0 and the particle displacement ψ. The partial derivatives are with respect to material coordinates (x,t). In Eq. (4) we have restricted to deformations where the divergence of the displacement field ψ is given by the history of p through the nonlinear condensation D. A typical model is given by $D = \kappa_0(p - n(p) - Lp)$, where $\kappa_0(p - n(p))$ corresponds to an equation of state at constant entropy, and L is a linear operator corresponding to relaxation corrections. This assumption gives the generalized Westervelt equation

$$\ddot{p} - \kappa_0^{-1} \nabla \cdot (\rho_0^{-1} \nabla p) = \ddot{n}(p) + \ddot{L}p. \tag{5}$$

The medium parameters κ_0, ρ_0, the nonlinearity function n, and the linear relaxation operator L all depend on the material point x and hence this generalized Westervelt equation describes waves in a heterogeneous medium. The model goes beyond second-order nonlinearity in that arbitrary nonlinearity is allowed in n. It is for instance possible to express n to arbitrary order in p from a virial expansion of the equation of state.[28] The model goes beyond viscous effects in that the relaxation operator L is allowed to be a convolution operator. This is necessary in order to model the observed frequency dependence of absorption in medical ultrasonics.[22] If a nonlinear relaxation operator is to be considered, then these considerations should also include a nonlinear heat equation. This can be combined with a material equation where the condensation D is also a function of temperature or entropy, but we will not pursue this further here.

The condensation D can in particular cases be found by inverting a more conventional material equation. It is sufficient to consider a plane deformation and we choose coordinates such that $\psi' = \nabla \cdot \psi$. A conventional material equation for a thermoviscous solid is $-p = (K + 4G/3)\psi' + [(3/2)(K + 4G/3) + A' + 3B' + C']\psi'^2 + \nu\dot{\psi}'$[11,38] (Ref. 10, p. 267).

Here, K and G are, respectively, the bulk and shear moduli, and A', B', C' are the third-order moduli used by Landau and Lifshitz. The coefficient ν includes bulk and shear viscosity, and the effect of thermal conductivity to first order. The Fourier transformation of the material equation in time implies that it may be viewed as a second-order material equation in small parameters given by the condensation $D=-\psi'$ and the frequency ω. The Fourier-transformed equation can be solved with respect to the condensation D and expanded to second order in p, ω. The result after an inverse Fourier transform is $D=\kappa_0 p-\beta_n(\kappa_0 p)^2-\nu\kappa_0^2\dot{p}$, with the identifications $\kappa_0=1/(K+4G/3)$ and $\beta_n=3/2+(A'+3B'+C')/(K+4G/3)$. In particular we have $\beta_n=3/2$ in the case of a linear stress–strain relation, because of the quadratic term in the definition of the strain. A similar argument for a liquid gives $\beta_n=1+B/(2A)$ and $\kappa_0=1/K$. In both cases, we get

$$n(p)=\beta_n\kappa_0 p^2, \quad Lp=\nu\kappa_0\dot{p}. \tag{6}$$

We see that Eq. (5) for a thermoviscous solid coincides with the generalized Westervelt equation given in the previous section. The model is, however, not equivalent since the partial derivatives in Eq. (5) is with respect to material coordinates, and we allow a separate equation of state for each material point. Additionally, Eq. (5) is an equation for the Piola–Kirchoff acoustic pressure. The above turns out also to be a justification of the model used by Wojcik *et al.*,[32,17,33] but we use an inverted material equation and have a somewhat different interpretation. A similar second-order equation in material coordinates and with the ordinary pressure was derived by Angelsen [Ref. 40, Eqs. (12.100), (12.144)]. He has an additional term $\kappa_0 p\nabla_\perp^2 p$, and in his derivation it is assumed that *curl* $\psi=0$. His argumentation is complementary to the arguments we have given.

We will briefly sketch how the above argumentation for a thermoviscous solid can be generalized. General conventional material equations are of the form $p=p(\partial_x\psi,\partial_x\dot{\psi})$. The dependence on $\partial_x\dot{\psi}$ includes in particular the possibility of modeling viscous effects. More generally, it is reasonable to assume a material equation of the form $p=p_{\mathrm{op}}(\partial_x\psi)$ where p_{op} is an operator which maps tensor-valued functions of time to tensor-valued functions of time. This corresponds to the intuition that the pressure tensor at a certain time t depends on the history of the deformation up to time t. The viscosity assumption corresponds intuitively to a material with short memory since the dependence on the past is just through the time derivative $\partial_x\dot{\psi}$ of the deformation $\partial_x\psi$. It is convenient to replace the material assumption $p=p_{\mathrm{op}}(\partial_x\psi)$ with an inverted material assumption of the form $-\partial_x\psi=D_{\mathrm{op}}(p)$. It seems no less intuitive to assume that the deformation depends on the history of the stress, than to assume that the stress depends on the history of the deformation. The operator D_{op} may also in some cases be obtained directly by inverting p_{op} (Ref. 38, p. 11). The trace of D_{op} gives the condensation $-\nabla\cdot\psi=D(p)$ at a certain time t as function of the history of the pressure p up to time t. This holds generally, but in particular when we restrict attention to waves where p is scalar.

Since material coordinates are unconventional in medical ultrasonics, we will briefly justify that conservation of momentum is given by Eq. (3). Let P be the Piola–Kirchoff pressure tensor so that $PS(dx)$ is the differential force acting on the oriented undeformed area element $S(dx)$. Conservation of momentum is then given by Newton's law $\partial_t\int_R\rho_0\dot{\psi}dx=-\int_{\partial R}PS(dx)+F$, where F is the total external force acting on the material region R. This includes in particular gravitational forces. If we assume that F is time independent, then the divergence theorem gives conservation of acoustic momentum as in Eq. (3). Here, we have introduced the acoustic Piola–Kirchoff pressure p by $\nabla\cdot P=\nabla\cdot P_0+\nabla p$, and used an equilibrium consideration to cancel the term with the ambient Piola–Kirchoff pressure tensor P_0 against the external field F.

In the above it must be understood that we have restricted the class of possible waves under consideration by assuming that the Piola–Kirchoff volume force is given as the gradient of a scalar p. This assumption is fulfilled if and only if the acoustic wave is irrotational in the sense that $curl(\rho_0\psi)=0$. In analogy with the acoustic mode known from linear theory, we refer to these waves as Piola–Kirchoff acoustic waves. This assumption differs from the conventional in that the ambient mass density is included, and in the case of a constant ambient mass density it still differs: *curl* with respect to material coordinates is different from *curl* with respect to spatial coordinates. The physical relevance of such waves depends on the boundary conditions.

IV. PRESSURE WAVES

In this section we will derive the ultrasound equation (5) in material coordinates as an approximate equation for the conventional pressure. We simplify the discussion by only considering materials for which the mass density is a function of the pressure, and will thereby also neglect energy loss. The loss terms represents no problem when they are only accounted for to linear order and can be taken into account by standard methods. We consider the case where external forces are canceled by gradients in the ambient pressure, and conservation of momentum is then given by[39,41]

$$\rho_0\ddot{\psi}=-\det(F)F^{-1}\nabla p. \tag{7}$$

The presence of the deformation gradient tensor $F=1+\partial_x\psi$ in Eq. (7) and not in Eq. (3) is because p in Eq. (7) is the ordinary acoustic pressure.

We assume that we have an equation of state on the form $\rho=\hat{\rho}(p,x)$. The dependence on the material coordinate x gives the possibility of having a different material in every material point, but an x dependence may also be due only to variations in the ambient state as for an inhomogeneous liquid. This assumed pressure–density relation can be turned into a stress–strain relation by conservation of mass, as we will explain. The Jacobi determinant change of variables theorem gives us that $\rho\det(F)$ is the mass per material volume. With $\rho_0(x)=\rho(x,0)$ we obtain the equation $\rho_0=\rho\det F$ for the conservation of mass, since $F(x,0)=1$. This equation holds under more general conditions than the cor-

responding differential equation in Euler coordinates, including in particular the possibility of discontinuities in ρ. The main point is, however, the simplification introduced by the removal of the unknown field ρ in the conservation equations. The conclusion is a stress–strain relation of the form (Ref. 10, p. 265)

$$\det(F) = \frac{\rho_0}{\hat{\rho}(p,x)}. \tag{8}$$

Equations (7) and (8) together with relevant boundary conditions determine the displacement field ψ and the pressure field p. The equations are exact and are comparable with the simpler and exact description given by Eq. (1). However, Eq. (1) is only valid for potential flow, and it is well-known that this is not sufficiently general for the heterogeneous case. An example is given by a boundary between two materials where shear waves may be created and as a result the total velocity field fails to be irrotational. It seems that Eqs. (7) and (8) are the simplest possible exact description in the heterogeneous case for general acoustic waves. In the compressible case it is typically possible to solve Eq. (8) with respect to the pressure p, and insertion into Eq. (7) gives a single vector equation for the displacement field. This equation is related to the arbitrary order equation for the displacement field known in elasticity [Ref. 10, p. 267, Eq. (15)].

However, given some assumptions on the geometry of the waves it is possible to simplify further. The standard procedure is to simplify the equation for the displacement field (Ref. 10, p. 267), but we choose instead to consider the pressure. For plane waves we have simply

$$\det(F) = 1 + \nabla \cdot \psi, \tag{9}$$

$$\det(F) F^{-1} \nabla p = \nabla p. \tag{10}$$

By combining Eqs. (8) and (9) we get a special case of Eq. (4). Equation (7) is simplified by Eq. (10) into Eq. (3). The conclusion is that the generalized Westervelt equation (5) is exact for plane waves as an equation for the acoustic pressure.

In order to arrive at the generalized Westervelt equation for more general waves, we can seek to obtain the condensation $D = -\operatorname{Tr} \partial_x \psi$ as a function of pressure. From Eq. (8) this means to obtain the trace $\operatorname{Tr} \partial_x \psi$ as a function of the determinant $\det(1+\partial_x \psi)$. This is clearly impossible for general deformations, and relates to a restriction on the derivation given in the previous section: It is reasonable to assume an invertible stress–strain relation, but it is not possible in general to recover the deformation gradient F from the strain $E = (F^T F - 1)/2$. If we restrict to particular waves, for instance such that F is positive, then F is given by E and the spectral theorem. In this section we do not, however, have an invertible stress–strain relation, but have rather started with a pressure–density relation. In this case, if we assume additionally that F is positive, then the condensation D is a function of the pressure if and only if we have a plane deformation. This leads us to consider the class of waves where Eqs. (9) and (10) are sufficiently good approximations. This class of waves include waves where the transversal coordinates are slow as in a sound beam. By introducing a small parameter in the scaling of the transversal coordinates, it follows that the generalized Westervelt equation (5) is a second-order equation for the acoustic pressure. We choose not to give the details leading to this conclusion, but find it more enlightening to consider a special case.

We finish this section by considering spherical waves to get insight into the approximate Eqs. (9) and (10). For spherical deformations we have

$$\det(F) = \frac{R^2}{r^2} R', \tag{11}$$

$$1 + \nabla \cdot \psi = R' + 2\frac{R-r}{r}, \tag{12}$$

$$\det(F) F^{-1} \nabla p = \left(\frac{R}{r}\right)^2 \nabla p, \tag{13}$$

where R' is the derivative of the radial particle position R with respect to the radial material coordinate r. For spherical waves we conclude that the geometric approximations (9) and (10) are given by ignoring the material displacement $R - r$ compared to r. This is a very reasonable approximation also when this is generalized to sound beams: The displacement for a 2.5-MHz pressure pulse in water with amplitude in the range -2 to 7 MPa is bounded by 0.2 μm and the radius of curvature is on the cm scale.

The generalized Westervelt equation (5) is not valid when the radius of curvature of the phase front is comparable with the material displacement. An example is given by the case of the modeling of the sound field from contrast agents which may have an equilibrium radius equal to 4 μm, and displacements as large as 5 times the radius.[42] This case is better modeled by Eq. (1).

V. CONCLUSION

We have argued that the commonly accepted use of material coordinates in elasticity is a good choice also in the case of soft tissue, because of heterogeneity. The natural models are then found in the theory of elasticity. Conventional theory gives differential equations for the displacement field, but we have obtained a simpler model for the pressure. In one approach we derived a generalized Westervelt equation for the Piola–Kirchoff acoustic pressure, which is force per undeformed area. The model is exact for a class of waves which in particular includes plane waves in a heterogeneous solid with relaxation loss and a stress–strain relation. In a second approach we derived the generalized Westervelt equation as an approximate model for the acoustic pressure in a heterogeneous material given by a pressure–density relation. The approximation is valid for sound beams when the radius of curvature of the phase front is much larger than the displacement field. The importance of the generalized Westervelt equation compared to the corresponding equations from underwater acoustics is that it includes the modeling of heterogeneous materials as is required for medical ultrasound in tissue. Compared to well-known models from elasticity, it is a simplification which is sufficiently general for many applications in medical ultrasonics.

ACKNOWLEDGMENTS

We are pleased to acknowledge most stimulating comments and discussions on the modeling and application of nonlinear medical ultrasound with S. Tjøtta (Bergen), T. Johansen (NTNU), B. Angelsen (NTNU), and H. Torp (NTNU). It was in particular B. Angelsen who introduced us to this field of research and suggested a material description of medical ultrasound. Thanks to A. C. Baker (Bergen) for comments regarding the introduction of tissue harmonic imaging.

[1] B. Granz, ''Measurement of shock wave properties after the passage through a tissue mimicking material,'' IEEE Ultrasonics Symposium, pp. 1847–1851 (1994).
[2] K. Hynynen, ''The effects of nonlinear propagation in ultrasound hyperthermia,'' IEEE Conference in Medicine and Biology, pp. 256–257 (1990).
[3] T. G. Muir and E. L. Carstensen, ''Prediction of nonlinear acoustic effects at biomedical frequencies and intensities,'' Ultrasound Med. Biol. **6**, 345–357 (1980).
[4] F. A. Duck, ''Acoustic saturation and output regulation,'' Ultrasound Med. Biol. **25**(6), 1009–1018 (1999).
[5] F. Tranquart, N. Grenier, V. Eder, and L. Pourcelot, ''Clinical use of ultrasound tissue harmonic imaging,'' Ultrasound Med. Biol. **25**, 889–894 (1999).
[6] B. Ward, A. C. Baker, and V. F. Humphrey, ''Nonlinear propagation applied to the improvement of resolution in diagnostic medical ultrasound,'' J. Acoust. Soc. Am. **101**, 143–154 (1997).
[7] T. Christopher, ''Finite amplitude distortion-based inhomogeneous pulse echo ultrasonic imaging,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**(1), 125–139 (1997).
[8] M. A. Averkiou, D. N. Roundhill, and J. E. Powers, ''A new imaging technique based on the nonlinear properties of tissue,'' IEEE Ultrasonics Symposium, pp. 1561–1566 (1997).
[9] V. P. Kuznetsov, ''Equations of nonlinear acoustics,'' Sov. Phys. Acoust. **16**, 467–470 (1971).
[10] M. F. Hamilton and D. T. Blackstock, *Nonlinear Acoustics* (Academic, New York, 1998).
[11] K. Naugolnykh and L. Ostrovsky, *Nonlinear Wave Processes in Acoustics* (Cambridge University Press, Cambridge, England, 1998).
[12] Z. You, M. Lusk, R. Ludwig, and W. Lord, ''Numerical simulation of ultrasonic wave propagation in anisotropic and attenuative solid materials,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **38**(5), 436–4445 (1991).
[13] S. I. Aanonsen, T. Barkve, J. N. Tjøtta, and S. Tjøtta, ''Distortion and harmonic generation in the nearfield of a finite amplitude beam.'' J. Acoust. Soc. Am. **75**, 749–768 (1984).
[14] P. T. Christopher and K. J. Parker, ''New approaches to nonlinear diffractive field propagation,'' J. Acoust. Soc. Am. **90**, 488–499 (1991).
[15] J. Tavakkoli, O. A. Sapozhnikov, R. Souchon, and D. Cathignol, ''A time-domain numerical model for simulating the acoustic field of a high-amplitude focused source in biological media,'' IEEE Ultrasonic Symposium, pp. 1737–1740 (1997).
[16] L. Filipczynski, T. Kujawska, R. Tymkiewicz, and J. Wojcik, ''Nonlinear and linear propagation of diagnostic ultrasound pulses,'' Ultrasound Med. Biol. **25**(2), 285–299 (1999).
[17] G. Wojcik, B. Fornberg, R. Waag, L. Carcione, J. Mould, L. Nikodym, and T. Driscoll, ''Pseudospectral methods for large-scale bioacoustic models,'' IEEE Ultrasonics Symposium, pp. 1501–1506 (1997).
[18] T. D. Mast, L. M. Hinkelman, M. J. Orr, V. W. Sparrow, and R. C. Waag, ''Simulation of ultrasonic pulse propagation through the abdominal wall,'' J. Acoust. Soc. Am. **102**, 1177–1190 (1997).
[19] Q. H. Liu, ''The pseudospectral time-domain (PSTD) algorithm for acoustic waves in absorptive media,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**(4), 1044–1055 (1998).
[20] X. Yuan, D. Borup, J. Wiskin, M. Berggren, and S. A. Johnson, ''Simulation of acoustic wave propagation in dispersive media with relaxation losses by using FDTD method with PML absorbing boundary condition,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**(1), 14–23 (1999).
[21] L. M. Hinkelman, T. D. Mast, L. A. Metlay, and R. C. Waag, ''The effect of abdominal wall morphology on ultrasonic pulse distortion. I. Measurements,'' J. Acoust. Soc. Am. **104**, 3635–3649 (1998).
[22] F. A. Duck, *Physical Properties of Tissue* (Academic, New York, 1990).
[23] Y. C. Fung, *Biomechanics: Mechanical Properties of Living Tissue* (Springer, Berlin, 1993).
[24] R. T. Beyer, *Nonlinear Acoustics* (Navy Sea Systems Command, Washington, D.C., 1974).
[25] E. A. Zabolotskaya and R. V. Khokhlov, ''Quasi-plane waves in the nonlinear acoustics of confined beams,'' Sov. Phys. Acoust. **15**, 35–40 (1969).
[26] P. J. Westervelt, ''Parametric acoustic array,'' J. Acoust. Soc. Am. **35**, 535–537 (1963).
[27] S. Tjøtta, ''Finite amplitude ultrasound beams,'' IEEE Ultrasonic Symposium, pp. 709–714 (1994).
[28] G. Taraldsen, ''Nonlinear ultrasound equations with arbitrary order terms,'' Technical report, University of Trondheim, 2000.
[29] D. Ding and J. Lu, ''Higher-order harmonics of limited diffraction Bessel beams,'' J. Acoust. Soc. Am. **107**, 1212–1214 (2000).
[30] I. M. Hallaj and R. O. Cleveland, ''FDTD simulation of finite-amplitude pressure and temperature fields for biomedical ultrasound,'' J. Acoust. Soc. Am. **105**, L7–L12 (1999).
[31] T. D. Mast, L. M. Hinkelman, L. A. Metlay, M. J. Orr, and R. C. Waag, ''Simulation of ultrasonic pulse propagation, distortion, and attenuation in the human chest wall,'' J. Acoust. Soc. Am. **106**, 3665–3677 (1999).
[32] G. Wojcik, J. Mould, F. Lizzi, N. Abboud, M. Ostromogilsky, and D. Vaughan, ''Nonlinear modeling of therapeutic ultrasound,'' IEEE Ultrasonics Symposium, pp. 1617–1622 (1995).
[33] G. Wojcik, J. Mould, S. Ayter, and L. Carcione, ''A study of second harmonic generation by focused medical transducer pulses,'' IEEE Ultrasonics Symposium, pp. 1583–1588 (1998).
[34] J. N. Tjøtta, E. Reiso, and S. Tjøtta, ''Nonlinear equations of acoustics in inhomogeneous fluids,'' J. Acoust. Soc. Am. **83**, S1–S5 (1988).
[35] R. S. Kulkarni, W. L. Siegmann, and M. D. Collins, ''Nonlinear wide-angle paraxial acoustic propagation in shallow-water channels,'' J. Acoust. Soc. Am. **102**, 224–232 (1997).
[36] E. Reiso, ''Nonlinear equations of acoustics in inhomogeneous, thermoviscous fluids,'' Technical Report No. 89, University of Bergen, 1991.
[37] O. V. Rudenko, A. K. Sukhorukova, and A. P. Sukhorukov, ''Equations of high-frequency nonlinear acoustics for inhomogeneous media,'' Acoust. Phys. **40**, 264–268 (1994).
[38] L. D. Landau and E. M. Lifshitz, *Theory of Elasticity* (Pergamon, New York, 1998).
[39] L. E. Malvern, *Introduction to the Mechanics of a Continuous Medium* (Prentice-Hall, Englewood Cliffs, NJ, 1969).
[40] B. A. J. Angelsen, *Waves, Signals and Signal Processing in Medical Ultrasonics* (Department of Physiology and Biomedical Engineering, NTNU, Norway, 2000).
[41] Miroslav Silhavý, *The Mechanics and Thermodynamics of Continuous Media* (Springer, Berlin, 1997).
[42] T. G. Leighton, *The Acoustic Bubble* (Academic, New York, 1994).

Vortex sound generation due to a flow impedance discontinuity on a flat surface

S. K. Tang
Department of Building Services Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China

K. M. Li
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China

(Received 8 July 1999; revised 9 October 2000; accepted 19 December 2000)

The sound generated by the unsteady motion of a vortex filament moving over a flat boundary with a sharp flow impedance discontinuity is studied theoretically. Theoretical results show that the vortex filament undergoes significant accelerating or decelerating motions and radiates sound at the instant when it moves across the plane of impedance discontinuity. The accelerations and decelerations of the vortex filament are shown to be the major mechanisms of sound generation. The sound so produced has a large low-frequency content such that the change in the flow impedance affects only the sound generation process but not the subsequent sound propagation to the far field. [DOI: 10.1121/1.1349186]

PACS numbers: 43.28.Ra, 43.28.Py, 43.50.Nm [LCS]

I. INTRODUCTION

Flows in air ducts are turbulent. It is well-known that this turbulence will react with reactive duct elements, such as bends and dampers, to produce sound.[1] However, though Curle[2] had introduced a formulation for the production of sound by a turbulent flow close to solid boundaries, theoretical study of this turbulent sound-production process is still difficult due to the unknown flow dynamics within the turbulence. Experimental investigation into this topic is also not easy, as the experimental setup has to be large and it is very difficult to measure flow velocities and the hydrodynamic pressure on or close to a solid surface without using intrusive devices. In addition, the magnitudes of the associated sound-pressure fluctuations are usually too low to be measured with reliable accuracy.

Vortices are believed to be the backbone of turbulent flows.[3] It is also possible to calculate their motions under the influence of solid boundaries. A typical example of this can be found in Milne-Thomson,[4] where the speed of a two-dimensional vortex filament over a rigid flat plate is related to its perpendicular distance from the plate. The theory of vortex sound of Powell[5] has significant contributions to the study of turbulent flow noise as its enables the estimation of the sound field once the vortex dynamics are obtained. Though vortices are drastic simplifications of the real turbulent flows, the vortex theory[5] still provides insights to many practical low Mach number aerodynamic noise problems. Some examples can be found in Obermeier,[6] Tang and Ko,[7] and Tang and Ffowcs Williams.[8]

Self-generated noise is the noise produced by a turbulent flow when it flows along a traditional dissipative silencer in an air duct.[1] This unwanted but currently unavoidable noise lowers the efficiency of the silencer. Its effect is especially substantial in high-attenuation silencers. However, its generation mechanism is not very clear, though it must be related to the turbulent flow. The results of Quinn and Howe[9] and Ffowcs Williams[10] have showed that sound can be produced by the interaction of the turbulent pressure fluctuations within the pores on the steel plate covering the porous material inside the silencer. However, the effects of the porous materials are not explored. If one models the turbulent flow as patches of incompressible vortices, it is then expected that the propagation speeds of the vortices will change when they enter or leave the silencer as the porous material offers an impedance to the fluid flow.[11] The vortex sound theory of Powell[5] then suggests a possibility of sound production by the silencer.

In the present study, the sound produced by a two-dimensional vortex filament in the presence of a flat surface with a sharp flow impedance discontinuity in an otherwise unbounded medium is investigated theoretically. An analytical method to determine the dynamics of the vortex is introduced and the sound so produced is then related to the accelerations of the vortex. Though this flow configuration is not equivalent to a real flow duct, it is hoped that the present investigation can provide a picture for the generation mechanism of the self-generated noise and clues for further investigation into this controversial topic.

II. THE THEORIES

A. Vortex dynamics

This section introduces an analytical method to determine the associated vortex velocities, which is crucial in the computation of sound production in later sections. Figure 1 is the schematic diagram showing the flow configuration considered in the present study. The vortex possesses an anticlockwise circulation Γ and is initially at a transverse distance y_{2o} from the line $y_2=0$ which represents the flat surface. The boundary at $y_1<0$, $y_2=0$ is rigid, while the porous material is located at $y_1>0$, $y_2<0$. The junction between the

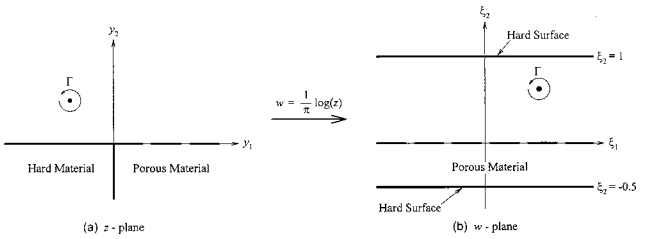


FIG. 1. The *z*- and *w* planes in the conformal mapping.

hard and porous materials is therefore at $y_1=0$, $y_2<0$ and the normal fluid velocity at this interface vanishes. The properties of the porous material are assumed to be characterized by the flow resistance R_f and the effective fluid density ρ_e inside its lattice.[12] For simplicity, m is introduced hereinafter to denote the ratio ρ_e/ρ where ρ is the fluid density in the unbounded medium. Also, the porous material is treated as a continuum as in Bear.[11] The equation governing the fluid flow inside the porous material is, according to Morse and Ingard[12]

$$\rho m \frac{\partial \mathbf{u}}{\partial t} + R_f \mathbf{u} = -\nabla p_h, \tag{1}$$

where $\mathbf{u}$ is the fluid velocity vector, p_h the hydrodynamic pressure, and t the time. It is assumed that the viscous effect of the fluid in the unbounded medium is negligible. However, since the tiny pores of the porous material are usually of very small dimensions, R_f cannot be neglected.

Direct calculation of the vortex dynamics from the flow boundary conditions specified in Fig. 1(a) is not easy because of the presence of the discontinuous flow boundary. The Wiener-Hopf method[13] is also not easy in this situation due to the anticipated unsteady vortex velocity. However, one can observe that the coordinate system in Fig. 1(a) can be transformed into a two-dimensional passage [Fig. 1(b)] through the use of complex variables.[14] Introducing $i=\sqrt{-1}$ and two complex planes z and w, where

$$z=y_1+iy_2 \quad \text{and} \quad w=\xi_1+i\xi_2,$$

the appropriate conformal mapping function for the transformation of the coordinate system in Fig. 1(a) to that in Fig. 1(b) is

$$z=f(w)=e^{\pi w}. \tag{2}$$

z is normalized by y_{2o}, w by the width of the two-dimensional passage. The hard surface backing the porous material in Fig. 1(b) is at $\xi_2=-0.5$. The properties of the porous material are unchanged during this transformation under the Cauchy-Riemann principle[14] (details are given in the Appendix). In the w plane, one can expect that the direction of the vortex velocity is parallel to the passage walls. Supposing this speed is V_w and assuming that it can be found, the velocity of the vortex at the point $z_\omega(=y_{1\omega}+iy_{2\omega})$ in the z plane V_z can then be obtained by using the Routh's correction[15]

$$V_z=\frac{1}{f'(w_\omega)}\left(V_w+\frac{\Gamma}{2\pi}\frac{f''(w_\omega)}{f'(w_\omega)}\right), \tag{3}$$

where $w_\omega=\xi_{1\omega}+i\xi_{2\omega}$ denotes the position of the image vortex in the w plane and $z_\omega=f(w_\omega)$. Though V_w is a real number as the velocity of the vortex in the w plane is parallel to the walls, V_z is complex, indicating that the vortex velocity in the z plane has a transverse component. Once V_z is found, a new z_ω can be found by using standard integration procedure. In the present study, the fourth-order Runge–Kutta method with adaptive step-size control was adopted. Then, a new w_ω and thus new V_w and V_z are calculated. The process is repeated and the vortex trajectory in the z plane can be obtained. Though the conformal mapping allows basically a one-to-one mapping between the z- and the w planes, one should note that the image vortex in the w plane is actually not a physically moving vortex. That is, the location of this vortex is determined solely by the mapping f, not by V_w directly such that its position cannot be determined by any time integration of V_w. Such a concept of using conformal mapping in the determination of vortex motions in the presence of complicated boundary geometry is well-known. Typical examples of conformal mapping application in fluid research include the works of Crighton,[16] Panaras,[17] and Conlisk and Rockwell.[18] Detailed description on the underlying theory can be found in Routh[15] or Woods.[19]

Assuming a stream function ψ for the fluid motion exists in the flow region $0<\xi_2<1$ in the w plane [Fig. 1(b)], one can write in the flow region

$$\nabla^2\psi=-\Gamma\,\delta(\xi_1-\xi_{1\omega})\,\delta(\xi_2-\xi_{2\omega}). \tag{4}$$

The target here is to determine the instantaneous V_w $(=\partial\xi_{1\omega}/\partial t)$. One should note that V_w depends only on $\xi_{2\omega}$ and the instantaneous vortex velocity in the ξ_2 direction is zero. Fourier transformation of Eq. (4) with respect to ξ_1 gives

$$\frac{\partial^2\psi_{\xi_1}}{\partial\xi_2^2}-k^2\psi_{\xi_1}=-\Gamma\,\delta(\xi_2-\xi_{2\omega})e^{ik\xi_{1\omega}}, \tag{5}$$

where ψ_{ξ_1} is the Fourier transform of

$$\psi\left(\psi_{\xi_1}=\int_{-\infty}^{\infty}\psi e^{ik\xi_1}\,d\xi_1\right).$$

It is straightforward to show that the corresponding transform of the stream function from an isolated vortex at $\xi_1=\xi_{1\omega}$, $\xi_2=\xi_{2\omega}$ is

$$\psi_\omega=\frac{\Gamma}{2|k|}e^{ik\xi_{1\omega}+|k|\xi_2-|k|\xi_{2\omega}}{}^{20}.$$

The velocity V_w can be obtained after performing an inverse Fourier transformation over k

$$V_w=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\partial}{\partial\xi_2}(\psi_{\xi_1}-\psi_\omega)\Big|_{\xi_2=\xi_{2\omega}}e^{-ik\xi_{1\omega}}\,dk. \tag{6}$$

Equations (5) and (6) and the fact that V_w does not depend on $\xi_{1\omega}$ together suggest that $e^{ik\xi_{1\omega}}$ is the only time-varying component in ψ_{ξ_1}.

Normal fluid velocities vanish on all the hard surfaces in the w plane [Fig. 1(b)]. Assuming another stream function ψ_p exists inside the porous material, as in Bear[11]

$$\left.\frac{\partial \psi_p}{\partial \xi_1}\right|_{\xi_2=-0.5}=0 \quad \text{and} \quad \left.\frac{\partial \psi}{\partial \xi_1}\right|_{\xi_2=1}=0. \tag{7}$$

Also,

$$\nabla^2 \psi_p = 0. \tag{8}$$

At the porous material interface $\xi_2=0$, and since the porous material is assumed to be a continuum, the hydrodynamic pressure and the normal fluid velocity are both continuous across this boundary

$$\rho m \frac{\partial}{\partial t}\left(\left.\frac{\partial \psi_p}{\partial \xi_2}\right|_{\xi_2=0}\right)+R_f\left.\frac{\partial \psi_p}{\partial \xi_2}\right|_{\xi_2=0}=\rho \frac{\partial}{\partial t}\left(\left.\frac{\partial \psi}{\partial \xi_2}\right|_{\xi_2=0}\right), \tag{9a}$$

and

$$\left.\frac{\partial \psi_p}{\partial \xi_1}\right|_{\xi_2=0}=\left.\frac{\partial \psi}{\partial \xi_1}\right|_{\xi_2=0}. \tag{9b}$$

By taking Fourier transform over ξ_1, it can be shown that both the transformed ψ_p and ψ take the form of $Ae^{-|k|\xi_2}+Be^{|k|\xi_2}$. The values of A and B in each region in Fig. 1(b) can be obtained by combining Eqs. (5), (7)–(9) and taking into account the vorticity jump across the vortex at $\xi_2=\xi_{2\omega}$.[20] Certainly, ψ is continuous in the region $0\leqslant\xi_2\leqslant 1$. It can be shown that

$$\begin{aligned}\psi_{\xi_1}&=\int_{-\infty}^{\infty}\psi e^{ik\xi_1}\,d\xi_1\\&=\frac{\Gamma}{2|k|}(e^{-|k|\xi_2}+Ge^{|k|\xi_2})\frac{e^{-|k|\xi_{2\omega}}-e^{|k|(\xi_{2\omega}-2)}}{G+e^{2-2|k|}}e^{ik\xi_{1\omega}},\end{aligned} \tag{10}$$

where $G=[ikV_w\rho+(ikV_wm\rho+R_f)\coth(|k|/2)]/[ikV_w\rho-(ikV_wm\rho+R_f)\coth(|k|/2)]$. The velocity of the vortex in the w plane can be obtained by taking the inverse Fourier transform over k

$$\begin{aligned}V_w&=\frac{1}{2\pi}\int_{-\infty}^{\infty}\left.\frac{\partial}{\partial \xi_2}(\psi_{\xi_1}-\psi_\omega)\right|_{\xi_2=\xi_{2\omega}}e^{-ik\xi_{1\omega}}\,dk\\&=-\frac{\Gamma}{4\pi}\int_{-\infty}^{\infty}\frac{Ge^{2|k|(\xi_{2\omega}-1)}+e^{-2|k|\xi_{2\omega}}}{G+e^{-2|k|}}\,dk.\end{aligned} \tag{11}$$

Since G contains V_w, Eq. (11) has to be solved by the iteration method in general. The velocity of the vortex in the z plane which contains the discontinuous boundary condition [Fig. 1(a)] can then be found using Eq. (3). The technique employed here is basically very similar to that of Panaras[17] and Conlisk and Rockwell,[18] but the present situation becomes more complicated due to the presence of the porous material which makes the vortex velocity in the w plane much less obvious. Time and all velocities presented in Sec. III are normalized by $4\pi y_{2o}^2/\Gamma$ and $\Gamma/4\pi y_{2o}$ respectively. Vortex accelerations are thus normalized by $\Gamma^2/16\pi^2 y_{2o}^3$.

B. Aeroacoustics

The formula that relates the production of sound to the unsteady low Mach vortex motions is due to Powell[5]

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial \tau^2}-\nabla^2 p=\nabla\cdot(\omega\times\mathbf{u}), \tag{12}$$

where ω is the vorticity, p the pressure fluctuation, c the ambient speed of sound, and τ the far-field observer time ($\tau=t+|\mathbf{x}|/c$). The coordinate system of Fig. 1(a) is adopted. In an unbounded medium, the far-field solution of Eq. (12) at a large distance $\mathbf{x}$ in the field field from the vortex filament is given by Lighthill[21]

$$p(\mathbf{x},\tau)=\frac{1}{4\pi c}\int\frac{x_i-y_i}{|\mathbf{x}-\mathbf{y}|^2}\frac{\partial}{\partial t}F_i(\mathbf{y},\tau-|\mathbf{x}-\mathbf{y}|/c)d\mathbf{y}, \tag{13}$$

where F_i is the ith component of the vortex force $(\Gamma\times\mathbf{u})$. In the two-dimensional case following the discussion in Tang and Ffowcs Williams,[8] Eq. (13) can be expressed as

$$p(\mathbf{x},\tau)=\frac{\Gamma}{4\pi c}\int_{-\infty}^{\infty}(-a_2\cos\theta+a_1\sin\theta)dy_3/|\mathbf{x}-\mathbf{y}|, \tag{14}$$

where a_i is the acceleration of the vortex in the y_i direction, θ is the azimuthal angle ($\tan\theta=x_2/x_1$), and y_3 is in a direction normal to paper in Fig. 1(a). The integrand is evaluated at the retarded time. The integration along y_3 can be transformed into a time integral as in Ffowcs Williams and Hawking[22] and be computed numerically as in Tang and Ko.[7] The form of Eq. (14) suggests that p has an amplitude that varies with $M^{2.5}$, where M is the characteristic Mach number of the problem. This is typical of two-dimensional vortex sound generation.[23]

The porous material offers an acoustic impedance z_a to the sound radiation. Though there are many models to accurately characterize the acoustical impedance of porous materials,[24,25] it is found that the simple model according to Morse[26] is sufficient for the present study. The model is

$$z_a=\rho c\sqrt{\frac{m+iR_f/2\pi\rho f_s}{\gamma\varepsilon}},$$

where f_s is the sound frequency, ε the porosity of the porous material, and γ the ratio of specific heat capacity. If the porous material is replaced by a nonpermeable liquid so that ε is zero, the surface of the liquid is acoustically hard. The far-field pressure fluctuation p is thus given by the summation of the contribution from the source in the unbounded medium $y_2>0$ and that from the acoustical image inside the materials. Thus, Eq. (14) becomes

$$p(\mathbf{x},\tau)=\frac{\Gamma}{2\pi c}\int_{-\infty}^{\infty}(-a_2\cos\theta+a_1\sin\theta)dy_3/|\mathbf{x}-\mathbf{y}|, \tag{15}$$

as the separation between the source and its image is infinitesimal when compared to the far-field distance $|\mathbf{x}|$ and the wavelength of the sound generated. Equation (15) suggests that the sound field is made up of two dipoles with time-dependent strengths. The longitudinal and transverse dipole

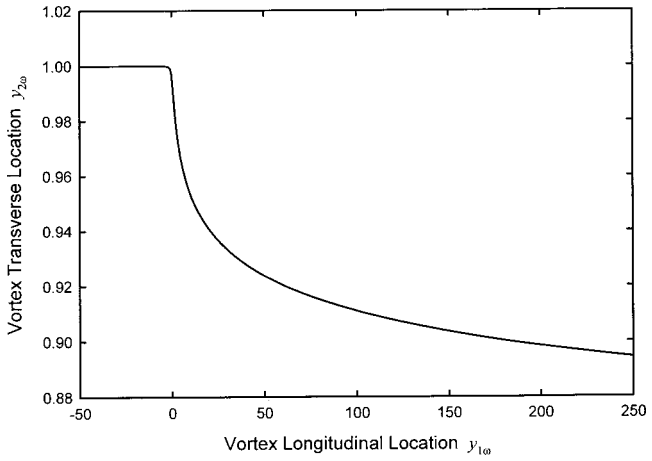


FIG. 2. Vortex flight path for $m=100$.

strengths, hereinafter denoted by D_1 and D_2, respectively, are therefore

$$D_1 = -\frac{\Gamma}{2\pi c}\int_{-\infty}^{\infty} a_2\, dy_3/|\mathbf{x}-\mathbf{y}|$$

and (16)

$$D_2 = \frac{\Gamma}{2\pi c}\int_{-\infty}^{\infty} a_1\, dy_3/|\mathbf{x}-\mathbf{y}|.$$

The resultant is also dipole but with a time-varying axis. Similar phenomena have been observed by Tang and Ffowcs Williams[8] and Tang and Ko.[7] Equation (15) also reveals explicitly that the accelerating motions of the vortex filament is the major source of sound in the present investigation.

If the vortex circulation Γ is weak, one can expect the frequency of the sound so produced is low and thus the acoustic impedance of the porous surface is therefore very large. As it will be shown later that the source strengths are of low frequency, this impedance is taken to be infinite for simplicity in the present investigation so that the sound field is again given by Eq. (15) even in the presence of the porous material.

III. RESULTS AND DISCUSSIONS

A. Hard surface/liquid

When the porous material in Fig. 1(a) is replaced by a liquid, which is immiscible with the fluid in which the vortex is propagating, the porosity ε is so small that the liquid surface becomes hard to acoustics. Then, the discontinuity of the flow boundary will only affect the flight path of the vortex, but not the propagation of sound to the far field.

Figure 2 shows the flight path of the vortex for $m=100$ with initial vortex position sufficiently upstream of the plane of discontinuity $y_1=0$. The vortex moves gradually towards the flat surface. Relatively large change in the vortex velocity is observed at the instant the vortex moves across $y_1=0$, suggesting significant sound generation.[5,7] The increase in m results in less significant change in the transverse position and the velocity of the vortex. This may imply a weaker generation of sound. Figure 3(a) illustrates the time variations of the longitudinal velocities of the vortex for $m=100$ and 1000. Those of the corresponding transverse velocities are shown in Fig. 3(b). Hereinafter, t_o represents the instant the vortex moves across the plane of discontinuity $y_1=0$. Figure 3 suggests that the vortex undergoes rapid changes in both its longitudinal and transverse velocities when it is close to the plane of boundary discontinuity. The larger the value of m, the less significant these changes will be. The time variations of the corresponding vortex accelerations for $m=100$, which are also proportional to the source strengths [Eq. (15)], are shown in Fig. 4.

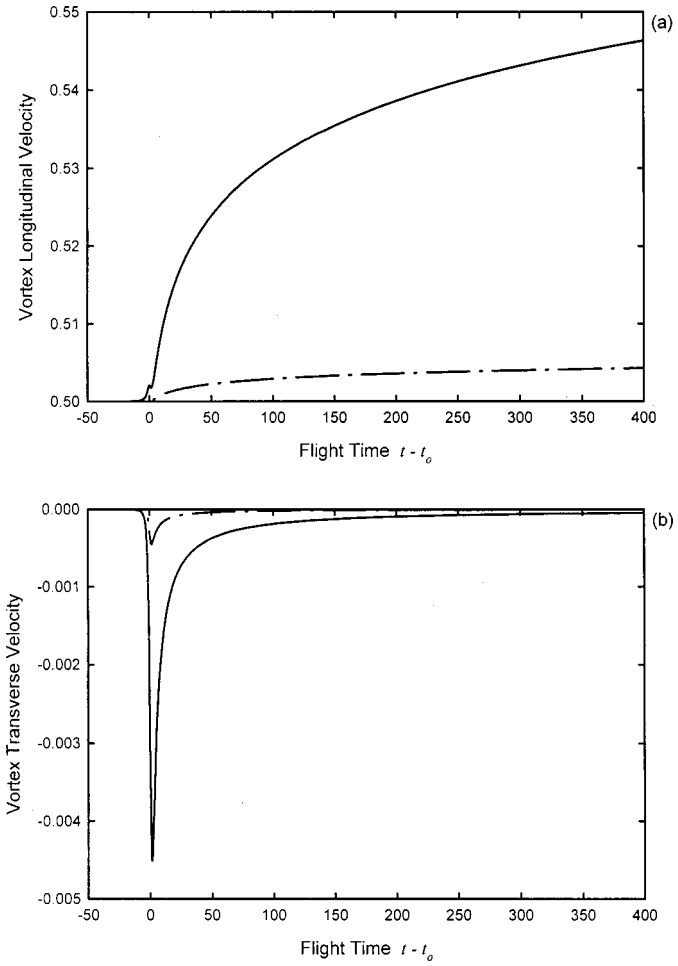


FIG. 3. Time variations of vortex velocities. (a) Longitudinal velocity; (b) transverse velocity. ———: $m=100$; $-\cdot-$: $m=1000$.

The time variations of the strengths of the two dipole sound fields, D_1 and D_2, are obtained by integrating the source strengths along the length of the vortex filament as depicted in Eq. (16). Results for $m=100$ are illustrated in Fig. 5. The increase in m lowers the dipole strengths but their time variations very much resemble those shown in Fig. 5 and thus are not presented. The transverse dipole sound so produced is basically of low frequency except at zero retarded time, which corresponds to the time when the vortex moves across the discontinuity plane. The longitudinal dipole sound is pulse-like and vanishes relatively quickly at $t > t_o$. The resultant far field will become a transverse dipole when the vortex is sufficiently distant from the plane of discontinuity $y_1=0$. However, the frequency of the sound is then very low.

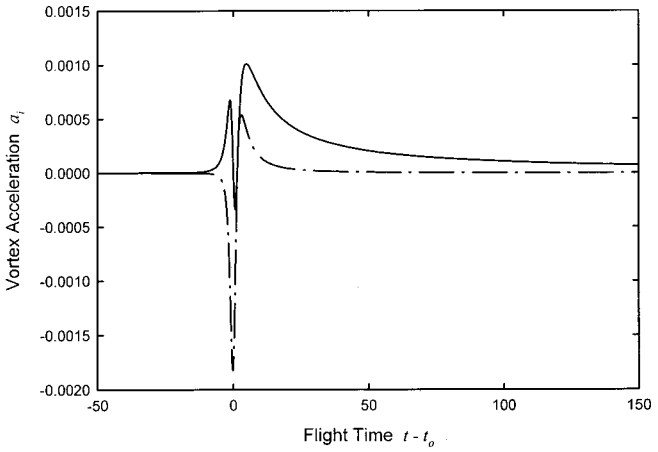


FIG. 4. Time variations of vortex accelerations. ———: longitudinal; –·–: transverse. $m=100$.

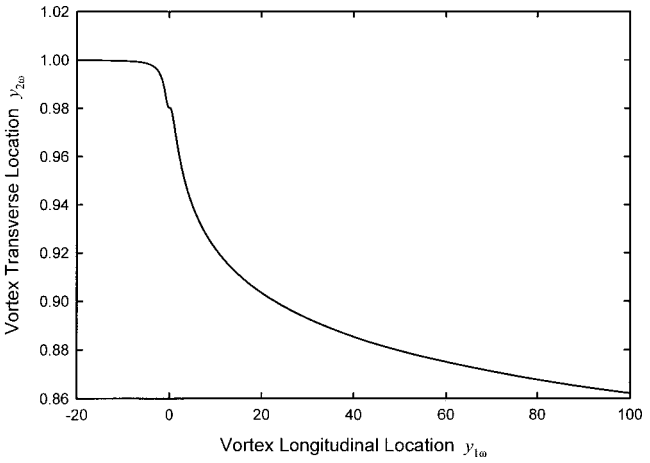


FIG. 6. Vortex flight path in the presence of a porous material. $R_f=10$, $m=3$.

B. Hard surface/porous material

The investigation in this section will be focused on the cases where $1<m<5$ as it is rare to have $m\geqslant 5$ or $m\leqslant 1$ in reality.[12] The flow resistance R_f and the sound frequency f_s hereinafter are normalized by $\rho\Gamma/4\pi y_{2o}^2$ and $\Gamma/2y_{2o}^2$, respectively. Figure 6 is a typical example showing that the vortex moves towards the surface of the porous material as it propagates downstream from the region with negative y_1($m=3$,$R_f=10$). Though the vortex flight path appears quite similar to those in the previous liquid case (Fig. 2), a ''kick'' is observed in Fig. 6 at $y_{1\omega}\approx 0$, suggesting that the vortex actually undergoes a series of accelerations and decelerations close to the instant when it crosses the plane of discontinuity. The increase in R_f or m reduces the magnitudes of the vortex acceleration/deceleration as illustrated in Figs. 7(a) and (b). The corresponding data for $R_f=100$ are shifted by 10 positive time units in order to avoid serious overlapping of data at $t\approx t_o$. At large m, the time span of significant acceleration variation increases due to a lower vortex propagation speed for $t>t_o$ as shown in Fig. 3. Results shown in Fig. 7 imply weaker sound radiation at increased magnitude of boundary impedance.

The time-variation patterns of the vortex accelerations and decelerations imply that the frequency of the sound so produced at or close to the instant when the vortex crosses the plane of discontinuity is higher than those in the previous section, where the porous material is substituted by a liquid. However, the source strengths, which are the rates of change of the vortex forces, are in general of very low frequency

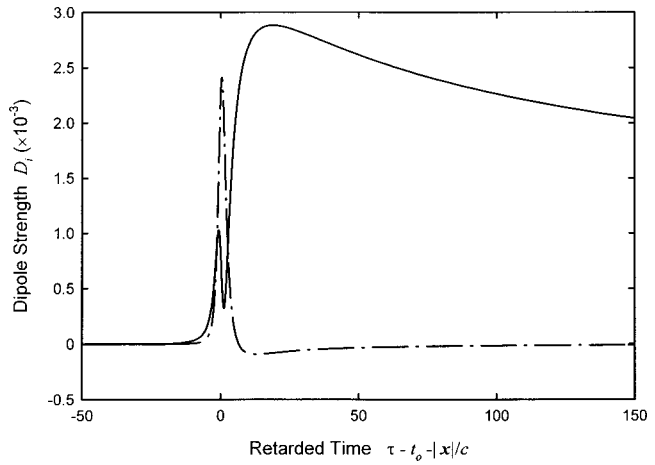


FIG. 5. Time variations of dipole strengths. ———: transverse; –·–: longitudinal. $m=100$.

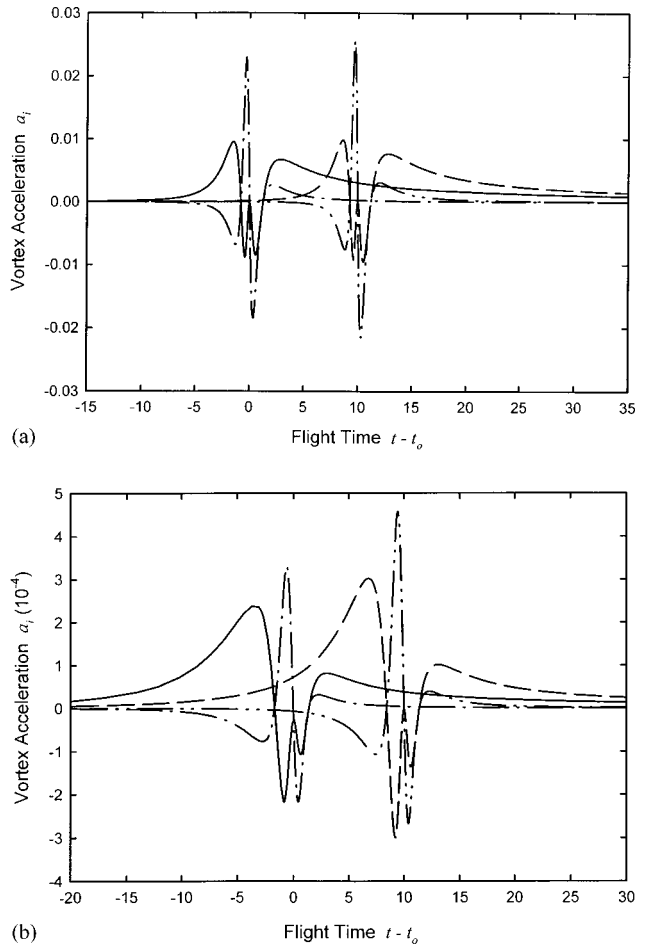


FIG. 7. Effects of relative fluid density and flow resistance on the time variations of vortex accelerations. (a) $R_f=10$; (b) $R_f=100$. ———: longitudinal acceleration, $m=3$; –·–: transverse acceleration, $m=3$; ---: longitudinal acceleration, $m=4$; –···–: transverse acceleration, $m=4$.

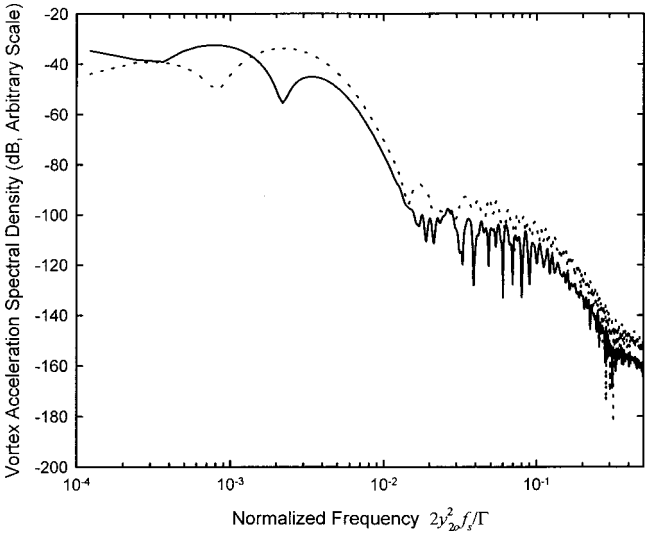


FIG. 8. Frequency spectra of source strengths. ———: Longitudinal dipole; ···: transverse dipole.

such that the hard surface approximation for the far field depicted by Eq. (15) works well in the present situation. A typical example of the vortex acceleration spectra, which also represents those of the source strengths as Γ is constant, is shown in Fig. 8 ($m=3, R_f=10$). The effect of the impedance surface is therefore mainly on the dynamics of the vortex filament rather than on the modification of sound propagation.

Figures 9(a) and (b) illustrate the time variations of the strengths of the longitudinal and transverse dipoles, respectively, at $R_f=10$, $2\leqslant m\leqslant 4$. The corresponding results at $R_f=100$ are shown in Fig. 10. It is observed that the magnitudes of the far-field sound increases with m but decreases with increasing flow resistance R_f. Since the variation of m is limited in reality, the far-field sound, especially the transverse dipole strength, depends very much on R_f. The majority of the sound energy is radiated close to the instant when the vortex filament moves across the plane of impedance discontinuity as expected. Though the magnitude of the transverse dipole sound D_2 does not die down quickly after this instant [Figs. 9(b) and 10(b)], its very low-frequency content renders it hard to be heard shortly after the vortex moves over the impedance discontinuity.

IV. CONCLUSIONS

The sound generated by a vortex filament moving over a boundary with a sharp flow impedance change is investigated theoretically by using the potential theory and vortex

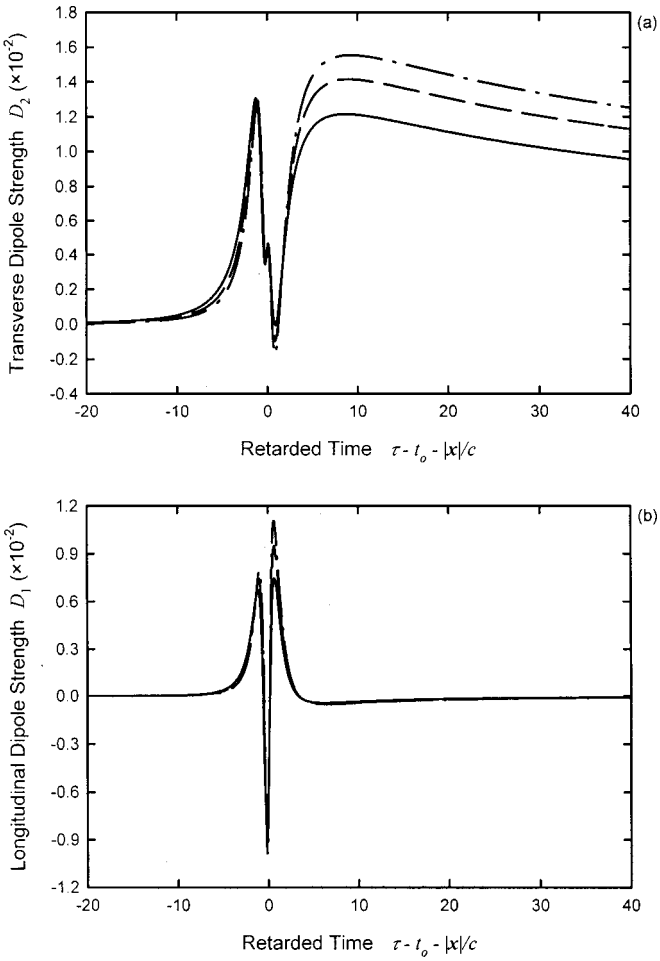


FIG. 9. Effects of porous material on time variations of dipole strengths for $R_f=10$. (a) Transverse dipole; (b) longitudinal dipole. ———: $m=2$; --: $m=3$; -·-: $m=4$.

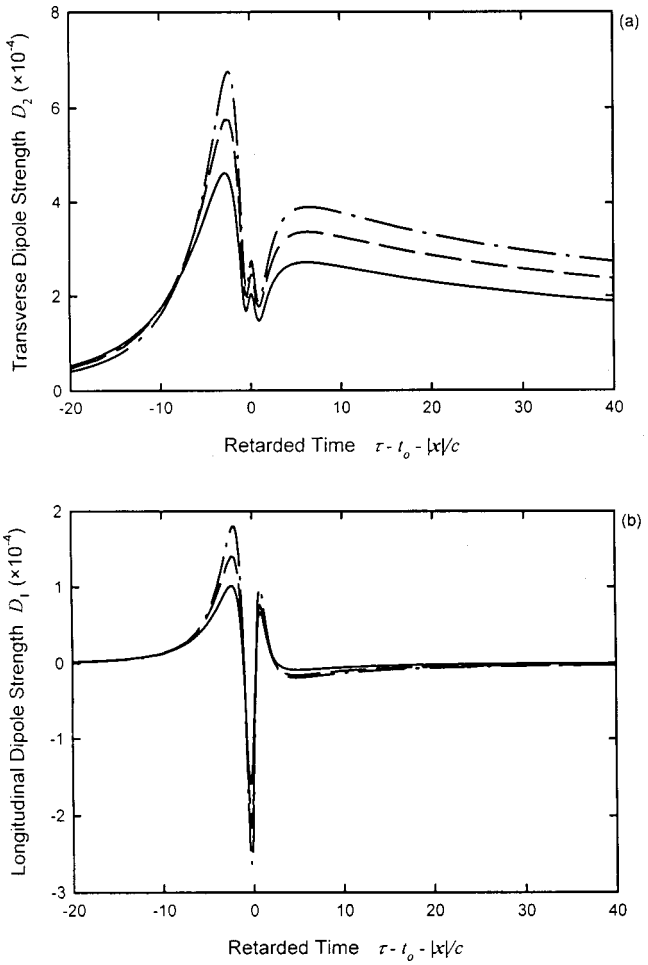


FIG. 10. Effects of porous material on time variations of dipole strengths for $R_f=100$. (a) Transverse dipole; (b) longitudinal dipole. ———: $m=2$; --: $m=3$; -·-: $m=4$.

sound theory. Both the flight paths of the vortex filament and the corresponding far-field sound pressure under different flow impedance changes are computed. The roles of the flow resistance and relative fluid density inside the boundary material lattice are also discussed.

The theory shows that sound is generated mainly through the accelerating motions of the vortex filament. The resultant far-field sound is dipole. Its magnitude varies with the two-and-a-half power of the characteristic velocity and thus, the vortex circulation, as in other two-dimensional vortex sound cases. The majority of the sound energy is radiated at the instant when the vortex filament moves across the plane flow impedance discontinuity. This instant is also the one at which the vortex filament undergoes substantial longitudinal and transverse acceleration. In the case where the boundary is made up of a hard solid and a liquid, the strength of the final dipole decreases as the density of the liquid increases. When the liquid is replaced by a porous material, the corresponding dipole strength decreases as the flow resistance inside the porous material increases. However, the opposite is observed when the relative fluid density in the porous material is concerned.

The source strengths, which are the vortex forces and thus the accelerations, are of very low frequency in the present investigation. The flow impedance change on the flow boundary therefore affects the propagation path of the vortex filament and thus the sound generation process only. The consequent sound propagation to the far field is not affected.

ACKNOWLEDGMENTS

S.K.T. would like to express his gratitude to Dr. N. Peake of the Department of Applied Mathematics and Theoretical Physics, The University of Cambridge, for his advices in solving the vortex dynamics. The financial support from the Hong Kong Polytechnic University and NAP Acoustics (Far East) Ltd. is gratefully acknowledged.

APPENDIX: POROUS REGION UNDER CONFORMAL MAPPING

The flow equation in the unbounded medium $y_2>0$ in Fig. 1(a) is not altered by the conformal mapping given by Eq. (2) as far as potential flow is concerned. Inside the porous material, the flow is governed by Eq. (1). Since $z=f(w)$, one obtains for any continuous quantity K that

$$\frac{\partial K}{\partial y_1}=\frac{\partial K}{\partial \xi_1}\frac{\partial \xi_1}{\partial y_1}+\frac{\partial K}{\partial \xi_2}\frac{\partial \xi_2}{\partial y_1}$$

and (A1)

$$\frac{\partial K}{\partial y_2}=\frac{\partial K}{\partial \xi_1}\frac{\partial \xi_1}{\partial y_2}+\frac{\partial K}{\partial \xi_2}\frac{\partial \xi_2}{\partial y_2}.$$

Recalling Eq. (1) and after resolving quantities in the two directions, one obtains

$$\rho m\frac{\partial}{\partial t}\frac{\partial \psi_p}{\partial y_1}+R_f\frac{\partial \psi_p}{\partial y_1}=\frac{\partial p_h}{\partial y_2}$$

and (A2)

$$\rho m\frac{\partial}{\partial t}\frac{\partial \psi_p}{\partial y_2}+R_f\frac{\partial \psi_p}{\partial y_2}=-\frac{\partial p_h}{\partial y_1}.$$

Therefore,

$$\rho m\frac{\partial}{\partial t}\left(\frac{\partial \psi_p}{\partial \xi_1}\frac{\partial \xi_1}{\partial y_1}+\frac{\partial \psi_p}{\partial \xi_2}\frac{\partial \xi_2}{\partial y_1}\right)+R_f\left(\frac{\partial \psi_p}{\partial \xi_1}\frac{\partial \xi_1}{\partial y_1}+\frac{\partial \psi_p}{\partial \xi_2}\frac{\partial \xi_2}{\partial y_1}\right)=\frac{\partial p_h}{\partial \xi_1}\frac{\partial \xi_1}{\partial y_2}+\frac{\partial p_h}{\partial \xi_2}\frac{\partial \xi_2}{\partial y_2}\quad\text{(A3a)}$$

and

$$\rho m\frac{\partial}{\partial t}\left(\frac{\partial \psi_p}{\partial \xi_1}\frac{\partial \xi_1}{\partial y_2}+\frac{\partial \psi_p}{\partial \xi_2}\frac{\partial \xi_2}{\partial y_2}\right)+R_f\left(\frac{\partial \psi_p}{\partial \xi_1}\frac{\partial \xi_1}{\partial y_2}+\frac{\partial \psi_p}{\partial \xi_2}\frac{\partial \xi_2}{\partial y_2}\right)=-\frac{\partial p_h}{\partial \xi_1}\frac{\partial \xi_1}{\partial y_1}-\frac{\partial p_h}{\partial \xi_2}\frac{\partial \xi_2}{\partial y_1}.\quad\text{(A3b)}$$

The Cauchy-Riemann principle[14] states that for the adopted conformal mapping

$$\frac{\partial \xi_1}{\partial y_1}=\frac{\partial \xi_2}{\partial y_2}\ \text{ and }\ \frac{\partial \xi_1}{\partial y_2}=-\frac{\partial \xi_2}{\partial y_1},$$

and it is straightforward to show by adding up Eqs. (A3a) and (A3b) that

$$\left(\rho m\frac{\partial}{\partial t}\frac{\partial \psi_p}{\partial \xi_1}+R_f\frac{\partial \psi_p}{\partial \xi_1}-\frac{\partial p_h}{\partial \xi_2}\right)J_1+\left(\rho m\frac{\partial}{\partial t}\frac{\partial \psi_p}{\partial \xi_2}+R_f\frac{\partial \psi_p}{\partial \xi_2}+\frac{\partial p_h}{\partial \xi_1}\right)J_2=0,\quad\text{(A4)}$$

where $J_1=(\partial \xi_1/\partial y_1)-(\partial \xi_2/\partial y_1)$ and $J_2=(\partial \xi_2/\partial y_1)+(\partial \xi_1/\partial y_1)$. Since Eq. (A4) is valid at any point in the porous material in the w plane and J_1 and J_2 are in general linear independent, the equations governing the flow in the porous material in the w plane are

$$\rho m\frac{\partial}{\partial t}\frac{\partial \psi_p}{\partial \xi_1}+R_f\frac{\partial \psi_p}{\partial \xi_1}=\frac{\partial p_h}{\partial \xi_2}$$

and (A5)

$$\rho m\frac{\partial}{\partial t}\frac{\partial \psi_p}{\partial \xi_2}+R_f\frac{\partial \psi_p}{\partial \xi_2}=-\frac{\partial p_h}{\partial \xi_1},$$

which are exactly Eq. (A2) with y_1 and y_2 replaced by ξ_1 and ξ_2, respectively.

[1] L. L. Beranek, *Noise and Vibration Control Engineering, Principles and Applications* (Wiley, New York, 1992).
[2] N. Curle, ''The influence of solid boundary upon aerodynamic noise,'' Proc. R. Soc. London, Ser. A **231**, 505–514 (1955).
[3] D. Küchemann, ''Report on I.U.T.A.M. Symposium on concentrated vortex motions in fluids,'' J. Fluid Mech. **21**, 1–20 (1965).
[4] L. M. Milne-Thomson, *Theoretical Hydrodynamics* (Macmillan, London, 1968).
[5] A. Powell, ''Vortex sound theory,'' J. Acoust. Soc. Am. **36**, 177–195 (1964).
[6] F. Obermeier, ''The influence of solid bodies on low Mach number vortex sound,'' J. Sound Vib. **72**, 39–49 (1980).
[7] S. K. Tang and N. W. M. Ko, ''Sound generation by interaction of two inviscid two-dimensional vortices,'' J. Acoust. Soc. Am. **102**, 1463–1473 (1997).
[8] S. K. Tang and J. E. Ffowcs Williams, ''Acoustic radiation from a vortex

approaching a circular cylinder with surface suction,'' Acustica **84**, 1007–1013 (1998).
[9] M. C. Quinn and M. S. Howe, ''On the production and absorption of sound by lossless liners in the presence of mean flow,'' J. Sound Vib. **97**, 1–9 (1984).
[10] J. E. Ffowcs Williams, ''The acoustics of turbulence near sound-absorbent liners,'' J. Fluid Mech. **51**, 737–749 (1972).
[11] J. Bear, *Dynamics of Fluids in Porous Media* (Dover, New York, 1972).
[12] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[13] N. Noble, *Methods Based on the Wiener-Hopf Technique* (Chelsea, New York, 1988).
[14] R. V. Churchill and J. W. Brown, *Complex Variables and Applications* (McGraw-Hill, Singapore, 1984).
[15] E. J. Routh, ''Some applications of conjugate functions,'' Proc. London Math. Soc. **12**, 73–89 (1881).
[16] D. G. Crighton, ''Radiation from vortex filament motion near a half plane,'' J. Fluid Mech. **51**, 357–362 (1972).
[17] A. G. Panaras, ''Pressure pulses generated by the interaction of a discrete vortex with an edge,'' J. Fluid Mech. **154**, 445–461 (1985).
[18] A. T. Conlisk and D. Rockwell, ''Modeling of vortex corner interaction using point vortices,'' Phys. Fluids **24**, 2133–2142 (1981).
[19] L. C. Woods, *The Theory of Subsonic Plane Flow* (Cambridge University Press, Cambridge, 1961).
[20] N. Peake, Private communication, 1995.
[21] M. J. Lighthill, ''On sound generated aerodynamically. I. General theory,'' Proc. R. Soc. London, Ser. A **211**, 564–587 (1952).
[22] J. E. Ffowcs Williams and D. L. Hawkings, ''Shallow water wave generation by unsteady flow,'' J. Fluid Mech. **31**, 779–788 (1968).
[23] J. E. Ffowcs Williams, ''Hydrodynamic noise,'' Annu. Rev. Fluid Mech. **1**, 197–222 (1969).
[24] K. Attenborough, ''Ground parameter information for propagation modeling,'' J. Acoust. Soc. Am. **92**, 418–427 (1992); see also R. Raspet and K. Attenborough, ''Erratum: Ground parameter information for propagation modeling,'' J. Acoust. Soc. Am. **92**, 3007 (1992).
[25] J. F. Allard, *Propagation of Sound in Porous Media: Modeling Sound Absorbing Materials* (Chapman & Hall, London, 1993).
[26] P. M. Morse, *Vibration and Sound* (The American Institute of Physics, New York, 1976).

A computational approach for flow–acoustic coupling in closed side branches

Paul M. Radavich and Ahmet Selamet[a)]
The Ohio State University, Department of Mechanical Engineering and Center for Automotive Research, 206 West 18th Avenue, Columbus, Ohio 43210

James M. Novak
Ford Motor Company, Powertrain Operations, Dearborn, Michigan 48121

(Received 23 October 1999; revised 22 December 2000; accepted 27 December 2000)

The quarter-wave resonator, which produces a narrow band of high acoustic attenuation at regularly spaced frequency intervals, is a common type of silencer used in ducts. The presence of mean flow in the main duct, however, is likely to promote an interaction between these acoustic resonances and the flow. The coupling for some discrete flow conditions leads to the production of both large wave amplitudes in the side branch and high noise levels in the main duct, thereby transforming the quarter-wave silencer into a noise generator. The present approach employs computational fluid dynamics (CFD) to model this complex interaction between the flow and acoustic resonances at low Mach number by solving the unsteady, turbulent, and compressible Navier–Stokes equations. Comparisons between the present computations and the experiments of Ziada [PVP-Vol. 258, ASME, 35–59 (1993)] for a system with two coaxial side branches show that the method is capable of reproducing the physics of the flow–acoustic coupling and predicting the flow conditions when the coupling occurs. The theory of Howe [IMA J. Appl. Math. **32**, 187–209 (1984)] is then employed to determine the location and timing of the acoustic power production during a cycle.
 [DOI: 10.1121/1.1350618]

PACS numbers: 43.28.Ra, 43.28.Py [LCS]

LIST OF SYMBOLS

c	speed of sound
d	side branch length in flow direction
f	frequency
h	height of main duct
L	side branch length perpendicular to flow
M	U/c, Mach number
P_{rms}	root mean squared acoustic pressure
St	fd/U, Strouhal number
T	$1/f$, acoustic period
$\mathbf{u}$	total velocity vector minus acoustic velocity
$\mathbf{u}'$	acoustic velocity vector
U	average mean flow velocity in main duct

Greek symbols

Π	acoustic sound power
ρ_0	mean density
ω	vorticity vector

I. INTRODUCTION

Quarter-wave resonators are used as acoustic silencers in numerous applications. This well-known silencer produces a large acoustic attenuation at frequencies where the length of the side branch is an odd multiple of a quarter wavelength, ($f=c/4L$, $3c/4L$, $5c/4L,\ldots$, with c being the speed of sound and L the length of the side branch). In the presence of mean flow, however, a shear layer is created between the moving fluid in the main duct and the stationary fluid in the side branch. Under certain flow conditions, instability in the shear layer creates oscillations, which can then go on to excite acoustic resonances in the side branch. The acoustic resonances then amplify the oscillations in the shear layer, and the whole process continues to amplify until large-amplitude vortices are formed. This coupling can lead to the production of both large wave amplitudes in the side branch and high noise levels in the main duct. Thus, this interaction causes the quarter-wave resonator to become a noise source rather than an acoustic silencer.

Numerous works have been conducted on the noise and high-pressure amplitudes generated by flow over rectangular cavities. These configurations can be categorized into two groups: deep cavities and shallow cavities, as illustrated in Fig. 1. The deep cavities have a length to diameter ratio, L/d, greater than 1, and include the side branches of the current investigation. In these cavities, flow-induced noise is produced primarily when oscillations in the shear layer create waves which travel along the length of the cavity, L, and reflect back to interact again with the shear layer. These long side branches are commonly investigated both in confined flows, where h is a finite dimension, and in half-plane flows where h is infinite. The shallow cavities, where L/d is less than or equal to 1, are similar in nature to deep cavities and share many common properties. For these short cavities, however, the stronger acoustic interactions take place over

a)Electronic mail: selamet.1@osu.edu

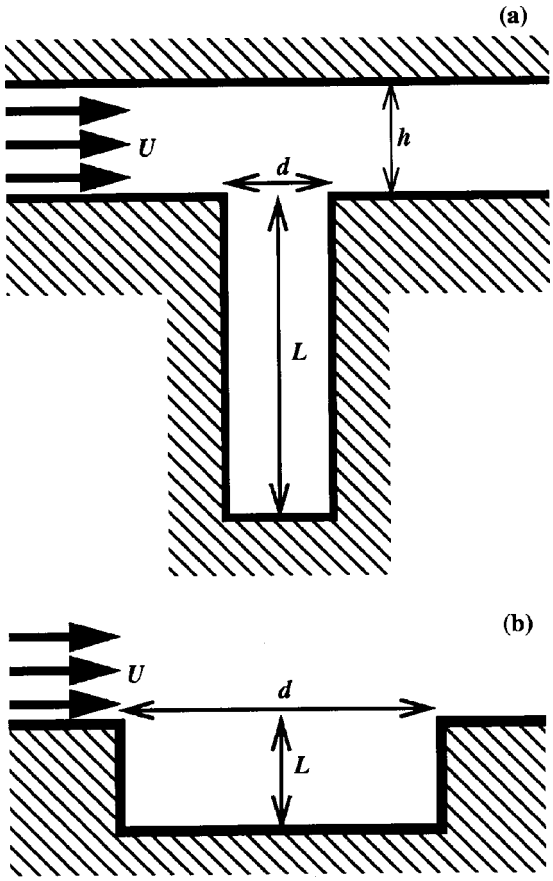


FIG. 1. Geometry for (a) deep cavities or side branches ($L>d$) and (b) shallow cavities ($L<d$).

the longer dimension of the cavity d. The shallow cavities have typically been investigated in half-plane flows. Detailed reviews of flow noise in cavities can be found, for example, in Naudascher (1967), Rockwell and Naudascher (1978), and Komerath *et al.* (1987).

Some of the early experimental work on deep cavities was performed by East (1966), who examined low Mach number flows over a two-dimensional (rectangular cross section) unconfined cavity. East showed that these deep cavities excite primarily discrete frequencies that occur near the fundamental acoustic resonance frequency of the side branch, and concluded that the tones are produced when oscillations are amplified by coupling between the shear layer fluctuations and the cavity acoustic modes. He also noted that the peak excitation occurred in two discrete ranges of Strouhal number $St=fd/U=0.3-0.4$ and $St=0.6-0.9$, suggesting two different modes of shear layer excitation. Visualization studies by Erickson *et al.* (1986), Erickson and Durgin (1987), and Ziada (1993) showed that the excitation near $St=0.4$ is characterized by a single vortex in the branch mouth, while the excitation near $St=0.8$ is characterized by two simultaneous vortices in the branch mouth. Later experimental work by Jungowski *et al.* (1987, 1989) for a circular side branch mounted to a circular duct showed excitation in similar ranges of Strouhal number.

Shallow cavities with $L/d<1$, as shown in Fig. 1(b), also exhibit flow–acoustic coupling similar in nature to that of the deep cavities. Several experimental works documenting the properties of these shallow cavities include Krishnamurty (1955), Ball (1959), Dunham (1962), Heller and Bliss (1975), Franke and Carr (1975), Shaw (1979), and Sarno and Franke (1990). These studies have shown that, with the long length of the cavity in the flow direction, as many as five different shear layer modes can be excited. The Strouhal number of the first vortex mode excitation occurs near 0.4, which is similar to that of deep cavities. Helmholtz resonators, which consist of a cavity connected to the main flow duct through a smaller orifice opening, also have distinct acoustic resonances and have shown excitation properties similar to the side branches (De Metz and Farabee, 1977; Anderson, 1977; Hersh *et al.*, 1978; Panton, 1988).

Due to the complexity of the problem with unsteady, nonlinear, viscous, compressible, and turbulent flow, the formulation of analytical methods becomes difficult and involves numerous assumptions. Covert (1970) used linear stability theory to couple mean flow with oscillations in a cavity. For shallow cavities, Tam and Block (1978) added the shear layer momentum thickness and acoustic wave propagation and reflection in the cavity to their linear model. They demonstrated that the shear layer momentum thickness and the length to depth ratio of the cavity were important in determining how the Strouhal number at excitation varies with Mach number. With estimated values for the momentum thickness, their results correlated well with experiments in determining the Strouhal number for excitation. Their acoustic model did not allow for wave reflections from the open end of the cavity, however; thus, this method would require modification to allow for cavity normal-mode resonances. Howe (1981, 1997) used linear theory to model a cavity connected to the mean flow through a small aperture. His predictions of excitation velocity also correlated well with experimental values. More complicated models by Bruggeman *et al.* (1991) and Durgin and Graf (1992) introduce concentrated vortices into the flow and are able to predict acoustic pressure amplitudes for cavity flows. These vortex models perform well in providing qualitative predictions of changes in the acoustic amplitudes, although the predicted amplitudes exceed experimental values by a factor of 4 in Bruggeman *et al.* and 5 in Durgin and Graf. In another vortex model by Kriesels *et al.* (1995), the interaction between acoustic waves and distributed vortex ''blobs'' was studied at the junction between a duct and a deep side branch. Comparisons between Schlieren visualization and computational predictions of the vortex growth and propagation using this method were excellent. The model requires that acoustic boundary oscillations be specified from measured values, however, and thus does not reproduce self-sustaining oscillations. Several authors including Nelson *et al.* (1983), Bruggeman *et al.* (1991), Durgin and Graf (1992), and Kriesels *et al.* (1995) have combined their vortex models with the theory of Howe (1984) to examine when and where in the acoustic cycle sound power is produced.

Recently, attempts have been made to use conventional CFD methods to solve for the acoustic field generated by flow over a shallow cavity. The majority of these works considers high Mach number flows greater than 0.95, except for Hardin and Pope (1995), who examine a low Mach number

of 0.1. Hardin and Pope used an uncoupled approach in which the acoustic field is solved from incompressible and laminar flow-field results. This method does not allow, however, for coupling or feedback between the acoustic and flow fields. The works of Hankey and Shang (1979), Rizzetta (1988), Baysal *et al.* (1988, 1990, 1994), Kim and Chokani (1990), Chokani and Kim (1991), and Tam *et al.* (1996) treat high Mach number flows by solving the compressible Navier–Stokes equations with a simple algebraic turbulence closure model. These works show that the computational method is adequate for predicting the properties of the mean flow field. Rizzetta (1988) also demonstrates the ability to predict the generated acoustic noise. The first three resonance frequencies of excitation are predicted in the cavity, although the pressure amplitudes are overpredicted across the frequency spectrum. A most recent work by Lilley *et al.* (1997) uses a large eddy simulation to close the turbulence equations for flow over a shallow cavity. No results concerning the acoustic pressure field are provided in this work, however.

The present approach solves the compressible Navier–Stokes equations with a turbulence closure model, similar to some of the previous computational investigations of flow over shallow cavities. Rather than focusing on modeling the mean flow field, however, the objective of the present work is to accurately reproduce the interaction between the acoustic waves in the cavity and the vortices that form in the shear layer between the flow duct and the cavity. While other researchers have studied similar shallow cavities at high Mach number using this method, the present work considers deep side branches which are capable of producing nonlinear wave amplitudes from low Mach number flow. As indicated earlier, other authors have used various types of simplified models to study the same interactions in deep side branches. Many of the assumptions and limitations of these models can be overcome using the present approach, however. By solving the full Navier–Stokes equations, the nonlinearities in the acoustic waves can be properly modeled, thereby circumventing the constraint of linear wave amplitude treatments. Also, rather than approximating the vortex strength, location, and path as it travels across the mouth of the side branch, as is done in many of the vortex methods, the present approach allows the vortex to develop naturally and interact with both the acoustic waves and the flow field as it travels over the side branch. Many of the analytical and vortex models also use the assumption of incompressible flow for low Mach numbers, which requires that the acoustic waves be introduced externally from experimental measurements. Such approach does not allow for a true interaction between the vortices and the acoustic field, as the vortices are unable to alter the fixed acoustic amplitude. Thus, this simplification makes prediction of variations of the acoustic pressure amplitude at different flow velocities impossible. By solving the compressible flow equations with the current method, the vortices and the acoustic waves are allowed to interact, with each affecting the other until a natural equilibrium is reached. This enables the present work to determine both when the flow–acoustic coupling will occur and how the acoustic pressure varies as the mean flow velocity and various duct dimensions are changed. The present computational approach also has the potential to include irregular-shaped geometry which would be difficult to include in many of the foregoing analytical models. Several authors, including Dunham (1962), Franke and Carr (1975), Jungowski *et al.* (1987, 1989), and Bruggeman *et al.* (1991) have shown that small changes in the geometry of the interface between the flow field and the cavity, such as the addition of a small ramp or radius, can help to reduce the acoustic pressure amplitudes or eliminate the coupling altogether. Finally, the present approach allows for reflections from the inlet and outlet boundaries of the main duct. Bruggeman *et al.* (1986) demonstrated experimentally that when the side branch is connected to another duct, as shown in Fig. 1(a), the reflections from the inlet and outlet ducts can have a major impact on the acoustic pressure amplitudes produced during the flow–acoustic coupling. They emphasized that the acoustic properties of the entire system, not just the side branch itself, need to be accounted for, and were able to reduce acoustic amplitudes by 20 dB by altering the termination length. This effect has been largely overlooked elsewhere in the literature, however.

The objective of the present study is to assess the feasibility of using computational fluid dynamics to solve flow–acoustic coupling problems in deep cavities at low Mach numbers <0.1. A two-dimensional investigation has been performed comparing the numerical predictions to experimental results obtained by Ziada (1993). His experimental configuration consists of a main duct with two side branches positioned at the same duct location opposed from each other. Results are provided for the acoustic pressure amplitude versus Strouhal number, as well as smoke visualization of the interaction between the main duct and a side branch. The detailed computations also allow the theory of Howe (1984) to be applied to the flow field to approximate the acoustic power produced by the flow–acoustic interaction. Following this Introduction, a brief description of the computational approach is given in Sec. II. A comparison of the computations with the experimental results of Ziada is presented in Sec. III, along with a detailed investigation into how acoustic power is produced by the coupling. Finally, results and concluding remarks are given in Sec. IV.

II. METHODOLOGY

The effect of acoustic propagation and bulk fluid flow can be obtained simultaneously if the unsteady viscous compressible flow equations are solved with sufficient accuracy. In the present problem, the acoustic waves and the mean flow are strongly coupled, with oscillations in the mean flow exciting acoustic waves which in turn feed energy back into the mean flow oscillations. This requires an accurate resolution of both acoustic waves and fluid flow in the time domain. To ensure accurate resolution of the acoustic waves, the effects of time-step size and grid spacing were investigated for a single side-branch configuration similar in dimension to the double side-branch configuration used for the present study. With the current resolution, acoustic amplitudes are underpredicted within about 2–3 dB at the quarter-wave resonance frequency of the side branch due to numeri-

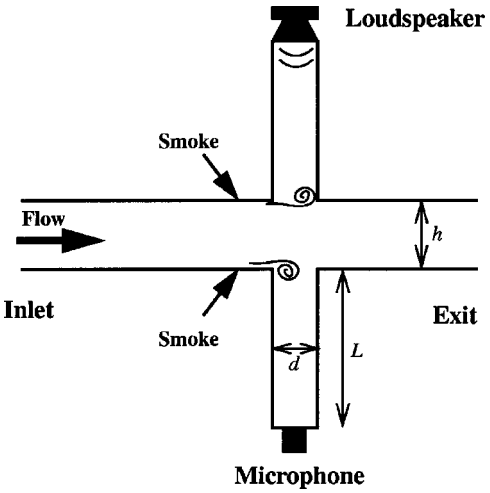


FIG. 2. Geometry for the experiments of Ziada (1993).

cal dissipation. Better accuracy can be obtained by increasing the temporal and spatial resolution further; however, the higher accuracy is not justified due to the excessive increase in computational time. The present computational approach solves the unsteady flow equations using the PISO algorithm (Issa, 1986; Issa *et al.*, 1986), which is an implicit, noniterative method for unsteady compressible flow equations. The standard *k*-epsilon turbulence model of Launder and Spalding (1974) is used to close the flow equations. This turbulence model requires the solution of only two additional differential equations, and has been documented extensively. Yet, it has not been used for flow–acoustic problems extensively, as limited research is available in this area. While it is almost certain that a better closure exists or can be developed specifically for flow–acoustic-type problems, such development is beyond the scope and main objective of the present work.

As discussed in the Introduction, the success thus far in modeling flow–acoustic interactions for flow over cavities has been limited, and the majority of the work is confined to shallow cavities. In order to evaluate the feasibility of the numerical approach, an investigation on deep side branches is undertaken by modeling the configuration used by Ziada (1993), who provides information on acoustic pressure amplitudes versus the Strouhal number and clear smoke visualization of the vortex as it travels across the branch mouth. The geometry, which is shown in Fig. 2, consists of a square main duct with two opposed side branches that are rectangular in cross section. The rectangular cross section allows for simple two-dimensional modeling. The main duct height is $h=0.092$ m, and the dimension of the side branch in the flow direction is $d=0.052$ m. Both side branches are of equal length $L=1.0$ m. The visualization study was performed at atmospheric pressure and low Mach number ($M<0.1$) with a loudspeaker placed at the end of one of the side branches, as shown in Fig. 2, to control the acoustic excitation level in the system. The effect of the Strouhal number on acoustic pressure amplitudes was also investigated (without the loudspeaker) by varying the flow velocity. In order to reduce the viscous dissipation of acoustic waves, experiments not requiring visualization were performed at an elevated pressure of 0.35 MPa (approximately 3.5 atm). Ziada and Buhlmann (1992) showed that increasing the mean pressure from 0.098 to 0.35 MPa increased the nondimensionalized acoustic pressure $2P_{\mathrm{rms}}/\rho_0 U^2$ by 60%.

The computational domain is two dimensional with a 2.0-m inlet duct and a 1.0-m outlet duct leading to a mesh of 17 250 cells. The inlet boundary condition reflects acoustic waves as a solid boundary would, while the outlet imposes a fixed pressure. The inlet length of 2.0 m was chosen to be twice the length of the side branch in order to reduce acoustic coupling between the two. A law of the wall model is used to simplify the turbulence model at the solid wall boundaries. Investigation of the specified turbulence intensity imposed at the inlet revealed that the intensity above 6% does not affect the acoustic pressure amplitudes. An inlet turbulence intensity of 10% was specified for the present runs. The inlet velocity was imposed using a 1/7 power-law velocity profile. All computational runs were initialized with a mean pressure of 0.35 MPa to match the pressure used by Ziada, and a temperature of 293 K. A mean velocity was also specified in the main duct initially, with zero velocity in the side branches. Ziada reported that the two side branches oscillated out of phase (the acoustic pressure at the top branch was maximum when the acoustic pressure at the bottom branch was minimum); therefore, a small pressure imbalance was introduced to the side branches at start-up to promote an out-of-phase response. The system was then run until a quasi-steady state was reached in which the only change in time was the acoustic oscillations. Computational time on a single SGI Origin 2000 processor for the above runs is approximately 2.4×10^{-5} CPU second/cell/time step or about 2 to 3 days per run. Details of the computations can be found in Radavich (2000).

Once the computations reach a quasi-steady state, the theory of Howe (1984) is used to relate the acoustic power generation

$$\Pi = -\int_{\mathrm{CV}} \overline{\rho_0 \mathbf{u}' \cdot (\omega \times \mathbf{u})}\, dV, \tag{1}$$

to the integral over a control volume, CV, of the time average of a triple product between the acoustic velocity $\mathbf{u}'$, the vorticity ω, and the flow velocity $\mathbf{u}$. Here, the acoustic velocity is the unsteady irrotational component of the velocity and the flow velocity $\mathbf{u}$ excludes the acoustic velocity. As shown by Jenvey (1989), Eq. (1) is a constant entropy, inviscid, low Mach number approximation to the sound power. For the present high Reynolds number, low Mach number flow, it provides a reasonable approximation to the sound power, and is convenient for studying the interaction between the flow and acoustic fields. Equation (1) is solved computationally by storing the velocity and density for each cell in the domain for an acoustic cycle. The vorticity is calculated from the total unsteady velocity, and the solenoidal velocity is solved from the vorticity. The irrotational velocity is simply the difference between the total velocity and the solenoidal velocity. Fast Fourier transforms (FFTs) of the irrotational velocity and density are then taken cell by cell to separate the acoustic irrotational velocity $\mathbf{u}'$ and the mean density ρ_0. As the mean component of the irrotational ve-

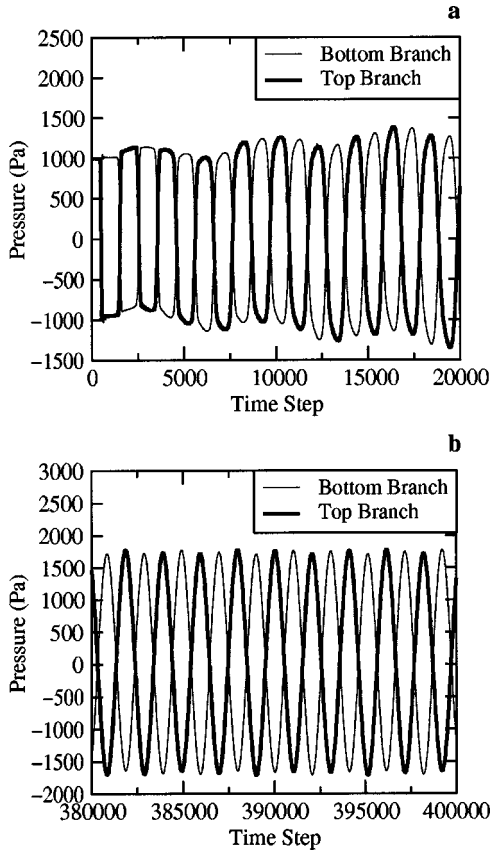


FIG. 3. Acoustic pressure amplitude during transient start-up at $St=0.4$; (a) start of run; (b) quasi-steady state.

locity is solenoidal (Batchelor, 1967), it has been added to the solenoidal velocity which yields the flow velocity **u** used in Eq. (1). The unsteady acoustic power is then integrated for each cell (or control volume) followed by time averaging once again using an FFT. Results of this approach along with experimental comparisons are discussed next.

III. RESULTS AND DISCUSSION

An example of the transient start-up of a run for $St=0.4$ is given in Fig. 3, which shows the acoustic pressure at the closed ends of the top and bottom side branches versus the time step. The acoustic amplitude at the first time step in Fig. 3(a) is determined by the pressure discontinuity specified in the initial conditions. This discontinuity promotes an out-of-phase oscillation in which vortices are shed alternately between the top and the bottom branches. Later in Fig. 3(b), the system has reached a quasi-steady state in which the acoustic pressure oscillations are repeating. If there were no coupling between the mean flow and the acoustic waves, the combined effect of viscosity and the numerical dissipation would cause the initial oscillations of Fig. 3(a) to gradually dissipate to zero amplitude, since there is no acoustic source in the system to sustain these oscillations. Instead, the amplitudes grow until a final amplitude is reached in Fig. 3(b) that is greater than the initial oscillation. For this to occur, energy must be transferred from the mean flow to the acoustic oscillations, demonstrating a true coupling between the acoustic field and the flow field.

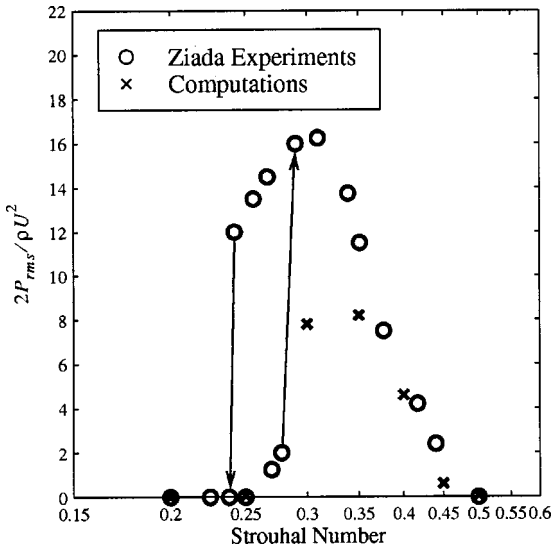


FIG. 4. Nondimensionalized acoustic pressure amplitude versus Strouhal number.

To determine how this coupling varies with flow conditions, a series of runs was performed by holding the geometry of the side branches fixed and varying the inlet flow velocity to obtain Strouhal numbers ranging from 0.2 to 0.5. Figure 4 demonstrates how the acoustic pressure amplitude at the resonance frequency of the side branches varies with the Strouhal number. Since both branches produce similar amplitudes, only results from the bottom branch are shown here. The computations show excitation in the region from $St=0.25-0.5$. The nondimensionalized pressure amplitude peaks near a value of 8 at $St=0.35$ and drops off as the velocity goes lower or higher. Outside of the region from $St=0.25-0.5$ the acoustic oscillations are so small that the nondimensionalized pressure approaches zero amplitude. The experimental results from Ziada (1993) are also shown in Fig. 4. According to Ziada, the loop in the experimental data is a nonlinear hysteresis effect which shows that as the flow velocity is gradually increased, there is a region where the feedback loop is ''locked in'' and the coupling continues above its normal value. Decreasing the flow through this region avoids the lock-in region and produces different results. It was not attempted to reproduce the loop in this study due to the additional computational time needed to slowly ramp up the velocity. In the experiments by Ziada, mufflers were placed upstream and downstream of the side branch, and because of this the reflection properties of the experimental inlet and outlet are unknown. In the computations, the inlet and outlet lengths were chosen so that they would not couple with the side branch at its quarter-wave frequency. If the mufflers used in the experiments are assumed to reduce the interaction between the main duct and the side branch by decreasing reflections at the boundaries, then the experimental pressure amplitudes should be similar to the computational results, which is observed to be true in Fig. 4. Although a direct comparison cannot be made between the experiments and the computations due to these boundary differences, it is important to emphasize that the computational method predicts the flow–acoustic coupling in the same

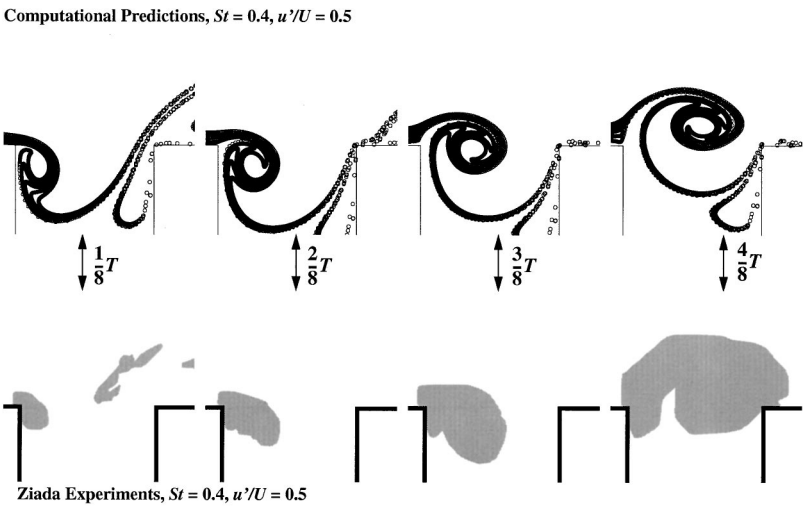


FIG. 5. Smoke visualization comparison between computations and experiments at four points in the acoustic cycle [experimental data sketched from Ziada (1993)].

Strouhal range of 0.25 to 0.5. In terms of sound-pressure level, the difference between the computations and the experiments is only about 6 dB at $St=0.3$. As discussed in Sec. II, however, increasing the resolution of the computations may increase its acoustic predictions by as much as 3 dB, which would drop the difference to only 3 dB.

The paper by Ziada also contains detailed smoke visualization of the formation and propagation of vortices at the interface between the main duct and the side branch. Sketches of these smoke figures are shown in Figs. 5 and 6 for $St=0.4$. This experimental work provides a series of eight figures that shows the formation and propagation of the vortex during a single acoustic cycle; however, neither the phase of the first figure nor the phasing between consecutive figures is specified. It is assumed here that the acoustic cycle was divided into equal increments with one figure at each increment. In order to compare with experiments, the present computational method released a large number of particles into the flow and tracked their positions to simulate smoke plots. The computational work divided the acoustic cycle into regular intervals and these figures were then matched to the experimental smoke pictures. Due to the uncertainty in the phase of the experimental figures, some phase differences may exist between the experiments and computations in Figs. 5 and 6. In the experiments performed by Ziada, a loudspeaker was used to increase the oscillation amplitudes for improved visualization (recall Fig. 2), and the experiments were conducted at atmospheric pressure to simplify the smoke injection process. The computations achieved similar high amplitudes by imposing an oscillating velocity at the top side-branch end. To equalize the comparisons, the ratio of acoustic velocity to mean flow velocity u'/U was matched, with both having a ratio of 0.5. The acoustic velocity is approximated in both the experiments and the computations by normalizing the pressure at the bottom closed branch end by $\rho_0 c$. The results are shown in Figs. 5 and 6 for one complete cycle where the acoustic period, T, has been divided into eight increments. Overall, the comparison between the computational predictions and the experiments is excellent, suggesting that the actual physics of the flow–acoustic coupling is captured in the computations. In the sequence of figures, the acoustic pulse pulls downward into the side branch at $T/8$, enhancing the roll-up of the vortex off of the upstream edge. The vortex continues to grow and is convected along by the mean flow at $2T/8$ and $3T/8$. The compressed fluid in the side branch then pushes back, forcing the vortex upwards from $3T/8$ to $5T/8$. The cycle then repeats itself as the acoustic pulse reverses at $6T/8$ and $7T/8$, causing the roll-up of another vortex. The simultaneous existence of two vortices generated one acoustic period apart at $7T/8$

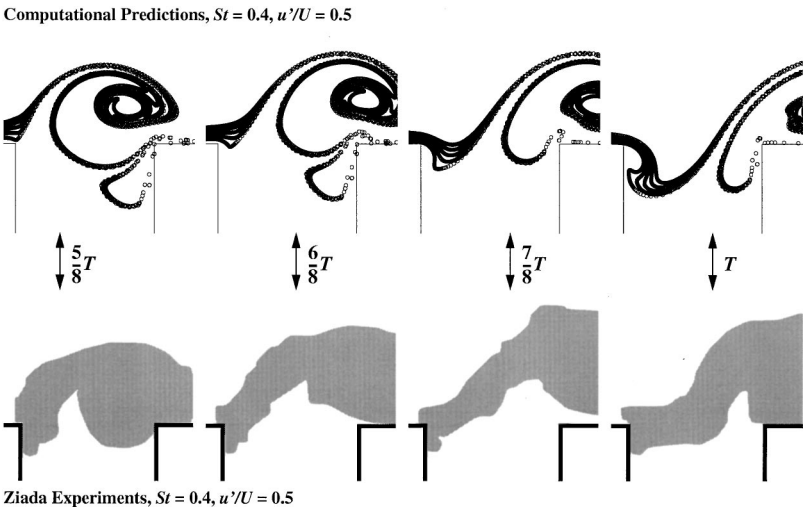


FIG. 6. Smoke visualization comparison between computations and experiments at four points in the acoustic cycle [experimental data sketched from Ziada (1993)].

FIG. 7. Acoustic source power time averaged over the acoustic cycle for $St=0.4$, $u'/U=0.5$ (black lines denote zero amplitude).

allows for an approximation of the average vortex convection velocity for the cycle. Dividing the distance (measured in the mean flow direction) between the estimated centers of the two vortices by the acoustic period gives an average vortex convection velocity of approximately 5 m/s. This yields a vortex convection velocity to mean flow velocity ratio near 0.45 for this case with $St=0.4$ and $u'/U=0.5$, whereas Ziada (1993) approximates the value to be closer to 0.6. This difference is due primarily to the difficulty in locating the vortex centers from the smoke plots; however, as Figs. 5 and 6 reveal, the vortex convection velocity is closely matched between the computations and experiments.

The approximate acoustic power produced over an acoustic cycle by the interaction between the vortex, the velocity field, and the acoustic velocity is shown in Fig. 7. Since the computations are two-dimensional, the power is displayed per unit depth in the third dimension. This plot, which focuses on the junction between the main duct and the lower side branch, was produced by integrating the triple product [Eq. (1)] over each computational cell and time averaging the results. The computations predict that a large net source of acoustic power is produced in the main duct just outside of the side branch, while a concentrated acoustic sink is located near the upstream corner. The details of where in the acoustic cycle this acoustic power is produced can be examined by solving for the instantaneous approximate sound power

$$\Pi_{\text{inst}}=-\int_{\text{CV}}\rho_0\mathbf{u}'\cdot(\omega\times\mathbf{u})dV, \tag{2}$$

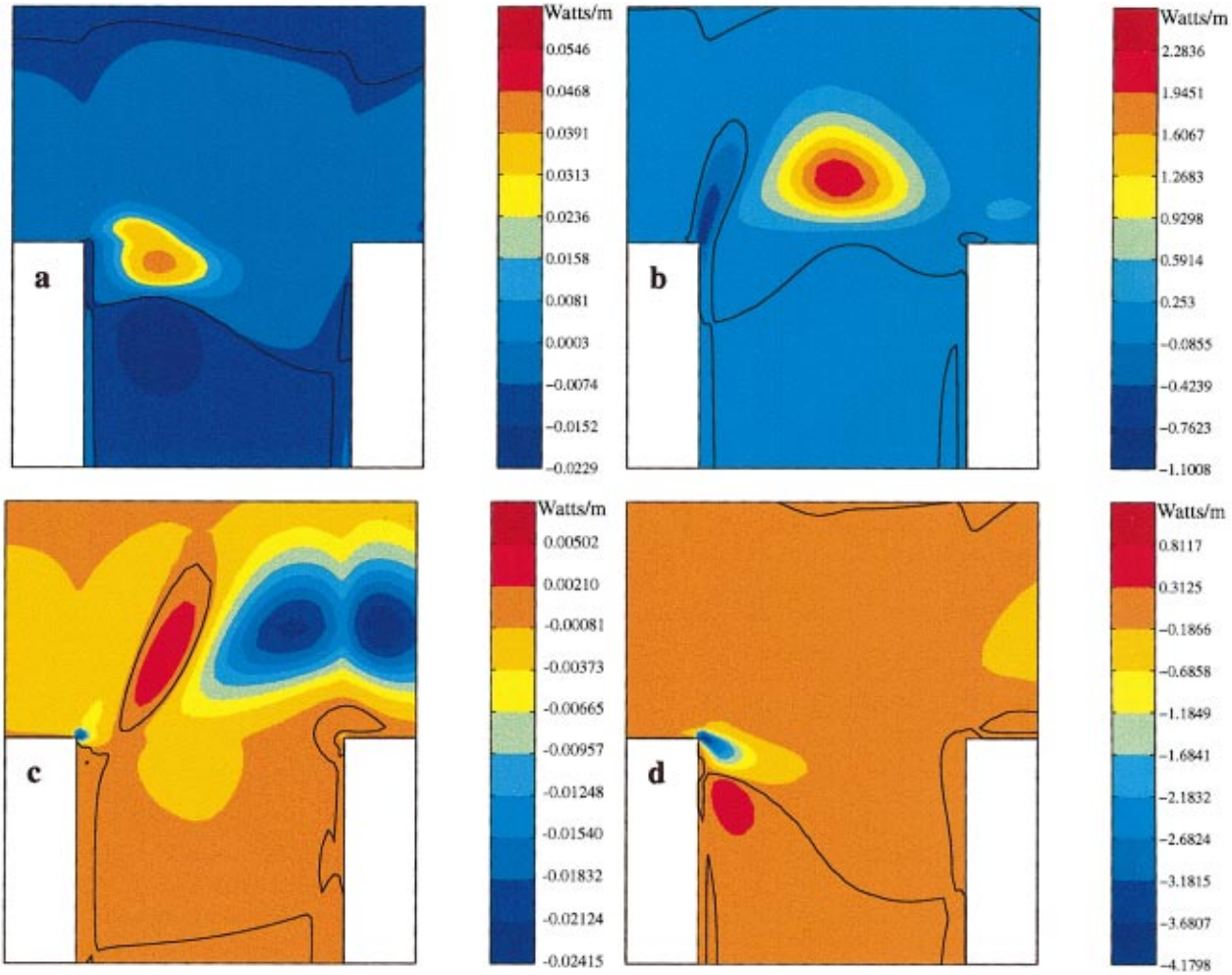


FIG. 8. Instantaneous sound power for $St=0.4$, $u'/U=0.5$, timing of acoustic period corresponds to Figs. 5 and 6; (a) $2T/8$; (b) $4T/8$; (c) $6T/8$; (d) T (black lines denote zero amplitude).

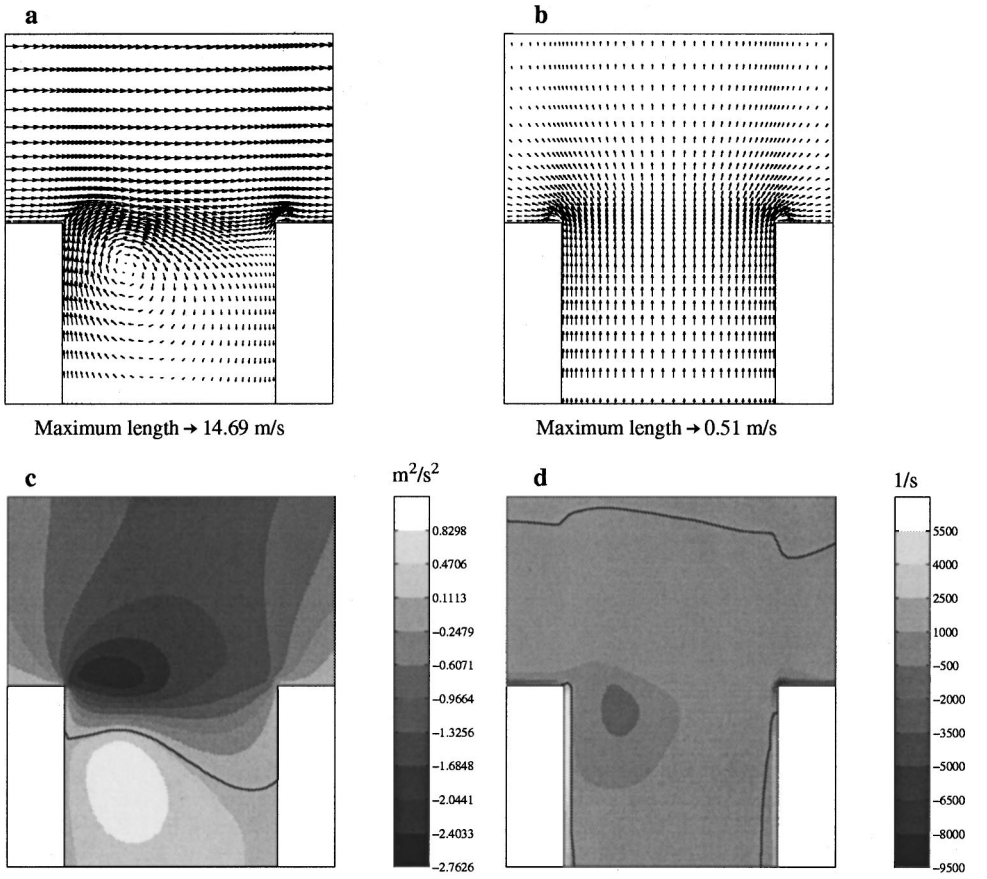


FIG. 9. Components of the triple product [Eq. (1)] for Fig. 8(a), $St=0.4$, $u'/U=0.5$; (a) $\mathbf{u}$; (b) $\mathbf{u}'$; (c) $-\mathbf{u}\times\mathbf{u}'$; (d) ω [black lines in (c) and (d) denote zero amplitude, white in (d) is due to amplitudes off of color bar].

taken at different points in the acoustic cycle before time averaging. The instantaneous power at four different times are shown in Fig. 8. These four times correspond to $2T/8$, $4T/8$, $6T/8$, and T in the smoke plots of Figs. 5 and 6. For each time instant in Fig. 8, further details of the flow field are given in Figs. 9–12, which display the instantaneous velocity vectors, acoustic velocity vectors, vorticity, and the cross product $-\mathbf{u}\times\mathbf{u}'$. The triple product in Eq. (2) can be rearranged as

$$\Pi_{\text{inst}}=-\int_{\text{CV}}\rho_0\omega\cdot(\mathbf{u}\times\mathbf{u}')dV, \tag{3}$$

indicating that the product of the vorticity with the cross product $-\mathbf{u}\times\mathbf{u}'$ is the main component of the acoustic power.

Focusing on the first time corresponding to $2T/8$ in Fig. 5, Fig. 8(a) shows a small source with peak amplitude of about 0.05 W/m located near the upstream corner of the junction between the main duct and lower side branch. A weaker sink is also located further down in the side branch. The instantaneous and acoustic velocity vectors at this time are shown in Figs. 9(a) and (b). The velocity in Fig. 9(a) shows a large recirculation near the upstream corner where the smoke plot rolls up at $2T/8$ in Fig. 5. The acoustic velocity in Fig. 9(b) shows that the acoustic pulse is weakly pushing up out of the side branch at this time. With the outward acoustic velocity being primarily perpendicular to the flow in the main duct, the cross product of the velocity and acoustic velocity in Fig. 9(c) reveals a negative contribution outside of the side branch. The presence of the vortex tends to augment the negative cross product near the upstream corner, while it creates a weaker positive contribution further down in the side branch. The negative vorticity in Fig. 9(d) is located near the transition from positive to negative in the cross product, thus via Eq. (3) producing a source near the upstream corner and a sink below this as shown in Fig. 8(a). The peak negative contribution of the cross product is greater in amplitude than the positive contribution, and thus the acoustic source at this time is stronger than the sink.

At the next time step of $4T/8$ in Fig. 5, the instantaneous power in Fig. 8(b) shows a large strong source with a peak amplitude near 2.3 W/m located in the main duct nearly centered on the junction. Comparison with Fig. 7 reveals that this portion of the acoustic cycle is a major contributor to the time-averaged acoustic source. Breaking down the triple product in Fig. 10 shows that the acoustic pulse is pushing strongly out of the side branch at this time. This strong upward pulse perpendicular to the flow in the main duct produces a large negative cross-product term in Fig. 10(c). This, in the presence of the vorticity in the junction, yields a strong contribution to the acoustic power. Comparison with Fig. 9 reveals that the magnitude of the vorticity near the center of circulation in Fig. 10(d) is approximately the same as the vorticity in Fig. 9(d). The amplitude of the cross product has increased greatly from Fig. 9(c) to Fig. 10(c) however, which causes the acoustic power in Fig. 8(b) to be much greater than the power in Fig. 8(a). Comparison between the location of the acoustic sources for these two times with their corresponding smoke plots in Fig. 5 illustrates that the location

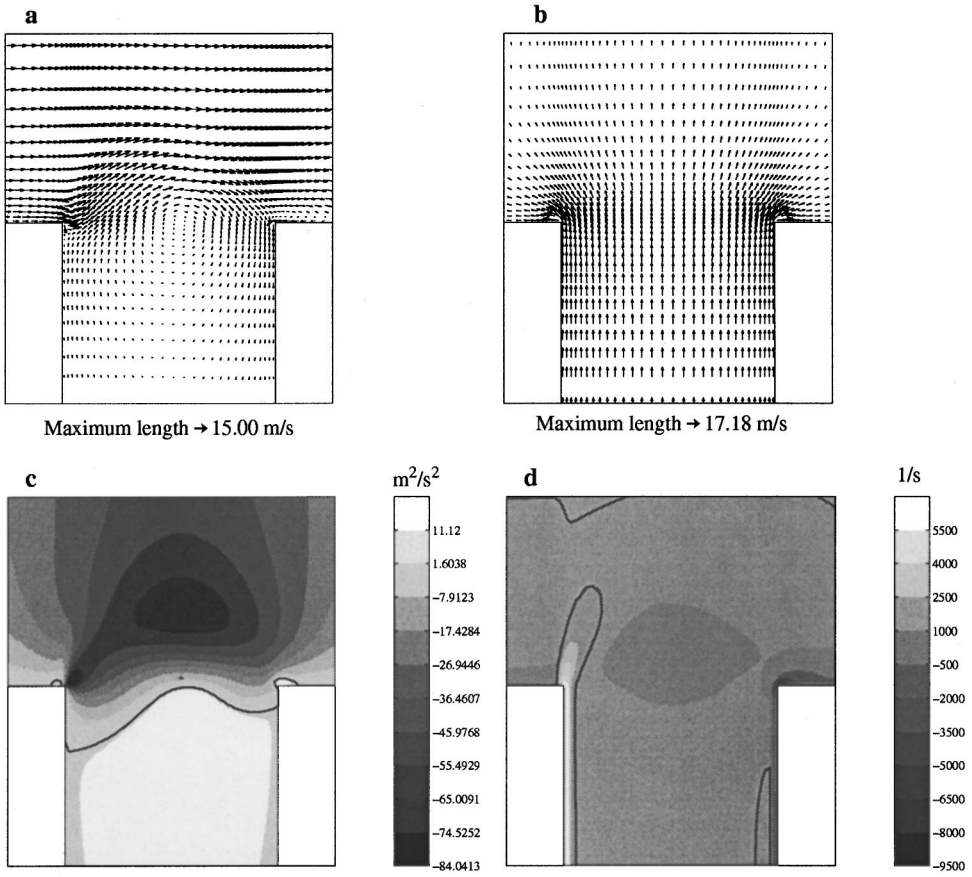


FIG. 10. Components of the triple product [Eq. (1)] for Fig. 8(b), $St=0.4$, $u'/U=0.5$; (a) $\mathbf{u}$; (b) $\mathbf{u}'$; (c) $-\mathbf{u}\times\mathbf{u}'$; (d) ω [black lines in (c) and (d) denote zero amplitude, white in (d) is due to amplitudes off of color bar].

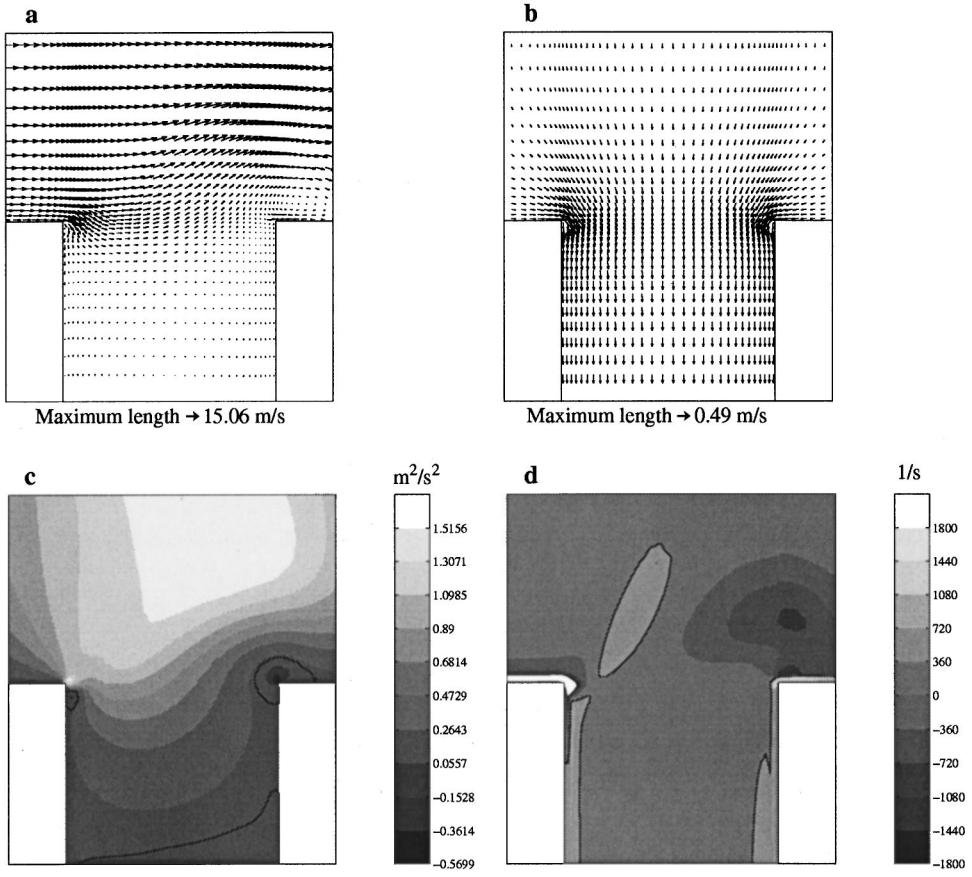


FIG. 11. Components of the triple product [Eq. (1)] for Fig. 8(c), $St=0.4$, $u'/U=0.5$; (a) $\mathbf{u}$; (b) $\mathbf{u}'$; (c) $-\mathbf{u}\times\mathbf{u}'$; (d) ω [black lines in (c) and (d) denote zero amplitude, white in (d) is due to amplitudes off of color bar].

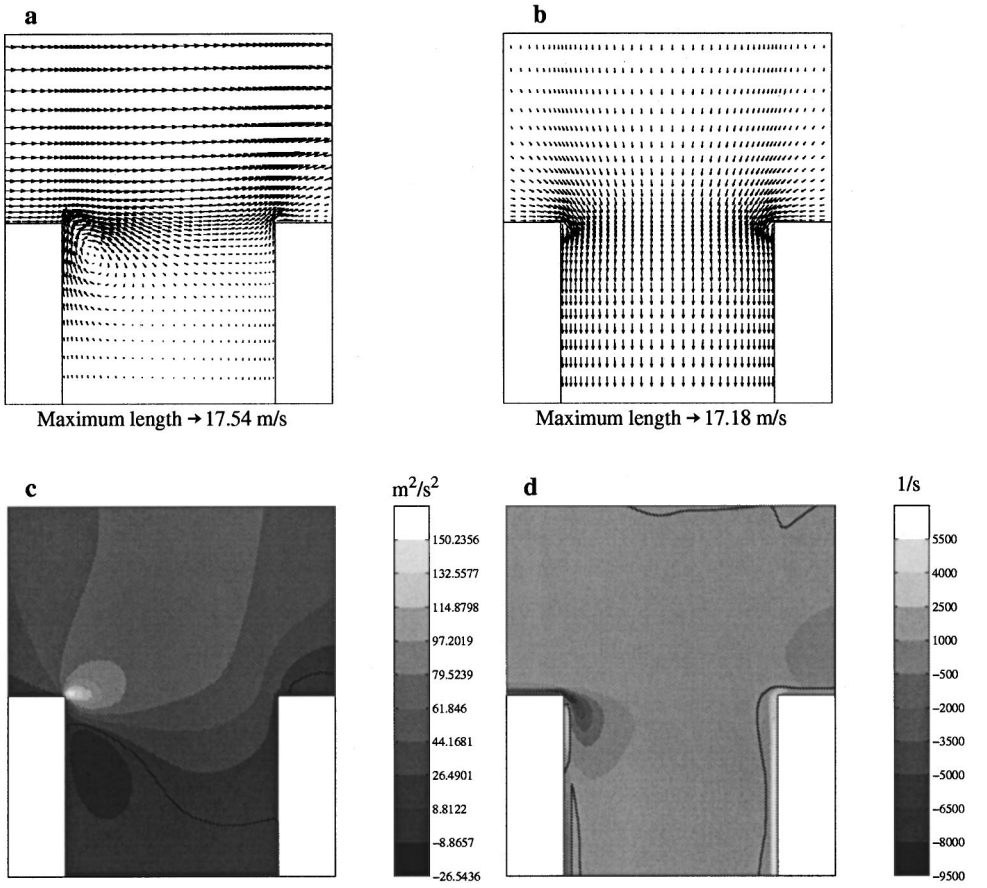


FIG. 12. Components of the triple product [Eq. (1)] for Fig. 8(d), $St=0.4$, $u'/U=0.5$; (a) $\mathbf{u}$; (b) $\mathbf{u}'$; (c) $-\mathbf{u}\times\mathbf{u}'$; (d) ω [black lines in (c) and (d) denote zero amplitude, white in (d) is due to amplitudes off of color bar].

and size of the sources closely resemble the location and size of the smoke vortices.

Next at $6T/8$ in Fig. 6, the acoustic source has turned to a large relatively weak sink in Fig. 8(c), with peak amplitude near -0.024 W/m. The vorticity in Fig. 11(d) still shows a large negative circulation above the downstream corner of the side branch, and the strength of the vortex has decreased only slightly. The acoustic velocity has now reversed direction in Fig. 11(b), and is now weakly pulling into the side branch. This downward pull creates a mostly positive cross-product contribution outside of the side branch, as shown in Fig. 11(c). The combination of the positive cross product with the negative vorticity now produces an acoustic sink over the downstream corner. However, the weak acoustic velocity does not give the cross product a strong contribution, and the sink is relatively weak because of this.

Finally, the last frame investigated corresponds to time T in Fig. 6. Figure 8(d) shows that the acoustic power is mostly negative during this portion of the acoustic cycle, with a local concentrated peak sink of approximately -4.2 W/m located at the upstream corner. The velocity in Fig. 12(a) reveals that the vortex has already begun to roll up near the upstream corner at this time. During this portion of the acoustic cycle, the acoustic pulse in Fig. 12(b) is drawing strongly down into the side branch. This produces a positive cross-product component outside of the side branch in Fig. 12(c). Similar to the observations in Fig. 9(c), but with the acoustic velocity in the opposite direction, the presence of the vortex near the upstream corner tends to augment the cross product near the upstream corner and create a negative component further down in the side branch. The presence of the negative vorticity in Fig. 12(d) combined with the strong positive cross product near the upstream corner creates a strong acoustic source near the upstream corner. Because the vorticity is highly localized near the upstream corner, however, the strong sink in Fig. 8(d) is confined to the vicinity of the upstream corner.

Comparing the figures at different times in the acoustic cycle, the acoustic source is produced as the acoustic velocity pulse pushes upward out of the side branch, while an acoustic sink is produced as the acoustic velocity pulls back into the side branch. Comparing these two conditions in Figs. 10 and 12 reveals that the peak amplitude of the negative values of the cross product from Eq. (3) in Fig. 10(c) is nearly half the peak amplitude of the positive values of the cross product in Fig. 12(c). The vorticity in Fig. 10(d) occurs over a broader area than the stronger, more condensed vorticity in Fig. 12(d), which results in the source of Fig. 8(b) being larger in size than the sink in Fig. 8(d). Examining the time-averaged acoustic source in Fig. 7 once again shows that the source and sink produced during the upward and downward acoustic pulses carry over into the time averaged power. Once the entire acoustic cycle has been accounted for, a large source remains over the junction between the main duct and the side branch with a localized sink near the upstream corner.

IV. CONCLUDING REMARKS

This work has demonstrated the ability of conventional CFD methods to solve coupled flow–acoustic problems for a

configuration with two coaxial side branches attached to a main duct. A true coupling is computationally possible between the flow field and the acoustic field, allowing acoustic oscillations to grow larger than their initial values at certain flow conditions. Computations with several different flow velocities demonstrated the ability to properly identify when the flow–acoustic coupling should and should not occur. Comparisons with the experimental smoke visualization work of Ziada (1993) show the ability to duplicate the motion of the vortex as it propagates across the side-branch opening and interacts with the acoustic waves in the side branch. Using the theory of Howe (1984), it was shown that a large net acoustic source is produced in the main duct just outside of the side branch due to the interaction between the vortices and the flow and acoustic fields.

Anderson, J. S. (**1977**). ''The effect of an air flow on a single side branch Helmholtz resonator in a circular duct,'' J. Sound Vib. **52**(3), 423–431.

Ball, J. W. (**1959**). ''Hydraulic characteristics of gate slots,'' J. Hydraul. Div., Am. Soc. Civ. Eng. **85**, 81–114.

Batchelor, G. K. (**1967**). *An Introduction to Fluid Dynamics* (Cambridge University Press, Cambridge, MA).

Baysal, O., Srinivasan, S., and Stallings, R. L. (**1988**). ''Unsteady viscous calculations of supersonic flows past deep and shallow three-dimensional cavities,'' AIAA Paper 88-0101.

Baysal, O., and Yen, G. (**1990**). ''Implicit and explicit computations of flows past cavities with and without yaw,'' AIAA Paper 90-0049.

Baysal, O., Yen, G., and Fouladi, K. (**1994**). ''Navier-Stokes computations of cavity aeroacoustics with suppression devices,'' J. Vibr. Acoust. **116**, 105–112.

Bruggeman, J. C., Wijnands, A. P. J., and Gorter, J. (**1986**). ''Self-sustained low frequency resonance in low-Mach-number gas flow through pipe-lines with side-branch cavities: A semi-empirical model,'' AIAA Paper 86-1924.

Bruggeman, J. C., Hirschberg, A., Van Dongen, M. E. H., Wijnands, A. P. J., and Gorter, J. (**1991**). ''Self-sustained aero-acoustic pulsations in gas transport systems: Experimental study of the influence of closed side branches,'' J. Sound Vib. **150**(3), 371–393.

Chokani, N., and Kim, I. (**1991**). ''Suppression of pressure oscillations in an open cavity by passive pneumatic control,'' AIAA Paper 91-1729.

Covert, E. E. (**1970**). ''An approximate calculation of the onset velocity of cavity oscillations,'' AIAA J. **8**(12), 2189–2194.

De Metz, F. C., and Farabee, T. M. (**1977**). ''Laminar and turbulent shear flow induced cavity resonances,'' AIAA Paper 77-1293.

Dunham, W. H. (**1962**). ''Flow-induced cavity resonance in viscous compressible and incompressible fluids,'' Fourth Symposium on Naval Hydrodynamics and Propulsion Hydroelasticity, Office of Naval Research, pp. 1057–1081.

Durgin, W. W., and Graf, H. R. (**1992**). ''Flow excited acoustic resonance in a deep cavity: An analytical model,'' ASME—Symposium on Flow-Induced Vibration and Noise, **7**, pp. 81–91.

East, L. F. (**1966**). ''Aerodynamically induced resonance in rectangular cavities,'' J. Sound Vib. **3**(3), 277–287.

Erickson, D. D., Durgin, W. W., Maguire, C. F., and Moeller, M. J. (**1986**). ''Shear layer coupling with side-branch resonators,'' Forum on Unsteady Flow, ASME Publication No. FED39, pp. 43–45.

Erickson, D. D., and Durgin, W. W. (**1987**). ''Tone generation by flow past deep wall cavities,'' AIAA Paper 87-0167.

Franke, M. E., and Carr, D. L. (**1975**). ''Effect of geometry on open cavity flow-induced pressure oscillations,'' AIAA Paper 75-492.

Hankey, W. L., and Shang, J. S. (**1979**). ''Analyses of pressure oscillations in an open cavity,'' AIAA J. **18**(8), 892–898.

Hardin, J. C., and Pope, D. S. (**1995**). ''Sound generated by flow over a two-dimensional cavity,'' AIAA J. **33**(3), 407–412.

Heller, H. H., and Bliss, D. B. (**1975**). ''Flow-induced pressure fluctuations in cavities and concepts for their suppression,'' AIAA Paper 75–491.

Hersh, A. S., Walker, B., and Bucka, M. (**1978**). ''Effect of grazing flow on the acoustic impedance of Helmholtz resonators consisting of single and clustered orifices,'' AIAA Paper 78-1124.

Hourigan, K., Thompson, M. C., Brocher, E., and Andrianantoandro, A. (**1993**). ''Coupling of vortex shedding with the fundamental resonant mode of a resonator tube,'' Noise Control Eng. J. **41**(2), 331–337.

Hourigan, K., Welsh, M. C., Thompson, M. C., and Stokes, A. N. (**1990**). ''Aerodynamic sources of acoustic resonance in a duct with baffles,'' J. Fluids Struct. **4**, 345–370.

Howe, M. S. (**1997**). ''Edge, cavity, and aperture tones at very low Mach numbers,'' J. Fluid Mech. **330**, 61–84.

Howe, M. S. (**1984**). ''On the absorption of sound by turbulence and other hydrodynamic flows,'' IMA J. Appl. Math. **32**, 187–209.

Howe, M. S. (**1981**). ''The influence of mean shear on unsteady aperture flow, with application to acoustical diffraction and self-sustained cavity oscillations,'' J. Fluid Mech. **109**, 125–146.

Issa, R. I. (**1986**). ''Solution of the implicitly discretized fluid flow equations by operator-splitting,'' J. Comput. Phys. **62**, 40–65.

Issa, R. I., Gosman, A. D., and Watkins, A. P. (**1986**). ''The computation of compressible and incompressible recirculating flows by a noniterative implicit scheme,'' J. Comput. Phys. **62**, 66–82.

Jenvey, P. L. (**1989**). ''The sound power from turbulence: A theory of the exchange of energy between the acoustic and nonacoustic fields,'' J. Sound Vib. **131**(1), 37–66.

Jungowski, W. M., Botros, K. K., Studzinski, W., and Berg, D. H. (**1987**). ''Tone generation by flow past confined, deep, cylindrical cavities,'' AIAA Paper 87-2666.

Jungowski, W. M., Botros, K. K., and Studzinski, W. (**1989**). ''Cylindrical side branch as tone generator,'' J. Sound Vib. **131**(2), 265–285.

Kim, I., and Chokani, N. (**1990**). ''Navier–Stokes simulation of unsteady supersonic cavity flow field with passive control,'' AIAA Paper 90-3101.

Komerath, N. M., Ahuja, K. K., and Chambers, F. W. (**1987**). ''Prediction and measurement of flows over cavities—A survey,'' AIAA Paper 87-0166.

Kriesels, P. C., Peters, M. C. A. M., Hirschberg, A., Wijnands, A. P. J., Iafrati, A., Riccardi, G., Piva, R., and Bruggeman, J. C. (**1995**). ''High amplitude vortex-induced pulsations in a gas transport system,'' J. Sound Vib. **185**(2), 343–368.

Krishnamurty, K. (**1955**). ''Acoustic radiation from two-dimensional rectangular cutouts in aerodynamic surfaces.'' NACA TN 3487.

Launder, B. E., and Spalding, D. B. (**1974**). ''The numerical computation of turbulent flows,'' Comput. Methods Appl. Mech. Eng. **3**, 269–289.

Lilley, G. M., Zhang, X., and Rona, A. (**1997**). ''Progress in computational aeroacoustics in predicting the noise radiated from turbulent flows,'' Int. J. Acoust. Vib. **2**(1), 3–10.

Naudascher, M. (**1967**). ''From flow instability to flow-induced excitation,'' J. Hydraul. Div., Am. Soc. Civ. Eng. **93**, 15–40.

Nelson, P. A., Halliwell, N. A., and Doak, P. E. (**1983**). ''Fluid dynamics of a flow excited resonance. II. Flow acoustic interaction,'' J. Sound Vib. **91**(3), 375–402.

Panton, R. L. (**1988**). ''Effect of orifice geometry on Helmholtz resonator excitation by grazing flow,'' AIAA J. **28**(1), 60–65.

Radavich, P. M., ''Coupling between mean flow and acoustic resonances of a deep side branch,'' Ph.D. dissertation, The Ohio State University, 2000.

Rizzetta, D. P. (**1988**). ''Numerical simulation of supersonic flow over a three-dimensional cavity,'' AIAA J. **26**(7), 799–807.

Rockwell, D., and Naudascher, E. (**1978**). ''Review—Self-sustaining oscillations of flow past cavities,'' J. Fluids Eng. **100**, 152–165.

Sarno, R. L., and Franke, M. E. (**1990**). ''Suppression of flow-induced pressure oscillations in cavities,'' AIAA Paper 90-4018.

Shaw, L. L. (**1979**). ''Suppression of aerodynamically induced cavity pressure oscillations,'' J. Acoust. Soc. Am. **66**, 880–884.

Tam, C. K. W., and Block, P. J. W. (**1978**) ''On the tones and pressure oscillations induced by flow over rectangular cavities,'' J. Fluid Mech. **89**(2), 373–399.

Tam, C. J., Orkwis, P. D., and Disimile, P. J. (**1996**). ''Algebraic turbulence model simulations of supersonic open-cavity flow physics,'' AIAA J. **34**(11), 2255–2260.

Ziada, S. (**1993**). ''A flow visualization study of flow-acoustic coupling at the mouth of a resonant sidebranch,'' PVP-Vol. 258, Flow-Induced Vibration and Fluid-Structure Interaction, ASME 1993, pp. 35–59.

Ziada, S., and Buhlmann, E. T. (**1992**). ''Self-excited resonances of two side-branches in close proximity,'' J. Fluids Struct. **6**, 583–601.

Matched field processing with data-derived modes

Paul Hursky,[a] W. S. Hodgkiss, and W. A. Kuperman
Marine Physical Laboratory, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093

(Received 15 November 1999; revised 20 September 2000; accepted 19 December 2000)

The authors demonstrate MFP using data-derived modes and the sound speed profile, using no *a priori* bottom information. Mode shapes can be estimated directly from vertical line array data, without *a priori* knowledge of the environment and without using numerical wave field models. However, it is difficult to make much headway with data-derived modes alone, without wave numbers, since only a few modes at a few frequencies may be captured, and only at depths sampled by the array. Using a measured sound speed profile, the authors derive self-consistent, complete sets of modes, wave numbers, and bottom parameters from data-derived modes. Bottom parameters enable modes to be calculated at all frequencies, not just those at which modes were derived from data. This process is demonstrated on SWellEx-96 experiment data. Modes, wave numbers, and bottom parameters are derived from one track and MFP based on this information is demonstrated on another track. [DOI: 10.1121/1.1353592]

PACS numbers: 43.30.Bp, 43.30.Wi [DLB]

I. INTRODUCTION

Matched field processing (MFP) (Refs. 1, 2, and 3) utilizes models that require accurate *a priori* knowledge of the physical properties of the ocean and its bottom. However, the costly and often unattainable environmental accuracy demanded by current MFP processes provides strong motivation to develop localization methods that do not require such complete knowledge of the environment and that provide some flexibility in setting the trade-off between robustness on the one hand and localization accuracy and processing gain on the other.

Our main result in this paper is to show how MFP can be done with the aid of a measured sound speed profile but with no *a priori* bottom properties. The bottom properties are usually required as inputs to acoustic models used to construct MFP replicas with which the measured data is matched. We will initially estimate the mode depth functions of the prevailing ocean waveguide directly from data at selected frequencies from a moving source, then make use of a measured sound speed profile to estimate the parameters of a half-space bottom. Having the bottom parameters enables us to use acoustic models, as is usually done, to construct pressure field replicas for MFP at all frequencies. We demonstrate this process on experimental data from a realistic shallow water coastal environment.

Mode shapes can be extracted from cross-spectral density matrices (CSDMs) measured on vertical arrays. This has previously been discussed in Refs. 4–9. In Sec. II we review how and when modes can be derived from data. Although it is remarkable that it can be done at all, it is unfortunate that the modes are extracted without their horizontal wave numbers. This and other difficulties (see Secs. II and III) make it challenging to make much practical use of data-derived modes without additional information.

Modes-only (without wave numbers) source depth estimators have previously been described in Refs. 10–15. Of these, Refs. 11, 13, 14, and 15 show results with experiment data. To overcome the inherent difficulties in working with data-derived modes using incoherent processing approaches such as these, we chose to relax our initial goal of making do with no *a priori* environmental information. In Sec. III, we show how to use a measured sound speed profile (our *a priori* information) and data-derived modes to derive a self-consistent set of modes, wave numbers and bottom parameters. In Sec. IV, we apply this technique to experimental data, and then use the resulting modes, wave numbers, and bottom in a MFP process on another track from the same experiment. We obtain results that are nearly identical to those achieved using MFP based on modes and wave numbers calculated by a normal mode model (see Ref. 16 for a description of the normal mode model) using the environmental information (including bottom properties not used in the data-derived MFP) obtained at great cost in an earlier experiment (see Ref. 17).

II. EXTRACTING MODE SHAPES FROM VERTICAL LINE ARRAY MEASUREMENTS

Various source distributions in range and depth previously have been reported to produce modes that are uncorrelated at the receiver (Refs. 18, 19, 20, and 9). In a modal context, this means that individual mode contributions to the pressure field at the receiver, considered as an individual time series, are uncorrelated. A consequence of this is that the eigenvectors of the cross-spectral density matrix (CSDM) measured on a vertical line array will correspond to the mode shapes. A number of previous reports have independently described extracting modes using this technique (Refs. 4, 6, 8, 9, and 21).

[a]Present address: Science Applications International Corporation, 888 Prospect Street, Suite 201, La Jolla, CA 92037. Electronic mail: paul.hursky@saic.com

A. Why the eigenvectors of the CSDM correspond to the modes

To describe how the eigenvectors of the CSDM measured on a vertical array can correspond to the normal modes, consider the modal expansion of **p**, the pressure at a vertical receiver array,

$$\mathbf{p} = U_r \mathbf{a}, \tag{1}$$

where vector element $p(i)$ is the pressure at the ith receive array element. The columns of U_r (indexed by m),

$$U_r(i,m) = u_m(z_i), \tag{2}$$

are the acoustic modes, sampled at the vertical array element depths z_i (i runs over the array elements), where $\mathbf{u}_m$ is a vector containing the mth mode.

Vector **a** depends upon the source location parameters, range r_s and depth z_s. In a range-independent propagation environment, the mth element of **a**,

$$a(m) = u_m(z_s) e^{jk_m r_s}, \tag{3}$$

is the contribution of the mth mode to the pressure field and is often called the mth mode amplitude.

The cross-spectral density matrix (CSDM) measured on this vertical line array is represented by

$$R = E\{\mathbf{p}\mathbf{p}^H\}, \tag{4}$$

in which $E\{\ \}$ indicates expectation. The CSDM R is typically estimated by time averaging $\mathbf{p}\mathbf{p}^H$.

The CSDM can also be represented by a modal expansion,

$$R = U_r E\{\mathbf{a}\mathbf{a}^H\} U_r^T. \tag{5}$$

The middle matrix factor in this expansion is the correlation matrix of mode amplitudes,

$$R_a = E\{\mathbf{a}\mathbf{a}^H\}. \tag{6}$$

An eigenvector decomposition of R produces the factorization VDV^H, in which unitary matrix V has columns that are the orthonormal eigenvectors of R, and diagonal matrix D has diagonal elements that are the eigenvalues of R. When R_a is diagonal, the columns of U_r are orthogonal (since R is Hermitian) and the eigenvector decomposition VDV^H corresponds factor for factor to the modal expansion of R, $U_r R_a U_r^T$, with the eigenvectors (columns of V) corresponding to the modes (columns of U_r).

In Appendix A we discuss the challenges faced by this approach with short or sparse arrays.

B. Scenarios that produce uncorrelated modes at the receive array

For the modes to be uncorrelated, the source distribution over depth and range must result in a diagonal matrix R_a. Refs. 18, 19, 20, and 9 have described source distributions in which this is the case.

In the waveguide theory for surface ambient noise discussed in Refs. 18 and 20, a modal expansion of the cross-spectral density is derived for a sheet of uncorrelated, uniformly distributed sources near the ocean surface. For typical values of horizontal wave numbers and modal attenuations, the predicted off-diagonal elements of the matrix R_a discussed in Sec. II A are small compared to the diagonal elements, with the result that the modes are uncorrelated. The critical parameters are the attenuation coefficients that determine whether the more distant larger surfaces contribute more than the less distant but smaller surfaces.

Reference 9 describes how a source moving over a long enough range interval will also result in a mode expansion in which the modes are uncorrelated. We have confirmed this in simulations and will demonstrate our processing on experiment data containing a moving source.

In both of these scenarios, the separation of the wave numbers plays a role. The closer the wave numbers, the longer their modal interference distance, and the larger the range interval the sources must be distributed over (or travel over) for the modes to be uncorrelated at the receive array. The spacing between wave numbers depends on a variety of factors, including the sound speed profile, ocean depth, and frequency. In general, as the depth of the ocean (measured in wavelengths) increases, the more modes the environment will support, causing more wave numbers to be packed into the same wave number interval (bracketed by the slowest sound speed in the water and the bottom sound speed). A deep water profile produces wave number spacings that are smaller at higher-order modes, while a shallow water profile produces wave number spacings that are smaller at the lower-order modes.

In Appendix A we describe how to compensate for array deformations such as tilt. Array deformations are a dilemma when extracting modes from ambient noise measurements. Such sources are typically distributed over large areas of the ocean at many different azimuths. Without separating such sources, as might be possible with a volumetric array, there is no way to compensate the data for array tilt, since each azimuth requires a different correction. Because the SWellEx-96 array had tilts of up to several degrees, we have worked on single moving sources for which array tilt can be compensated.

C. Getting real modes from complex eigenvectors

Given a less than fully spanning array and the resulting noisy and incomplete (not all the modes are found) modes that are eventually estimated, it is not immediately obvious how to identify which eigenvectors actually correspond to modes. In addition, the eigenvectors are typically complex and must be converted to mode depth functions that are real.

The technique that we have used is to progressively eliminate candidate eigenvectors throughout our processing sequence by checking their consistency with the physical constraints being imposed throughout our process. Initially, we eliminate eigenvectors having high-frequency content. We convert the remaining eigenvectors to real vectors by identifying their zero crossings and then taking their magnitude and applying an alternating sign to each interval delimited by zero crossings. During the shooting method, we eliminate candidate modes that are not consistent with the sound speed profile. While matching a bottom half-space model with the data-derived modes, we eliminate candidate modes that are not consistent with any bottom model.

III. MODE EXTRAPOLATION AND WAVE NUMBER INTERPOLATION

Practical difficulties making use of data-derived modes include the following:

(1) Incomplete arrays provide noisy estimates of modes.
(2) Modes are estimated only at depths spanned by the array so we cannot construct replicas for sources at other depths.
(3) Having modes but not wave numbers results in modes-only processes that provide only marginal gain.
(4) Only some of the modes are captured and only at frequencies emitted by the moving source of opportunity we are exploiting.

In this section, to overcome these difficulties, we relax the ambitious initial goal of making do with no *a priori* environmental information. We assume the sound speed profile has been measured over the water column, but that the bottom properties are not known. These are modest assumptions. The sound speed is routinely measured in most experiments, but the bottom properties are typically much more difficult to obtain. Yet bottom properties are an essential ingredient if acoustic models are to be used to calculate pressure field replicas for MFP. In Sec. III A, we present the data-dependent shooting method. We use a measured sound speed profile and the data-derived modes to extrapolate individual modes from the array depths to the entire water column. Significantly, the wave numbers are estimated in this process as well. In Sec. III B, we embed the data-dependent shooting method in a larger optimization process that produces a self-consistent set of bottom parameters, modes, and wave numbers.

A. Data-dependent shooting method—mode extrapolation

We can use the depth-separated wave equation to propagate (or shoot) a mode depth function $\Psi(z)$ from initial values at the surface to the bottom. A finite difference representation of this differential equation is

$$u_{i+1} = -u_{i-1} + \left\{ 2 - h^2 \left[\frac{\omega^2}{c^2(z_i)} - k^2 \right] \right\} u_i, \tag{7}$$

in which i is an index over depth. This depends on the frequency ω, depth-dependent sound speed $c(z)$, and the horizontal wave number k (the parameter we seek for each mode). Density and attenuation must be added if we propagate through the bottom. Given $c(z)$ in the water and two initial values (we use $u_0 = 0$ and $u_1 = 1$), $u(z)$ can be calculated for all values of k. However, this does not identify which discrete set of wave numbers corresponds to the normal modes of the waveguide. A well-known method used by normal mode models to find the discrete set of normal modes (and their k_m) is to vary k and look for modes (produced by shooting at those values of k) that satisfy a boundary condition at the bottom (see Ref. 20, Chap. 5). The boundary condition is based on the properties of the bottom. However, in our case, we do not need to match a boundary condition requiring bottom properties. We already have data-derived modes in the water column that we can match our shooting modes to. We can shoot modes at trial wave numbers until we produce a mode that matches our data-derived mode. This matching can be performed without knowing the bottom properties, since the shooting method must propagate the modes only to depths spanned by our data-derived modes, which are confined to the water column. This provides the modes over the entire water column (and their horizontal wave numbers!) without knowing the bottom information.

To find the wave number corresponding to a particular data-derived mode $\hat{u}_i$ (where i runs over the array element depths), we seek a global minimum of the function

$$F(k) = \sum_i \, (\hat{u}_i - u_i(k))^2, \tag{8}$$

in which Eq. (7) is used to calculate a mode shape $u_i(k)$ at each candidate k. This function is the squared difference between the data-derived mode $\hat{u}_i$ and the mode $u(k)$. The mode and wave number at the global minimum of this function are adopted as estimates of the full water column mode and its wave number that correspond to the particular data-derived mode $\hat{u}_i$ that they were matched with.

The shooting method produces denoised modes, extrapolated from the span of the array to the full water column, and estimates wave numbers in the process. However, we still have the same subset of the modes we started with, and only at the frequencies at which we were able to extract modes from data.

B. Matching a bottom model to data-derived modes—Wave number interpolation

The shooting method finds modes that are consistent with the measured sound speed profile, but does not take advantage of what we know about how modes relate to each other. For example, modes are orthogonal and they all see the same bottom (expressed as a boundary condition based on the bottom properties). Admitting only modes that meet these criteria collectively would reduce the overall search space. We initially imposed an orthogonality constraint, but got poor results. We speculate this is because orthogonality and consistency with the sound speed profile do not uniquely specify a mode set. Although it was attractive to simply sidestep having to deal with unknown bottom properties, forcing the modes to be consistent with a physical bottom model yielded much better results. We varied the parameters of a bottom model until we found the set of bottom parameters that produced modes that matched the data-derived modes. Since the shooting method is a Sturm–Liouville system, all modes and wave numbers calculated as we varied the bottom properties were implicitly orthogonal, as well as consistent with a single bottom model.

The way we implemented our optimization is similar to a technique reported in Ref. 22, in which the intent was to avoid having to completely recalculate modes and wave numbers every time a dynamic ocean environment changed. Only a small fraction of the water column near the surface was changing, while most of the sound speed profile and the

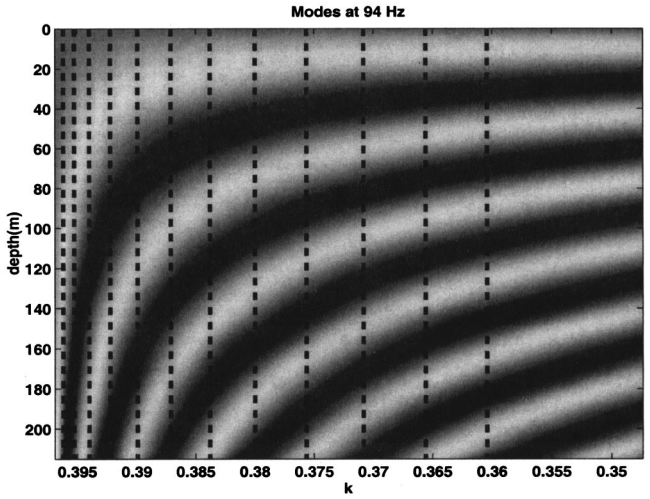


FIG. 1. Mode depth functions versus wave number at 94 Hz in the SWellEx-96 environment. The black dashed lines indicate the discrete modes calculated by a normal mode model, given the bottom information.

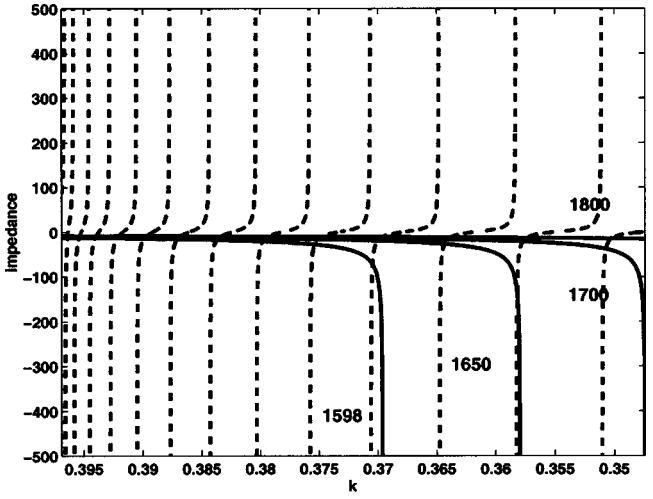


FIG. 2. Impedance calculated from mode depth functions (dashed lines) and the impedances calculated for several bottom half-space models at progressively higher compressional wave speeds (1598, 1650, 1700, and 1800, all solid lines, all speeds in m/s). The two impedances (one based on modes, the other based on the bottom) intersect at the discrete normal mode wave numbers. At 94 Hz in the SWellEx-96 environment.

bottom remained fixed. The shooting method was used to precalculate modes for the fixed part of the environment (by shooting from the bottom up to the dynamic part of the profile) over a fine grid of horizontal wave numbers. With each change in the small uppermost part of the sound speed profile, the shooting method was used to shoot modes from the surface, using the changed profile, down to the fixed part of the profile. The discrete set of modes was selected by matching the impedances of the lower modes with the (new) upper modes. The modes and impedances for the fixed lower part of the profile did not have to be recalculated with every change in the upper part of the profile.

In our case, we are interested in finding the bottom that is most consistent with the data-derived modes and the prevailing sound speed profile, which we regard as fixed. So, instead of precalculating the modes from the bottom up to a dynamic surface layer, the modes in the water column (assumed fixed in our optimization) were precalculated down to the bottom, whose parameters were varied during the optimization. For every candidate set of bottom parameters, modes and wave numbers were found by matching the impedance of the precalculated modes with the impedance of our candidate bottom. These modes were compared with the data-derived modes. The bottom parameters corresponding to the modes that best matched the data-derived modes were selected.

We chose a half-space bottom model with two parameters: compressional wave speed and density. This resulted in a particularly simple form for the impedance (see Ref. 20, Chap. 5),

$$Z_{\text{bottom}} = \frac{\rho_b}{\sqrt{k^2 - (\omega/c_b)^2}}. \tag{9}$$

The impedance from the precalculated mode depth functions (obtained by shooting from the surface down to the bottom) is

$$Z_{\text{water}} = \left.\frac{\psi}{\delta\psi/\delta z}\right|_{z=z_{\text{bottom}}}. \tag{10}$$

Figure 1 shows the precalculated continuum of modes versus horizontal wave number (for the SWellEx-96 environment). For reference, the black dashed lines indicate the discrete or normal mode wave numbers (as calculated by a normal mode model). Figure 2 shows two sets of impedances: the dashed line curves are calculated from the mode depth functions using Eq. (10) and the solid line curves show impedances calculated from several half-space bottom models using Eq. (9). Bottom impedances for wave speeds of 1550, 1600, 1700, and 1800 m/s are shown, all with a bottom density of 1.83 g/cm^3. The horizontal wave numbers of the normal modes occur where the impedance calculated from the mode-depth functions intersects the impedance calculated for a particular bottom.

To find a self-consistent set of modes, wave numbers, and bottom parameters, the measured sound speed profile was used by the shooting method to precalculate modes over a fine grid of wave numbers. At each wave number k, the impedance at the bottom (looking up into the water) was calculated using Eq. (10), producing a curve like the dashed line curve shown in Fig. 2. A two-dimensional search grid was formed over the two bottom parameters. At each set of bottom parameters, the impedance looking into the bottom is calculated for each k using Eq. (9), producing a curve like the solid line curves in Fig. 2. The normal modes for this set of bottom parameters occur at the k where these two impedance curves intersect. The modes are calculated at these discrete wave numbers (by shooting) and compared with the data-derived modes. This process is repeated for every candidate set of bottom parameters until the minimum summed squared difference between the calculated and data-derived modes is found. This resolves the bottom model that best matches our data-derived modes.

Note that to be compared with modes based on the bottom model, the data-derived modes must be assigned specific modes to be matched up with. This is done by counting zero crossings in the modes calculated by the shooting method that have been matched with the data-derived modes. This was adequate with the sound speed profile for which the method was demonstrated, but may be more challenging for more exotic profiles. For profiles where the number of zero crossings does not unambiguously identify which mode is which, multiple assignments would be evaluated, with the best result at each candidate set of bottom parameters kept for comparison. We have adapted the Hungarian algorithm (see Ref. 23), originally formulated for square assignment problems, for assigning a lesser number of data-derived modes to a subset of the larger complete set of modes in simulations, where the intent was to compare the data-derived with modes calculated by a normal mode model.

Note that the impedance looking up into the water only needs to be calculated a single time, since it depends on the sound speed, which we regard as fixed. The only calculations that must be repeated as bottom parameters are varied are the following.

(1) Use Eq. (9) to calculate the impedance looking into the bottom (again, over a finely spaced grid of wave numbers).
(2) Find the wave numbers at which the bottom impedance matches the precalculated impedance looking up into the water.
(3) Calculate (by shooting) the modes for these wave numbers (where the two impedance curves intersect).

We have embedded the data-dependent shooting method in a larger optimization, in which modes are sought that are consistent with the sound speed profile and the data-derived modes as before, and also consistent with a single physical bottom model. Although a simple bottom model was used, this technique can easily be generalized to more complex bottom models. For a bottom consisting of an arbitrary number of horizontal layers, we would shoot the wave equation up through all the layers, and calculate the impedance from the resulting wave function, using Eq. (10). However, this would impose more of a computational load than the half-space bottom model.

IV. PROCESSING SWellEx-96 EXPERIMENT DATA

In this section we will present results of processing data from the SWellEx-96 experiment, conducted in May 1996 ten kilometers off the coast of San Diego in California. Figure 3 shows the experimental configuration for the SWellEx-96 experiment (see Ref. 24). The summer sound speed profile has a thermocline at the surface and a nearly isovelocity profile below 50 m. The array spans the lower part of the water column (from 94 to 212 m). Figure 4 shows the bathymetry in the area of the experiment and two tracks, S5 (broken up into two legs, labeled S5a and S5b) and S9. The vertical line array was deployed from the FLIP research vessel, whose location is indicated by the black circle. Track S5 runs at 5 knots South to North along an isobath at a depth

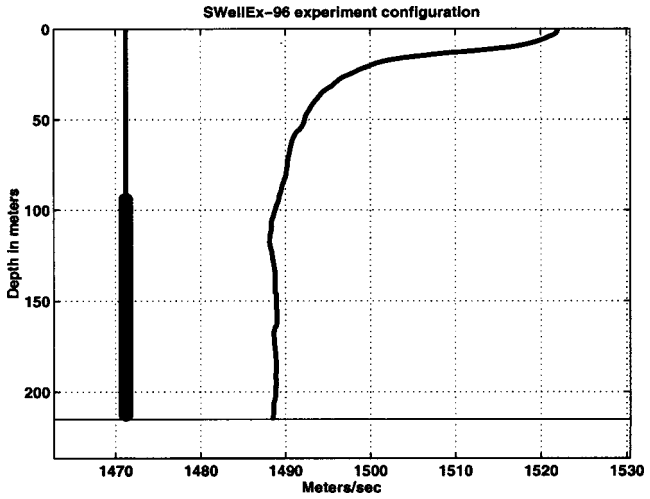


FIG. 3. The SWellEx-96 experiment environment: vertical line array and sound speed profile.

of 60 m. Two legs of Track S5 are shown. The first leg, labeled S5a, starts at a range of 8–9 km away from FLIP and closes to a range of 3–4 km. The second leg, labeled S5b, starts at a range of 3 km away from FLIP, closes to a closest point of approach (CPA) at 1 km, and then opens out to a range of 3 km on the North side of the CPA. The second track, Track S9 (only the half of the track that we processed is shown), runs nearly perpendicular to track S5 and the isobath lines from 3 to 5 km away from FLIP, traveling upslope from a point where the ocean depth is 200 m to a point where the ocean depth is less than 100 m. The bathymetry contours are spaced 20-m apart, with the heavier lines at 100-m increments.

In Sec. IV A, the techniques described in Secs. II and III

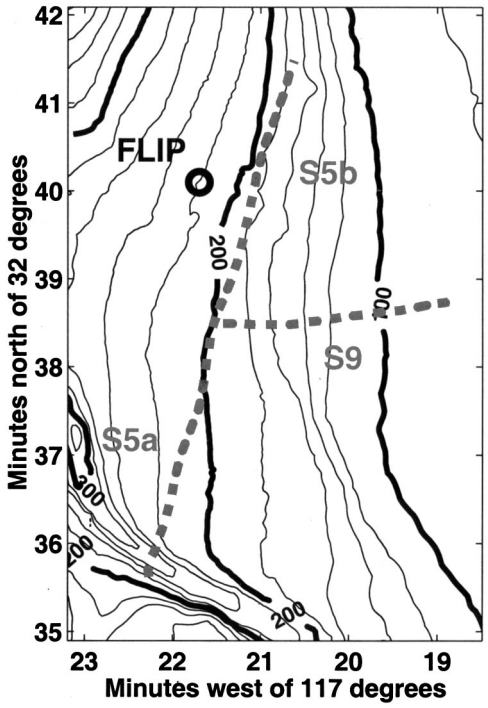


FIG. 4. Topographical map of the SWellEx-96 experiment site showing tracks S5 and S9.

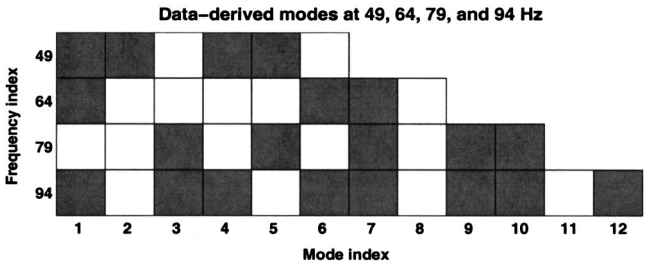


FIG. 5. The subset of modes extracted via eigenvector analysis from the SWellEx-96 S9 data at 49, 64, 79, and 94 Hz. The frequency index is shown along the *y* axis. The mode index is shown along the *x* axis. Gray indicates that a mode was extracted.

will be applied to SWellEx-96 Track S9 data to find a self-consistent set of modes, wave numbers, and bottom parameters. In Sec. III B, a fully coherent MFP process will be applied to SwellEx-96 Track S5 using the modes and wave numbers derived from Track S9. These methods rely upon *a priori* knowledge of the sound speed profile in the water, but assume no knowledge of the geoacoustic properties of the bottom, which would otherwise be needed to obtain modes and wave numbers from a normal mode model.

A. Modes, wave numbers, and bottom properties derived from SwellEx-96 Track S9

As the source travels up-slope along Track S9, a rich diversity in modal excitations is created at the receive array, because all of the different ocean depths generate different mode structures along the source track. This diversity in excitations at the receive array serves to produce the uncorrelated modes that enable us to extract modes via an eigenanalysis of the CSDM. As shown in Ref. 25, because of adiabatic mode propagation in this environment, the progressively shallower depths along the S9 track served to create ''mirages'' of the actual source locations at greater ranges and depths than the actual ranges and depths. So, although the actual range interval traveled by the source was several kilometers, in terms of the adiabatically propagating modes that arrived at the receiver, the receiver saw an apparent range interval that was several times that. Likewise, although the actual source depth was nearly constant, at 60 m, the apparent depth varied from 60 to 120 m, because of the changing bathymetry. From a MFP perspective, the range and depth bias errors resulting from these mirages are a nuisance, but for our modes-from-data process they were a great benefit.

We applied our mode extraction process to the CSDMs averaged over the part of the S9 track shown in Fig. 4. A significant number of modes was extracted from the tones at 49, 64, 79, and 94 Hz. Figure 5 shows four rows, corresponding to these four frequencies. The number of elements in each row corresponds to the number of modes at its frequency [as calculated by a normal mode model (see Ref. 16)]. The rows corresponding to lower frequencies have fewer modes. The modes that were extracted are gray. The modes that were not extracted are white.

Unfortunately, at higher frequencies (the same source also transmitted a series of tones from 112 to 388 Hz), only a few modes were extracted. This was consistent with our simulations of the SWellEx-96 experimental configuration, in which the array only spanned half the water column. We speculate that, because the number of propagating modes increases with frequency, the spacing between wave numbers tends to decrease and, as a result, the length of the range interval required for the modes to become uncorrelated increases with frequency.

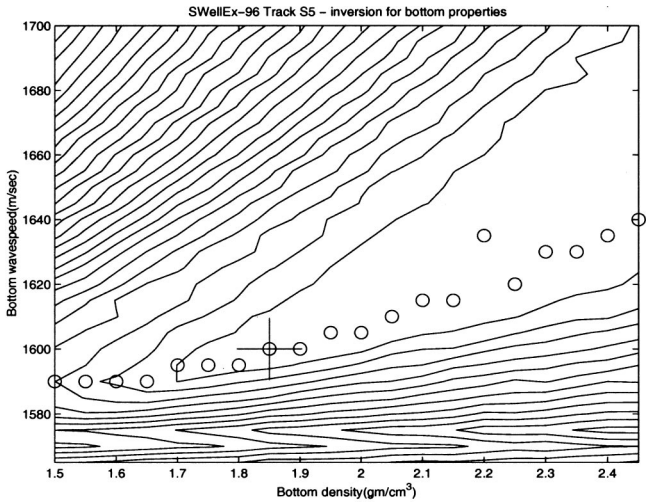


FIG. 6. Ambiguity surface for matching data-derived modes to a half-space bottom model based on bottom compressional wave speed and bottom density.

We applied the optimization described in Sec. III B to the modes extracted at these four frequencies. Figure 6 shows the ambiguity surface for matching a half-space bottom model to the modes derived from the experiment data. The circles, minimum points along the columns of the ambiguity surface, highlight the path through the lowest part of a valley running parallel to the horizontal density axis. Clearly, the modes are more sensitive to wave speed than density. The plus mark indicates the lowest point in the valley, and

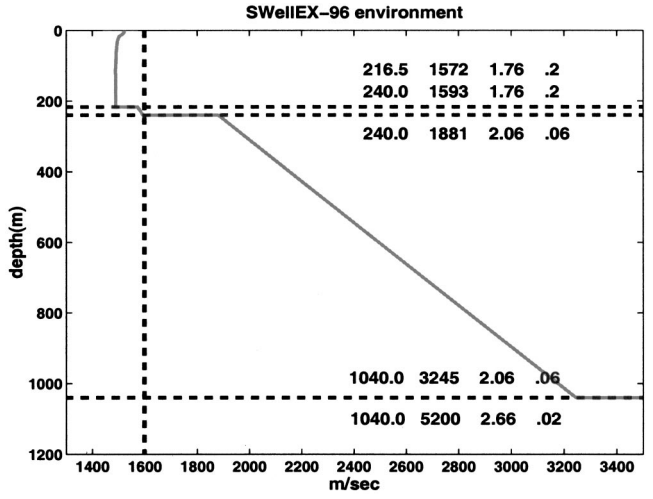


FIG. 7. Previously reported bottom parameters for the SWellEx-96 environment, based on direct measurements and joint source localization and bottom inversion. Our optimization produced a wave speed of 1598 m/s and a density of 1.83 g/cm^3. The 1598 m/s wave speed is indicated by the dashed vertical line. The sets of four parameters listed at each interface are depth m, compressional wave speed (m/s), density (g/cm^3), and attenuation (dB/km Hz).

occurs at a compressional wave speed of 1598 m/s and a bottom density of 1.83 g/cm^3. Previous measurements and inversion efforts for this environment have resolved the three layer model shown in Fig. 7. Our optimal values are comparable to values at the bottom of the first layer of the three layer model. Figure 8 shows the modes that were resolved by this optimization at 49 Hz. The data-derived modes are shown as solid lines (only modes 1, 2, 4, and 5 were recovered). The modes calculated from matching a bottom model to the data-derived modes are shown as dashed lines. The ''true'' modes (calculated by a normal mode model, using the best available environmental information) are shown as dotted lines.

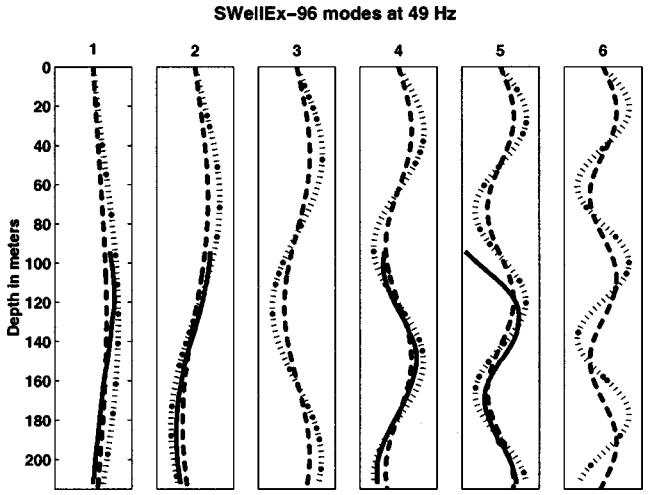


FIG. 8. Modes at 49 Hz resulting from our processes: subset of modes derived from eigenvectors of CSDM (limited to the span of array) are solid lines; modes calculated using resolved bottom half-space parameters are dashed lines; modes calculated by a normal mode model using the best-available environmental parameters are dotted lines.

B. Source localization results

We constructed replicas using the modes and wave numbers calculated in Sec. III A and performed MFP on both legs of the S5 track shown in Fig. 4.

The data was bandshifted, low pass filtered in two stages from 1500 Hz to 30 Hz, and spectrum analyzed using FFT periods of 256 points with Hanning windowing and no overlap. The spectra were incoherently averaged over all elements and the tones of interest were tracked using a peak picker confined to follow the peak within a small window. CSDMs were averaged over four spectrum snapshots (34.13 seconds per CSDM). A Bartlett matched field processor was used to generate all the results shown. MFP was performed at each individual frequency and the resulting individual ambiguity surfaces were normalized and averaged across frequency to form one composite ambiguity surface (see Ref. 26). To provide a benchmark against which to compare these results, we also performed MFP using modes and wave numbers calculated by a normal mode model (see Ref. 16), although these results will not be shown.

To show the relative processing gain provided by the various stages in our processing. Figures 9, 10, and 11 show three sets of MFP results. All plots show the time interval corresponding to the second half of Track S5, labeled S5b in Fig. 4. The source goes through its closest point of approach halfway through this interval, which was 40 min long and produced 70 range-depth MFP ambiguity surfaces. All plots show time in terms of the CSDM index along their horizontal axis (34.13 s per CSDM). The images shown in these figures contain time-evolving range (depth) tracks, as images in which each column is a slice versus range (depth) from a single range-depth MFP ambiguity surface. Each MFP range-depth surface contributed a slice versus range and a slice versus depth, both intersecting the maximum correlation peak of the range-depth surface.

Figures 9, 10, and 11 show results based on using only four of the available frequencies, those at which modes were extracted from data. Three processors are being compared in these figures: a modes-only depth estimation based on modes extrapolated by the shooting method and using only the subsets of modes extracted from data; a Bartlett MFP process based on modes and wave numbers obtained by the shooting method using complete sets of modes as a result of a polynomial interpolation of the wave numbers at the missing modes; a Bartlett MFP process based on modes and wave numbers calculated using the half-space bottom parameters resolved by the optimization process described in Sec. III B. Of these three, only the third would be able to exploit frequencies other than those at which modes were extracted from data. To provide a fair comparison between these processors, the third processor was limited to the same frequencies the other two processors were limited to. Later on, results based on all 13 tones and on 9 tones from a completely different source will be presented as well.

Figure 9 shows modes-only depth estimation results. Only a depth track is shown, because without wave numbers, we cannot form replicas that are functions of the source range. The data-dependent shooting method described in

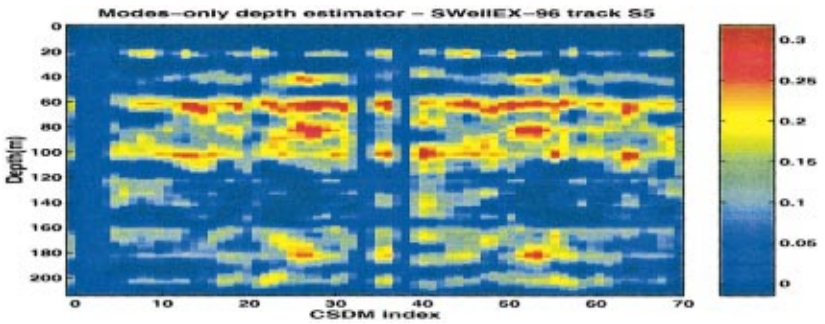


FIG. 9. The modes-only depth estimate track for the second half of SWellEx-96 track S5. Results are integrated over the four frequencies (49, 64, 79, and 94 Hz) at which modes were derived from the data in track S9. Modes have been extrapolated from the span of the array to the entire water column using the shooting method described in Sec. II. Only the modes derived from data were used (i.e., the modes drawn in gray in Fig. 5).

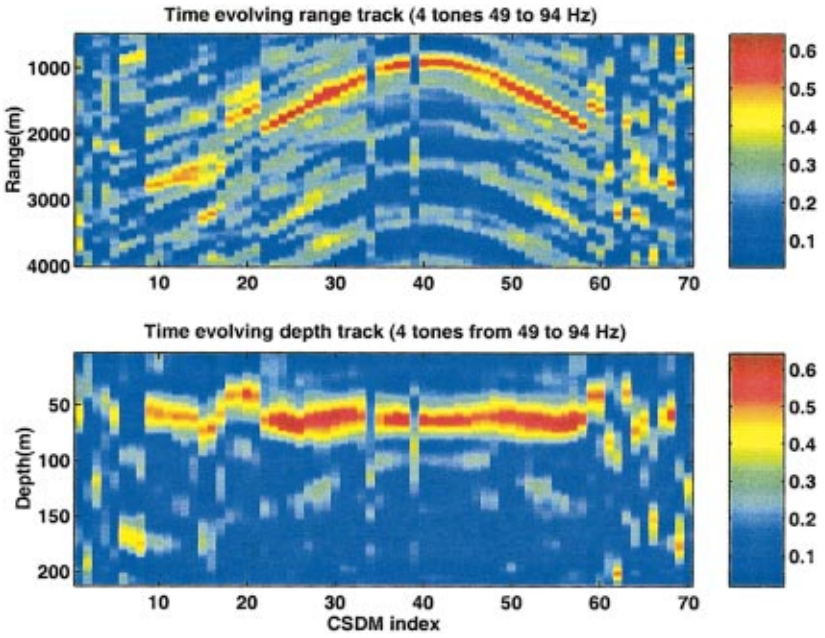


FIG. 10. MFP range and depth tracks for the second half of SWellEx-96 track S5. Results are integrated over the four frequencies (49, 64, 79, and 94 Hz) at which modes were derived from data in track S9. Full sets of modes were used, after performing polynomial interpolation on the wave numbers to get the missing modes (i.e., the modes drawn in white in Fig. 5).

Sec. III A, was used to extrapolate the data-derived modes from the span of the array to the entire water column. Only the subset of modes shown in Fig. 5 was used to generate these results. Note that, although the data-dependent shooting method produces wave numbers, we are not using these wavenumbers in order to show a qualitative comparison between incoherent (modes only) and coherent methods. With a fully spanning array, it would have been possible to obtain these results without a measured sound speed profile, which was only used to extrapolate the modes from the span of the array to the depth of the source.

Figure 10 shows coherent MFP results using modes and wave numbers calculated using the shooting method prior to matching a bottom model to the data-derived modes. The missing modes were calculated by fitting the wave numbers from the data-derived modes to a polynomial and using the polynomial to interpolate the missing wave numbers. Although such an interpolation is clearly *ad hoc* and could produce spurious results in other environments, in this environment, it served to complete the subsets of modes that had been extracted from data, and yielded better results than if we had only used the modes extracted from data. The tracks produced by this process extend only out to 2 km (it may be difficult to distinguish where these tracks drop out, because there are gaps in the source tones, but we have also processed the other half of this track where source ranges extend out to 9 km to corroborate these ranges).

Figure 11 shows coherent MFP results using modes and wave numbers derived using the optimization approach described in Secs. III and IV A. The improvement over the previous two processes is very evident, and it is clear that forcing the modes and wave numbers to be consistent with a bottom model resulted in significant processing gains, even if by all previous accounts it was a simplistic bottom model

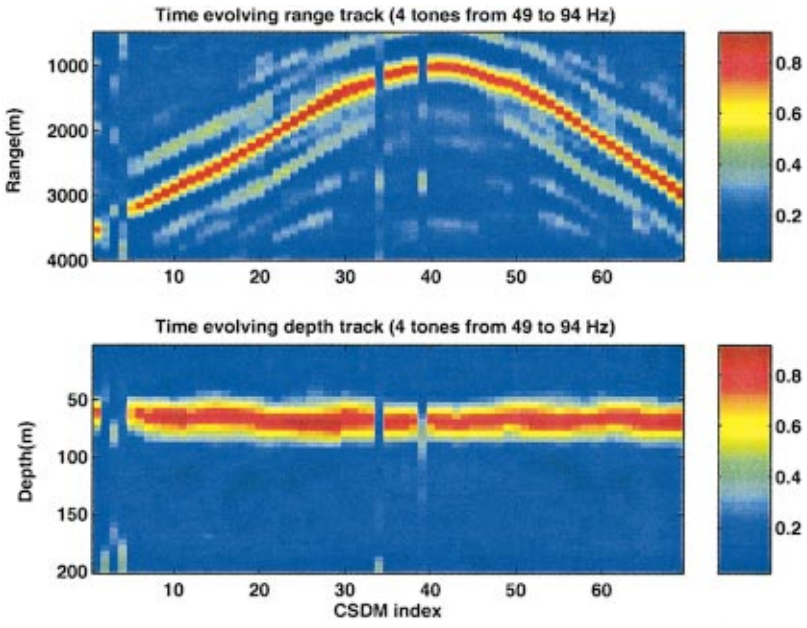


FIG. 11. MFP range and depth tracks for the second half of SWellEx-96 track S5. Results are integrated over the four frequencies (49, 64, 79, and 94 Hz) at which modes were derived from data in track S9. Full sets of modes were calculated using a normal mode model for the bottom half space matched to the data-derived modes, as described in Sec. II B.

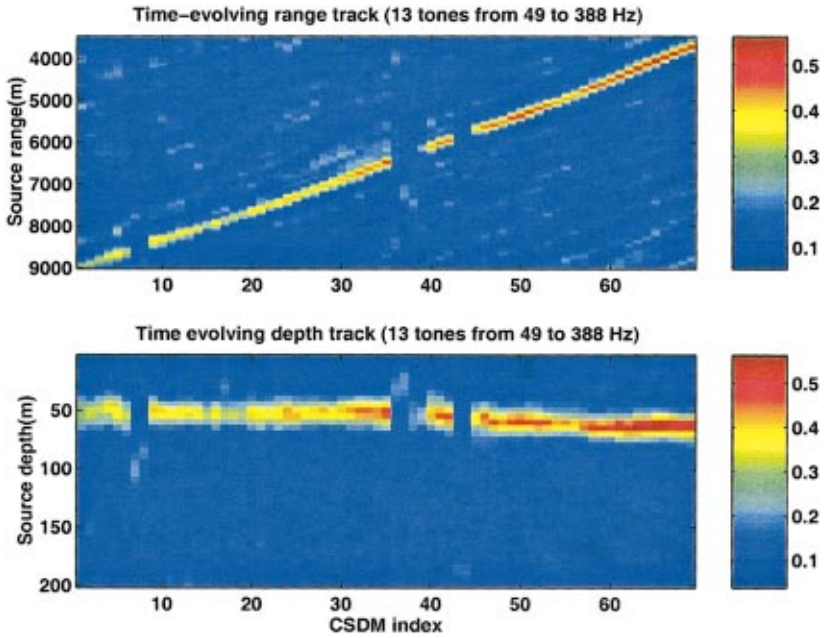


FIG. 12. MFP range and depth tracks for a deep source (at 60-m depth) from the first half of SWellEx-96 track S5. Thirteen frequencies were used. Modes and wave numbers were calculated using half-space bottom parameters resolved by matching data-derived modes.

(i.e., a half-space model compared to the three layer models that were measured and inverted for in other efforts).

Note that the shooting method was an essential part of both of the coherent MFP processes. The shooting method required a measured sound speed profile, but no *a priori* bottom properties were used in either process.

As mentioned above, in the third processor, resolving the bottom parameters enable modes and wave numbers to be calculated across the entire spectrum. There were two sources being towed during Track S5, a shallow source at 10-m depth emitting nine tones from 109 to 385 Hz, and a deep source at 60-m depth emitting 13 tones from 49 to 388 Hz. Modes and wave numbers were calculated for the measured sound speed profile and the half-space bottom model resolved by our optimization for all of these tones. To show the benefit of being able to calculate modes at all frequencies, we present results for both sources using all available tones. Figures 12 and 13 show MFP tracks for much longer ranges out to 9 km along Track S5a (the first half of Track S5) for both the deep and shallow sources, respectively, in which all available tones were processed. For the shallow source, all nine tones are at frequencies at which no modes had been extracted. For the deep source, 9 of 13 tones are at frequencies at which no modes had been extracted. The correlation values produced by MFP based on data-derived modes (without *a priori* bottom information) were nearly identical to those produced by MFP using modes calculated by a normal mode program (normal mode model documented in Ref. 16) using environmental information obtained in an earlier experiment (including the bottom information we did not use in our process based on data-derived modes).

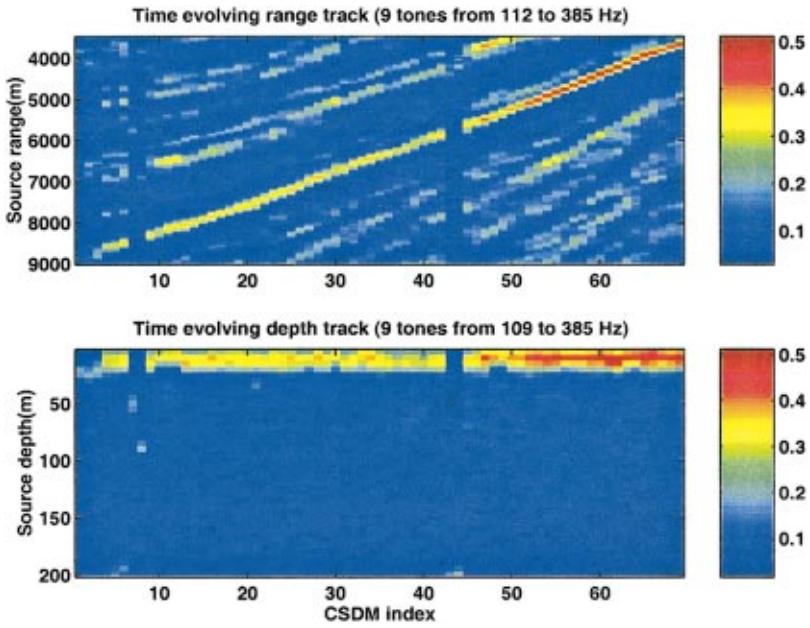


FIG. 13. MFP range and depth tracks for a shallow source (at 10-m depth) from the first half of SWellEx-96 track S5. Nine frequencies were used, all distinct from frequencies at which modes were derived from data. Modes and wave numbers were calculated using half-space bottom parameters resolved by matching data-derived modes.

V. SUMMARY AND CONCLUSIONS

We have discussed how and when it is possible to extract modes from data in Sec. II. Unfortunately, having modes without wave numbers only enables incoherent processes with only limited processing gain. However, using even incomplete sets of modes in conjunction with limited environmental information improved prospects considerably. In Sec. III, starting with a measured sound speed profile as *a priori* information (bottom properties were assumed to be unknown) and using data-derived modes limited to the span of the array, an optimization process was presented to obtain complete sets of modes sampled over the entire water column, their wave numbers, and the parameters of a half-space bottom model. Having the bottom parameters enabled modes and wave numbers to be calculated across the entire spectrum, not just at the frequencies at which modes were extracted from the data. In Sec. IV, this optimization process was demonstrated on data from SwellEx-96 experiment track S9. Subsets of modes at four frequencies being emitted by a towed source were recovered, and with the aid of a measured sound speed profile were used to obtain self-consistent modes, wave numbers, and bottom parameters. Using these bottom parameters, modes and wave numbers were calculated for tones over roughly three octaves (49–388 Hz). Most of these tones were distinct from the ones at which modes were initially extracted. These modes and wave numbers were used to construct replicas for a MFP process that was able to unambiguously resolve two sources at depths of 10 and 60 m out to ranges of 9 km. The resulting range and depth tracks, shown in Figs. 12 and 13, were consistent with the known source locations. Moreover, the MFP process based on limited environmental data (sound speed only) produced correlation values nearly identical to those produced by a MFP process using modes and wave numbers calculated by a normal mode model supplied with the best available environmental information.

ACKNOWLEDGMENTS

SWellEx-96 was planned and carried out jointly by NRaD/NCCOSC, the Naval Research Laboratory, and the Marine Physical Laboratory. Newell Booth (NRaD/NCCOSC) was SWellEx-96 Chief Scientist. Jim Murray (MPL) provided valuable information about the experiment data. Helpful suggestions also were provided by Michael Porter (SAIC) regarding the computational aspects of normal mode models. This work was supported by the Office of Naval Research, Code No. 321 US.

APPENDIX A: SPARSE ARRAYS AND SHORT ARRAYS

If the array spans the water column and adequately samples the modes, and we average our CSDM over an interval that contains a suitable source distribution (as discussed in Sec. II B), the eigenvectors of the CSDM will correspond to the normal modes of the ocean waveguide.

If the array is too sparse, the higher-order modes that have a higher spatial frequency content will not be adequately sampled and their higher spatial frequencies will alias into lower spatial frequencies. Even more fundamentally, with less elements than the number of modes in the environment, there will not be enough eigenvectors for there to be a one-to-one relationship between modes and eigenvectors.

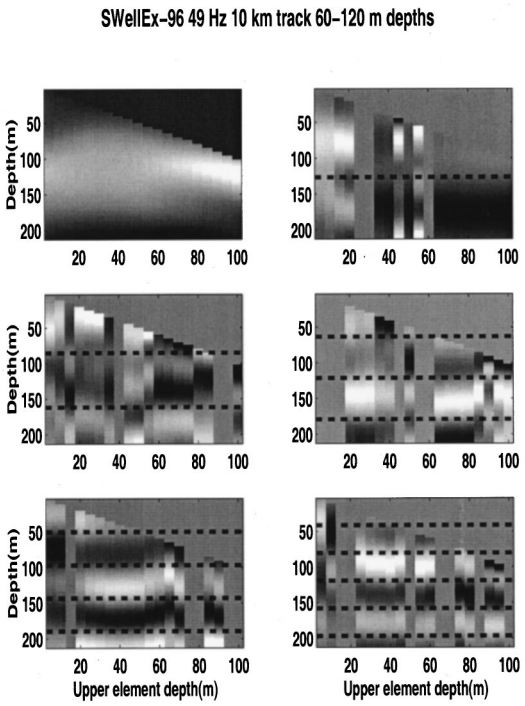


FIG. A1. Modes from eigenvectors of synthetic CSDMs from progressively shorter arrays (the lowest element at the bottom, the upper element depth varying along the horizontal axis). This simulation was of a 49 Hz source traveling over a 10-km track changing the depth from 60 to 120 m in the 215-m deep SWellEx-96 environment. Each subplot shows eigenvectors that match a particular mode, with six subplots showing modes 1–6. The black dashed lines show the zero crossings of the true modes. Some of the modes were not found at particular array lengths, as indicated by the gaps in the images for modes 3, 4, 5, and 6.

We have considered arrays that have enough elements sufficiently closely spaced that spatial aliasing of the modes is not an issue. This is the case in the experiment we show results from in Sec. IV. The SWellEx-96 array has a half-lambda spacing at 400 Hz. It has 64 elements, of which we have used 63. At 94 Hz, the highest frequency at which we extracted modes from data, there are 12 propagating modes, so clearly we have more eigenvectors than modes.

If the array does not span the water column, even with adequately sampled modes, the modes will be orthonormal over the entire water column but not necessarily over the interval spanned by the array. The eigenvectors by contrast are guaranteed to be orthonormal over this interval, since the CSDM is Hermitian. Therefore, since the eigenvectors are orthonormal and the modes are not, the two sets of vectors cannot be the same. This is a serious issue, since our results indicate that the method works, despite expectations to the contrary.

Why do we ultimately get results? It is a question of the degree by which the eigenvectors deviate from the modes with progressively shorter arrays. Apparently, the eigenvectors are enough like the modes that, when we apply our

process, which basically seeks modes that are closest to the eigenvectors and also consistent with a physical environment (measured sound speed and a bottom half-space), the modes and wave numbers of the prevailing environment are resolved.

Figure A1, the results of a simulation, shows how the modes remain represented in the eigenvectors of CSDMS, even with short vertical arrays. The modes-from-CSDM process was applied to a series of progressively shorter arrays. All arrays terminated at the bottom at 215 m. To construct progressively shorter arrays, the upper array element depth was varied from 5 to 100 m. A CSDM was synthesized at each of these vertical array lengths for a 49-Hz source traveling from 1 to 10 km in range and from 60 to 120 m in depth in the SWellEx-96 environment (see Sec. IV). Figure A1 contains six subplots, corresponding to the six modes at 49 Hz in this environment. The depth of the upper array element is indicated by the horizontal axis values, starting with the longest array on the left. Each column in the images shown corresponds to a different length array and contains the eigenvector that best matches the mode being depicted in that subplot. As the arrays get shorter, some of the higher-order modes are no longer found among the eigenvectors, and there are gaps in the modes at these shorter arrays. Only the part of the mode that overlaps the array can be captured by the eigenvector, so the image columns get shorter as the arrays get shorter moving from left to right in each subplot. To show the correspondence between the eigenvectors depicted in these images and the modes they estimate, the zero crossings of the normal modes in this environment are overlaid as black dashed lines. The eigenvectors capture the fundamental structure of the modes very well, even with as much as 100 m of the water column missing from the array.

The capacity for eigenvectors to represent the modes with progressively shorter arrays will undoubtedly depend on the particular environment. In the shallow water summer profile we were in, the thermocline at the surface traps most of the acoustic energy in the isovelocity layer below, creating a shorter effective waveguide concentrated in the lower part of the water column, where our array is.

We have performed many simulations to assess the feasibility of extracting modes from the eigenvectors of a CSDM under varying conditions, and have found that spanning a larger portion of the water column consistently improves the prospects for this method. Certainly, with a fully spanning array, not only would the modes be estimated over the entire water column, but a larger percentage of the modes would be extracted and with greater accuracy. This would greatly improve the capabilities of a modes-only depth estimator.

APPENDIX B: ARRAY TILT

Processing our experiment data, we found that uncompensated array shape changes such as tilt can seriously impact not only MFP, but also the culling of modes from data via an eigenvector analysis of the CSDM. When individual array elements are not aligned along the vertical with their peers, at least in the direction of the source, the individual mode contributions to the received field will have phase shifts proportional to their wave number and the horizontal offset in the direction of the source. Because this phase offset depends on the mode, the array element, and the direction of the source, it is not immediately clear how to correct for such an offset.

Because the phase shift caused by array tilt varies across the modes, a complete compensation would require the signal to be separated into its modal components. This is not possible unless the modes are already known. However, because the spread between the wave numbers is so much smaller than the average wave number, it is possible to use the average wave number to compensate for the array tilt across all modes, at least for small deviations from vertical. This is enough to bring all the phase shifts close enough to zero that they do not seriously impact the correspondence between the eigenvectors and the modes.

[1] H. P. Bucker, ''Use of calculated sound fields and matched-field detection to locate sound sources in shallow water,'' J. Acoust. Soc. Am. **59**, 368–373 (1976).
[2] A. B. Baggeroer, W. A. Kuperman, and H. Schmidt, ''Matched field processing: Source localization in correlated noise as an optimum parameter estimation problem,'' J. Acoust. Soc. Am. **83**, 571–587 (1988).
[3] A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky, ''An overview of matched field methods in ocean acoustics,'' IEEE J. Ocean Eng. **18**, 401–424 (1993).
[4] S. N. Wolf, ''Experimental determination of modal depth functions from covariance matrix eigenfunction analysis,'' J. Acoust. Soc. Am. **81**, S64 (1987).
[5] T. C. Yang and R. Kille, ''Broadband mode decomposition,'' J. Acoust. Soc. Am. **88**, S182 (1990).
[6] S. N. Wolf, Douglas K. Cooper, and B. J. Orchard, ''Environmentally adaptive signal processing in shallow water,'' in *Oceans '93, Engineering in Harmony with Ocean Proceedings*, Piscataway, NJ, 1993, IEEE, held 18–21 October in Victoria, BC, Canada, Vol. 1 of 3.
[7] M. Badiey, I. Jaya, and A. H.-D. Cheng, ''Shallow-water acoustic/geoacoustic experiments at the New Jersey Atlantic generating station site,'' J. Acoust. Soc. Am. **96**, 3593–3604 (1994).
[8] P. Hursky, W. S. Hodgkiss, and W. A. Kuperman, ''Extracting modal structure from vertical array ambient noise data in shallow water,'' J. Acoust. Soc. Am. **98**, 2971 (1995).
[9] T. B. Neilsen, E. K. Westwood, and T. Udagawa, ''Mode function extraction from a VLA using singular value decomposition,'' J. Acoust. Soc. Am. **101**, 3025 (1997).
[10] E. C. Shang, ''Source depth estimation in waveguides,'' J. Acoust. Soc. Am. **77**, 1413–1418 (1985).
[11] T. C. Yang, ''A method of range and depth estimation by modal decomposition,'' J. Acoust. Soc. Am. **82**, 1736–1745 (1987).
[12] G. R. Wilson, R. A. Koch, and P. J. Vidmar, ''Matched mode localization,'' J. Acoust. Soc. Am. **84**, 310–320 (1988).
[13] K. D. Heaney, ''Inverting for source location and internal wave strength using long range ocean acoustic signals,'' Ph.D. thesis, University of California San Diego, 1997.
[14] P. Hursky, W. S. Hodgkiss, and W. A. Kuperman, ''Comparison of mode-shapes-only source depth estimation with conventional matched field source localization,'' J. Acoust. Soc. Am. **105**, 1105 (1999).
[15] K. D. Heaney and W. A. Kuperman, ''Very long-range source localization with a small vertical array,'' J. Acoust. Soc. Am. **104**, 2149–2159 (1998).
[16] M. B. Porter, ''The KRAKEN normal mode program,'' SACLANT Undersea Research Centre, La Spezia, Italy, 1991.
[17] P. A. Baxley, ''Matched-field replica model optimization and bottom property inversion in shallow water,'' Master's thesis, University of California, San Diego, 1997.
[18] W. A. Kuperman and F. Ingenito, ''Spatial correlation of surface generated noise in a stratified ocean,'' J. Acoust. Soc. Am. **67**, 1988–1996 (1980).
[19] M. J. Buckingham, ''A theoretical model of ambient noise in a low-loss, shallow water channel,'' J. Acoust. Soc. Am. **67**, 1186–1192 (1980).
[20] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computa-*

tional Ocean Acoustics (AIP Press, Woodbury, NY, 1994).
[21] G. B. Smith, ''Through the sensor environmental estimation,'' J. Acoust. Soc. Am. **101**, 3046 (1997).
[22] W. A. Kuperman, M. B. Porter, J. S. Perkins, and R. Evans, ''Rapid computations of acoustic fields in three-dimensional ocean environments,'' J. Acoust. Soc. Am. **89**, 125–133 (1991).
[23] G. Carpaneto and P. Toth, ''Algorithm 548: Solution of the assignment problem [*H*],'' ACM Trans. Math. Softw. **6**, 104–111 (1980).
[24] N. O. Booth. ''SWellEx-96 preliminary data Report,'' NRaD/NCCOSC Code 881, San Diego, CA, May 1996.
[25] G. L. D'Spain, J. J. Murray, W. S. Hodgkiss, N. O. Booth, and P. W. Schey, ''Mirages in shallow water matched field processing,'' J. Acoust. Soc. Am. **105**, 3245–3265 (1999).
[26] N. O. Booth, P. A. Baxley, J. A. Rice, P. W. Schey, W. S. Hodgkiss, G. L. D'Spain, and J. J. Murray, ''Source localization with broad-band matched-field processing in shallow water,'' IEEE J. Ocean Eng. **21**, 393–401 (1996).

Acoustic transmission across a roughened fluid–fluid interface

Raymond Lim
Coastal Systems Station, Code R22, 6703 West Highway 98, Panama City, Florida 32407-7001

Iris C. Paustian and Joseph L. Lopes
Coastal Systems Station, Code R21, 6703 West Highway 98, Panama City, Florida 32407-7001

(Received 28 July 2000; accepted 30 November 2000; revised 19 December 2000)

A set of tank experiments was performed to investigate acoustic transmission across a roughened fluid–fluid interface with the intention to test heuristic Bragg scattering predictions used to explain observations of anomalous transmission in field experiments. In the tank experiments, two immiscible fluids (vegetable oil floating on glycerin) formed the layers. Small polystyrene beads were floated at the interface to simulate roughness. An array of hydrophones placed in the bottom layer (glycerin) was used to measure the acoustic levels transmitted across the interface. This array was also employed as a beamformer to determine the apparent angle and sound speed of the scattered signals. Data were acquired at subcritical grazing angles in the frequency range of 100–200 kHz for three different bead diameters and for various configurations in which the locations of the beads floating on the interface were varied. Results of these measurements demonstrated that a significant amount of acoustic energy can be scattered into the bottom layer by beads floating at the interface. The scattered levels increased with increasing bead diameter. However, discrepancies occurred between observed propagation properties and the Bragg predictions. By comparing the processed tank data to a computer simulation of the same it was determined that these discrepancies are a consequence of near-field reception of the scattering by the bead array and ignoring the directionality of the scattering by the beads. Consequences to observations made in field experiments are discussed. [DOI: 10.1121/1.1350450]

PACS numbers: 43.30.Hw, 43.30.Ma [DLB]

I. INTRODUCTION

Recent studies of acoustic transmission into fast, ocean sediments[1–7] reveal in-sediment sound pressure levels that are unexpectedly high when the field incident on the bottom approaches at a shallow grazing angle. The enhancement is especially pronounced when the angle is shallower than the ''critical'' angle below which the Green function for a flat bottom predicts high reflection and very little transmission. The most notable mechanisms proposed to explain this phenomenon are the excitation of Biot's slow wave in the poroelastic bottom,[4] scattering of sound into the bottom by surface roughness,[8–11] and scattering of the evanescent wave by volume inhomogeneities.[12,13] Due to difficulties in assessing properties of ocean sediments and in performing ocean acoustics measurements, the appropriate choice of mechanism for modeling the transmission of sound is still somewhat uncertain.

In the present paper, a set of tank experiments is described that are meant to test the roughness scattering mechanism of transmission. Surface roughness effects are isolated from effects attributed to bottom poroelasticity and bottom inhomogeneity by modeling the bottom as a homogeneous fluid. Therefore, the experiments involve a tank filled with two immiscible fluids (vegetable oil over glycerin), where floating polystyrene beads of various sizes at the intervening interface control the roughness of the interface. With the specified combination of fluids, a critical grazing angle of about 40° results for the flat interface. Sound transmitted through the bead layer into the lower fluid is recorded at a hydrophone array and analyzed using techniques analogous to those reported by Chotiros and collaborators.[4,14]

A number of assumptions are implicit in the use of this setup. First, it is assumed that the surface scattering effects proposed to occur at lower frequencies in field measurements can be scaled to the dimensions of the tank by raising the frequency. Second, although anomalous sound pressure levels have most often been observed in transmission into sandy sediments, the sound speed of these sediments is generally lower and the density generally higher than that of glycerin. Therefore, it is assumed that the dominant effects due to roughness scattering are not affected by variations in these parameters. In this regard, it is also assumed that mode converted shear waves produced on transmission into sand sediments do not contribute significantly to these scattering effects. Given the low shear speeds characteristic of sand sediments, this is not expected to be unreasonable. Finally, the use of polystyrene beads to mimic roughness is assumed to be feasible even though the material parameters of polystyrene may differ somewhat from those of either oil or glycerin. The rationale is that scattering by roughness features is the essential ingredient for producing enhanced sound transmission. Varying the bead size relative to the acoustic wavelength is expected to produce scattering levels associated with typical roughness features and any anomalous transmission due to scattering will be exhibited.

This paper is organized as follows. In the next section, a description of the experimental setup is provided. In Sec. III, a model that provides a qualitative but realistic representation of tank data is described. In Sec. IV, results of the tank measurement and model are presented and compared to the

TABLE I. Properties of vegetable oil, glycerin, and polystyrene.

Material	Density (g/cm^3)	Speed of sound (m/s)
vegetable oil	0.992	1455
99.5% glycerin	1.261	1905
polystyrene	1.05	2309 at 20 °C

predictions based on Bragg scattering. In Sec. V, the model is used to compare predictions of subcritical transmission with that reported from a field experiment.[4] The last section is a summary of relevant points.

II. EXPERIMENTAL SETUP

Prior to conducting transmission experiments, time-of-flight and attenuation measurements of vegetable oil and glycerin were made in a small 88-gallon test tank using four International Transducer Corporation (ITC) 1089D hydrophones at preset positions. One hydrophone was driven as a projector while the others were employed as receivers. Measurements were made in the 100–300 kHz frequency range. When compared to data collected with this setup in fresh water, the attenuation measurements indicated that there were no differences observed in the signal levels collected with the vegetable oil and there was, at most, a 10% reduction in the signal levels acquired with the glycerin. Thus little signal attenuation due to absorption by either of these fluids is expected. The time-of-flight measurements determined the sound speeds of vegetable oil and glycerin to be 1455 m/s and 1905 m/s, respectively, with no noticeable dispersion. This results in a 40° critical grazing angle for transmission from oil to glycerin. Table I summarizes important properties of the vegetable oil, glycerin, and polystyrene beads.

The setup for the transmission experiments is depicted schematically in Fig. 1. A 1.22-m-deep×1.22-m-wide×2.13-m-long wooden tank was filled with the two fluids, first with glycerin to a depth of 48.5 cm and then with a 69.1-cm layer of vegetable oil. Since these fluids are immiscible, this two-layer system remains stable and polystyrene beads, which have a mass density between that of either fluid, can float at the oil/glycerin interface to simulate roughness. The distribution of the polystyrene beads was confined in an area about 65-cm long and 60-cm wide over a hydrophone array located in the glycerin. Confinement was maintained by bounding this area of the interface with light nylon monofilament line. Initially, the beads were distributed in a square ''lattice'' arrangement by dropping them through holes of a template resting above the tank. The beads were spaced approximately 1.5 cm along the length and 1.5 cm along the width of the tank.

An array of four ITC 1089D hydrophones was placed in the glycerin. The hydrophones were supported by Delrin tubes that extended from aluminum braces attached to the sides of the tank. The tubes were machined to hold each hydrophone securely in place and to lead the transducer cable to a side of the tank in a manner that minimized reverberation problems and kept the sensor's sensitivity pattern up and/or towards the source. (When calibrated, the hydrophones exhibited a flat response in the 20–250 kHz frequency band and were omnidirectional aside from a small dip in sensitivity in the direction of the cable.) Two additional ITC 1089D hydrophones that could be removed during data collection were also employed. One of these hydrophones was placed at mid-depth in the oil and 1 m from the source; this hydrophone was used to collect signals transmitted by the source. The second hydrophone was attached so that the sensing element lay on the oil/glycerin interface at 5-cm from the mid-width of the tank and approximately 76.8 cm from the back of the tank. It was used to help aim the beam of the source transducer. Relative to an origin placed on this ''surface'' hydrophone, the (x,y,z) positions of the four hydrophones in the array were given by: (1) (5.0, −3.2, −4.0) cm; (2) (14.1, −9.1, −7.0) cm; (3) (14.4, −5.8, −11.7) cm; and (4) (14.5, −7.49, −19.6) cm. Here, directions are specified such that, when looking down the length of the tank from the source to the receive array, the x direction is forward, the y direction is from right to left, and the z direction is up. These coordinates were determined by using a computational code that resolved uncertainties in the actual

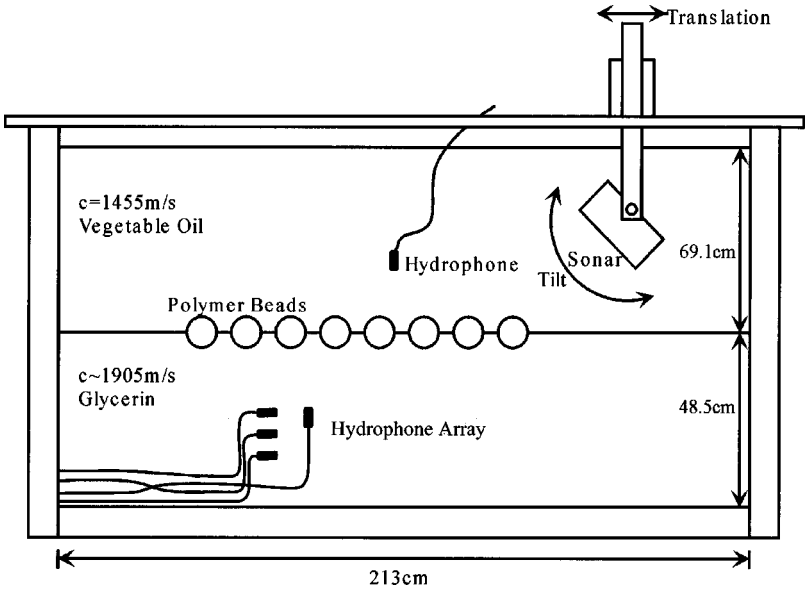


FIG. 1. Schematic representation of experimental setup.

FIG. 2. Photograph of hydrophone array in the tank.

hydrophone positions by optimizing a fit of these coordinates to expressions for the propagation distances between hydrophones. Inputs to the code included the measurable coordinates of each hydrophone after the array was placed in the tank as well as propagation times obtained from time-of-flight measurements from one hydrophone to the other three. The hydrophones and their corresponding mounting brackets were arranged to minimize any scattering from one to another. A photograph of the hydrophone array in the tank is shown in Fig. 2.

The source is a circular piston transducer with a 5.08-cm-diam aperture. A geared housing, constructed to hold the source, allowed the transducer face to be tilted manually. A pendulum tilt sensor was attached to the housing to allow the tilt angle of the projected beam to be monitored during data collection. The transducer housing with the attached tilt sensor was suspended via a telescoping arm from aluminum tracks attached to the top of the tank. The aluminum tracks allowed translation of the source along the length of the tank while the telescoping arm permitted the source to be positioned at selected heights between 35 and 62 cm above the oil/glycerin interface. To minimize wall reverberations in the present experiments, the source was not operated much below 100 kHz; at this frequency, a conical beam is produced with a 3-dB-down angular width of 15.5°. Measurements were made in the 100–200 kHz band. A photograph of the transducer mounting is shown in Fig. 3 and part of the aluminum tracks that the mounting was suspended from are shown at the upper left of Fig. 2.

A block diagram of the drive and receive electronics is shown in Fig. 4. The projector was driven using 0.1-ms duration sinusoidal pulses that had a 0.04-ms taper on the leading and trailing edge of the pulse. This taper minimized ringing in the waveform generated by the source. An ENI Model 240L power amplifier amplified the generated signals and the amplified signals were used to drive the projector. Signals acquired by the hydrophones were conditioned using Wavetek brickwall amplifier/filters along with waveform multiplier circuits when it was desired to record field intensities; the multiplier circuits squared the signals acquired by the hydrophones. In all instances the signals were high passed at a frequency of 600 Hz. The conditioned signals were then ping averaged and digitized at a sample rate of 5 MHz using a Phillips PM3394 digital oscilloscope. The digitized data were recorded with the aid of a computer.

Data were collected at shallow grazing angles in the

FIG. 3. Photograph of transducer mounting structure.

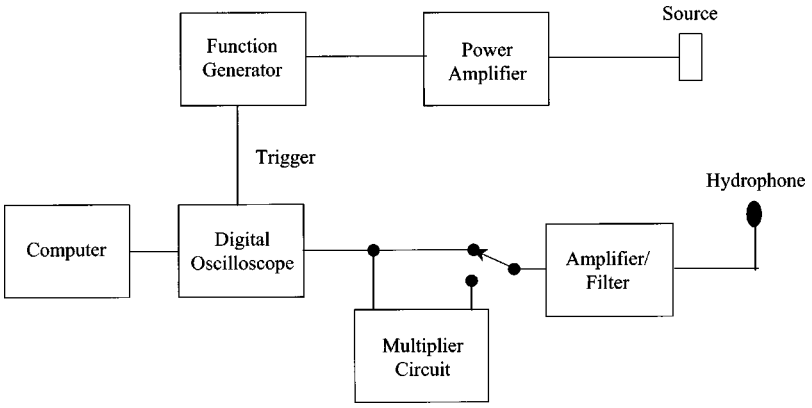


FIG. 4. Block diagram of the drive and receive electronics.

frequency range of 100–200 kHz for bead diameters of 6.35 mm, 4.76 mm, and 2.38 mm. The source was placed approximately 35 cm above the oil/glycerin interface and was positioned at the end of the tank furthest from the hydrophone array. This arrangement yielded an 18° grazing angle with the axis of the main projected beam crossing an area of the interface directly over the hydrophone array. This setup also insured that the incident field on the interface directly over the hydrophone array was in the far field of the source. Here the far field is estimated with an $A/2\lambda$ criterion, where A is the radiating area of the source and λ is the wavelength of the acoustic beam in vegetable oil.

Data were collected two distinct ways. First, full waveform data were acquired with and without the beads by vertically tilting the source in one-degree increments from a grazing angle of 14°–37°. At each angle, data were recorded for source pulses configured as described above and centered on the following frequencies: 80, 100, 115, 130, 145, 160, 180, 200, and 250 kHz. Waveforms at each angle and center frequency were averaged over 64 pings and the beads were only deployed in the square lattice arrangement. These tilt runs were used as an initial assessment of the level of roughness-induced transmission enhancement possible as a function of frequency, time, and grazing angle.

In the second set of measurements, the grazing angle was held constant at 18° and data were acquired for different bead arrangements within the 65-cm long and 60-cm wide deployment area. While the beads were initially arranged as a square lattice, subsequent arrangements were produced by manually stirring the beads at the oil/glycerin interface using a 1-m long, $\frac{1}{4}$-in diameter wooden dowel. Data collection proceeded after the beads and the interface settled. The data were acquired in two formats: as full waveform signals and as squared (intensity) signals via the multiplier circuitry, each signal corresponding to a different bead configuration. Of the full waveform signals, 30 waveforms at each hydrophone corresponding to 30 bead configurations and associated with source pulses centered on 100, 145, 200, and 250 kHz were recorded. Each waveform was averaged over 64 pings received by the hydrophones. Of the squared signals, 96 sets associated with 96 different bead configurations were collected by the Phillips PM3394 oscilloscope and averaged. The averaged intensity history was then recorded. Due to difficulties in making sure the interface had settled between pings, the squared signals were collected for 100-kHz centered source pulses only. Both of these types of data were processed to investigate the direction of the signal propagating in the glycerin.

III. MODEL

Although the tank measurements are intended to test predictions based on Bragg scattering, the predictions are formulated with several assumptions that are not maintained in our tank. For example, the roughness is not of infinite extent, reverberation from tank walls may not be insignificant, the incident field is not a monotone (cw) plane wave, and, due to the low attenuation of the glycerin, the sound scattered from the roughness may have more than downward propagating components at the receiver. A simulation of the tank measurements was performed to provide a qualitative but realistic representation of data, accounting for tank-wall reverberation, effects due to the dimensions of the roughness and the proximity of the source, the incident pulse shape, and the placement of the receive hydrophones. The simulation is created by using the source, hydrophone, and bead locations to determine time delays for pulses propagated from the source to each bead to the hydrophone. These pulses are superimposed in the time domain at the hydrophone position with the appropriate time delay and propagation loss factors imposed to create the simulated signal. Including contributions from mirror images of the beads and the source accounts for reverberation from tank walls.

The simulation may be represented by the following formulas:

$$\langle I(\mathbf{r},t)\rangle = \frac{1}{N}\sum_{j=1}^{N} p_j(\mathbf{r},t)^2, \tag{1}$$

where

$$p_j(\mathbf{r},t) = \sum_{n=1}^{N_b} \frac{2A(\xi)|\mathbf{r}_s|J_1(k_1 a \sin\vartheta)e^{-\mathrm{Im}(k_2)|\mathbf{r}-\mathbf{r}_b(n,j)|}}{N_b k_1 a \sin\vartheta |\mathbf{r}-\mathbf{r}_b(n,j)||\mathbf{r}_b(n,j)-\mathbf{r}_s|} \times S\left(t - \frac{|\mathbf{r}-\mathbf{r}_b(n,j)|}{\mathrm{Re}(c_2)} - \frac{|\mathbf{r}_b(n,j)-\mathbf{r}_s|}{c_1}\right). \tag{2}$$

In Eq. (1), $\langle I(\mathbf{r},t)\rangle$ is the intensity history at $\mathbf{r}$ averaged over an ensemble of N roughness configurations and $p_j(\mathbf{r},t)$ is the pressure history at $\mathbf{r}$ for the jth configuration. Each roughness configuration results from using a random number generator to specify uniformly distributed bead positions within

a specified area. The pressure history is specified by Eq. (2) as a sum over N_b copies of the incident pulse history, $S(t)$, where the time argument has been shifted to account for propagation to the nth bead in the simulation and then from that bead to $\mathbf{r}$. $S(t)$ is assumed known and, in our simulations, is a numerically generated copy of the experimental pulse shape. The amplitude of the pulse history for each bead is also adjusted relative to the others by multiplying by some extra factors. An exponential factor accounts for medium attenuation in the glycerin and is specified by adding an imaginary part to the wave number k_2 in glycerin. Two factors take into account spherical spreading loss as a consequence of propagation to each bead and from there to $\mathbf{r}$. A factor involving a cylindrical Bessel function of the first kind J_1 adjusts the field amplitude incident on each bead to simulate a circular piston source.[15] For this source, the amplitude at each bead depends on the wave number in the oil layer k_1, the radius of the source aperture a, and the angle ϑ from the main axis of the piston to the vector $\mathbf{r}_b(n,j)-\mathbf{r}_s$, where $\mathbf{r}_b$ and $\mathbf{r}_s$ specify the bead and source positions, respectively. Finally, an angular scattering amplitude factor $A(\xi)$ is used to adjust the pulse history of each bead according to the angle ξ between the vectors $\mathbf{r}_b(n,j)-\mathbf{r}_s$ and $\mathbf{r}-\mathbf{r}_b(n,j)$. Thus $A(\xi)=1$ for an omnidirectionally scattering bead. For beads that scatter with more complicated angular patterns a look-up table of the amplitude-angle dependence was generated, although noncylindrically symmetric patterns were not considered.

Although the simulation described above is meant to be qualitative due to assumptions inherent in its use, a few of the more significant assumptions are worth mentioning. First, pulse dispersion effects are ignored and when medium attenuation is included, it is kept constant over the bandwidth of a given pulse. The angular factor $A(\xi)$ is also assumed to describe the scattering pattern of the beads over the entire bandwidth of the pulse. Both of these assumptions can affect the fidelity of transmission spectrum simulations because attenuation and scattering patterns can have significant frequency dependence. Modifications to $A(\xi)$ as a consequence of refraction at the oil/glycerin interface are also ignored. Another assumption is that multiple scattering can be ignored, both between the beads and between the beads and the oil/glycerin interface. Multiple scattering can be expected to lengthen the pulse histories created in the simulation. The consequences of these assumptions will be discussed in the ensuing analysis.

IV. BRAGG PREDICTIONS AND ANALYSIS OF RESULTS

The tank and simulation data are studied to assess the enhanced transmission at shallow grazing angles based on Bragg theory. Prior to these studies, the origins of the basic predictions of Bragg scattering are first reviewed.

A. Review of predictions based on Bragg scattering

The predictions based on Bragg scattering can be deduced from the simple picture given in Fig. 5. A broadband pulse is incident at a grazing angle α on a patch of surface with small-scale roughness. When translated into the frequency domain, the dominant Bragg wave vector of the field scattered into the bottom by the roughness is found by summing the wave vector of the incident field with the wave vectors representing the dominant spatial components of the roughness. Assuming roughness characterized by a single spatial frequency $1/\lambda_r$ (i.e., a regular lattice with spacing λ_r), the dominant transmitted acoustic frequency f_d is predicted by requiring that this component of the scattered field add in phase in the bottom medium as the incident pulse travels past each lattice point. This suggests that the spacing of the wave fronts propagating at the angle β away from the patch of roughness should correspond to the wavelength of the dominant Bragg component in that direction. Assuming waves that scatter omnidirectionally from each lattice point, this leads to the following expression for the dominant frequency

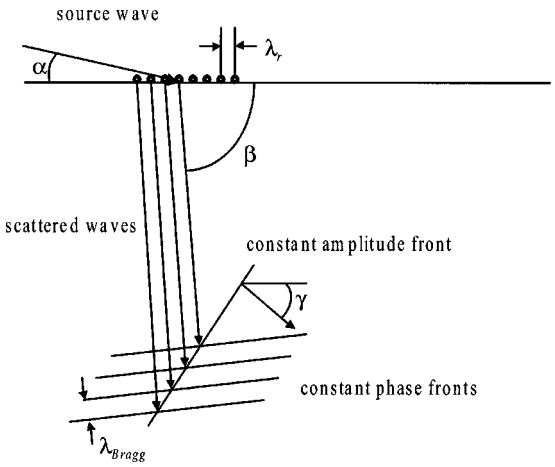


FIG. 5. Depiction of Bragg scattering.

$$f_d=\left[\frac{\lambda_r\cos\alpha}{c_{\text{oil}}}-\frac{\lambda_r\cos\beta}{c_{\text{gly}}}\right]^{-1}, \tag{3}$$

where c_{oil} and c_{gly} are the oil and glycerin sound speeds, respectively.

In regard to transmission of acoustic energy, it is noted that the frequency components predicted to scatter into the glycerin layer by Eq. (3) do so without regard to the grazing angle α being below the critical grazing angle that characterizes a flat interface. Therefore, enhanced acoustic penetration at angles shallower than the critical grazing angle can be expected if roughness is present. Also, for a receiver placed well below the evanescent region of the interface it is usual to assume the most energy would be detected directly beneath the rough patch due to lower propagation losses. By setting $\alpha=18°$, $\beta=90°$, $\lambda_r=1.5$ cm and using the parameters from Table I, the frequency estimated by Eq. (3) to dominate the energy detected at a hydrophone submerged in the glycerin below an illuminated rough patch is $f_d=102$ kHz.

Although energy will be scattered from the roughness overhead, the waves detected are not expected to be homogeneous. The constant phase fronts will not be aligned with the contours of the transmitted pulse because the transmitted pulse is a superposition of waves excited at different scattering centers as the source field propagates past them. For constant phase wave fronts propagating in a direction deter-

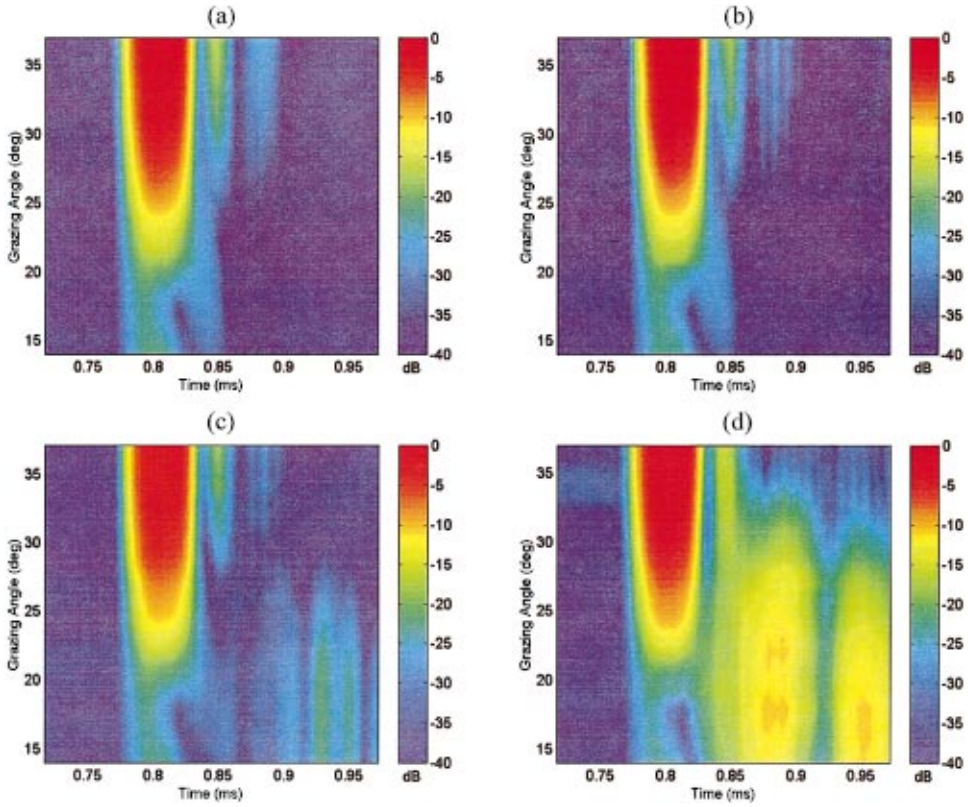


FIG. 6. Waterfall plot for $f = 100$ kHz. Beads deployed: (a) none, (b) 2.38 mm, (c) 4.76 mm, (d) 6.35 mm.

mined by β, the constant amplitude fronts defining the transmitted pulse at a given time are tilted by an angle γ that was shown by Moe[10] to be given by

$$\tan(\beta-\gamma)=\left(\frac{c_{\text{gly}}\cos\alpha}{c_{\text{oil}}}-\cos\beta\right)\Big/\sin\beta. \tag{4}$$

Since the constant phase fronts are propagating with the speed c_{gly}, the constant amplitude fronts will appear to propagate at the reduced speed

$$c_{\text{ca}}=c_{\text{gly}}\cos(\beta-\gamma). \tag{5}$$

For $\beta=90°$ (the case considered most often[8–11]) and $\alpha=18°$ this leads to a tilt of about $\gamma=39°$ from the interface for the propagation direction of the constant amplitude front and an apparent speed of $c_{\text{ca}}=1193$ m/s for this front.

The estimates for f_d, γ, and c_{ca} constitute the basic predictions made in regard to Bragg theory for acoustic energy scattered by roughness into a fast bottom layer. The goal of the subsequent analysis will be to examine the tank data for enhanced subcritical transmission of acoustic energy and test the predictions for f_d, γ, and c_{ca} associated with this transmission. By doing so and studying when and why discrepancies with these predictions can exist, the transmission data from field measurements can be better understood.

B. Assessment of enhanced subcritical transmission

The tank data used to assess enhanced subcritical transmission are displayed in a waterfall format. This format is a plot of processed ping trace data versus time for successive source tilts, with the tilt measured in terms of grazing angle. The data are processed by enveloping the recorded waveform, normalizing with respect to the maximum level obtained with deployment of the largest beads, converting the amplitude for each point into a decibel (dB) level, and mapping the resulting levels onto a color scale. For each set of processed data over the entire collection of angles, the maximum amplitude level is determined and assigned the color red; this level is referenced as 0 dB. The lowest displayed level in the waterfall plots corresponds to the color blue and is 40 dB below the maximum level.

Figures 6, 7, and 8 contain waterfall plots at 100, 145, and 200 kHz, respectively, showing the impact of the beads (i.e., a rough interface configured as a square lattice) on the acoustic energy transmitted into the glycerin layer. Each figure displays processed transmission data corresponding to no beads as well as with the three different bead sizes. In addition, all of these figures refer to signals recorded with the hydrophone located 11.7 cm below the oil/glycerin interface. In each figure, the processed data is displayed in the time window between 0.72 and 0.97 ms. This window is a time interval containing scattering effects due solely to the beads as observed at the interface hydrophone; i.e., the period of time between arrival of the transmitted signal and tank wall reverberation.

Several signals appear at various times and grazing angles in these figures. One signal is seen in all four plots of each figure at a ping time beginning at about 0.78 ms and extending to around 0.83 ms, except at 145 kHz where it extends to about 0.87 ms. The amplitude of this signal increases with increasing grazing angle, reaching a maximum level at about 41° in data not shown in the plots. At this angle, just above the critical grazing angle, the acoustic beam can reach the recording hydrophone by refraction across the oil/glycerin interface. Therefore, this first signal is a refracted

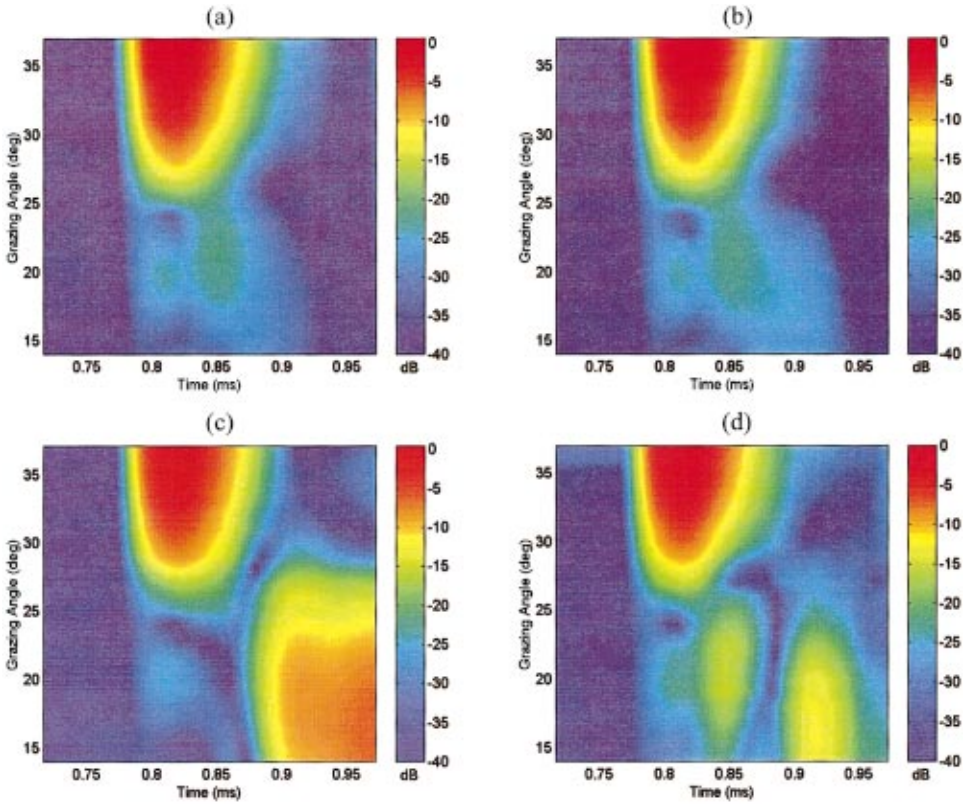


FIG. 7. Waterfall plot for $f = 145$ kHz. Beads deployed: (a) none, (b) 2.38 mm, (c) 4.76 mm, (d) 6.35 mm.

wave that appears at angles shallower than critical by virtue of the width of the main acoustic beam. Note that, except at 145 kHz, the duration of this wave appears shorter than the 0.1-ms duration of the source drive signal because of the 0.04-ms taper applied to each end of the pulse to reduce source ringing. Some spurious weak signals following the main refracted pulse after about 0.84 ms and above 27° appear nevertheless and are possibly an artifact of the poor response of the source. (The spurious signals could also be caused by reverberation among the hydrophone elements or their mounts but this was not seen to be a problem in preliminary tests of the experimental setup.) At 145 kHz, the

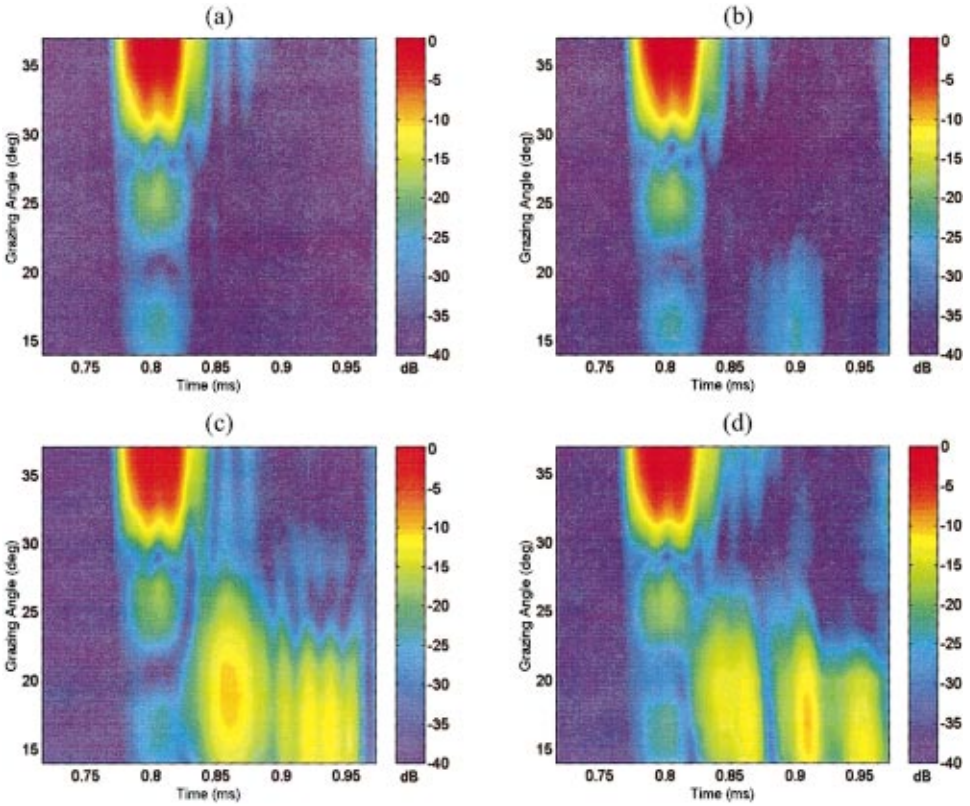


FIG. 8. Waterfall plot for $f = 200$ kHz. Beads deployed: (a) none, (b) 2.38 mm, (c) 4.76 mm, (d) 6.35 mm.

refracted wave is stretched because the source has a resonance near this frequency, which caused it to ring beyond its cutoff time. As frequency increases, oscillations in the decay of the refracted energy with decreasing grazing angle become evident due to refraction of the source's side lobes. This was confirmed with additional data showing this energy exhibited the correct main lobe beamwidth, the proper angular location of the side lobe, and the correct amplitude of side lobe levels for the source used in this experiment. The refracted side lobe pulses often appear distorted, which we attribute to angular dispersion of the signal.

For ping times greater than 0.85 ms and grazing angles less than 25°, little acoustic energy is observed in those waterfall plots where no beads or 2.38-mm-diam beads are deployed. However, acoustic transmission into the glycerin at grazing angles between 14° and 22° increases significantly with bead size when the 4.76-mm or the 6.35-mm-diam beads are floated at the interface. The enhanced energy appears to be well extended in length compared to the transmitted pulse. In some instances, the observed signal levels exceed the background noise floor of the no-bead case by almost 40 dB. This energy is clearly a consequence of scattering from the beads into the glycerin. The angular range of enhancement is consistent with an 18° grazing angle required for the projected beam to illuminate an area of the interface directly over the hydrophone array and the timing is consistent with propagation of the scattered energy directly from the beads to the recording hydrophone.

The results of these transmission measurements may be compared with other reported work. In a laboratory experiment, Simpson and Houston[5] showed similar increases in subcritical transmission levels for a deliberately roughened water-sediment interface when compared to an interface that was carefully scraped flat. The authors estimated the root-mean-square interface roughness to be between 1 and 2 cm, which is larger than the roughness associated with the largest diameter beads used in our transmission measurement. In addition, enhanced subcritical transmission into sediments due to interface roughness has been predicted in modeling by Thorsos *et al.*[8,9] In their model, the sediment is treated as an attenuating liquid and the roughened interface is modeled in terms of a roughness spectrum with a roughness parameter defined as the product kh where k is the acoustic wave number in water and h is the root-mean-square roughness height. Their model has shown that a significant amount of energy can be scattered into the sediment for $kh=0.66$ at a frequency of 20 kHz; this corresponds to $h=8.07$ mm. If this were scaled to a frequency of 100 kHz, one would need $h=1.61$ mm, which is consistent with using the 4.76-mm-diam beads used in our measurements. On the other hand, little transmission is seen with the smallest sized beads (2.38 mm diam) in our tank. This is to be contrasted with the enhanced penetration observed by Boyle and Chotiros[2] and Chotiros[4] in another tank experiment with a sand bottom. Since their sand surface was reported to be scraped flat, the small beads in our tank were expected to be a reasonable model of the residual interface roughness in their tank experiment. Therefore, the low signal level observed in our tank suggests a different mechanism may be responsible for the penetration reported in Refs. 2 and 4.

Although past results are not in complete accord with the present measurements, the larger bead sizes did produce considerable transmission into the glycerin layer and, when scaled to the frequencies used, correspond to reasonable dimensions for bottom roughness features observed in field experiments. Therefore, it is likely that at least some of the transmission observed in field experiments is caused by interface roughness. However, it is also notable that the transmission enhancements shown in Figs. 6–8 did not follow the expected trend as a function of frequency for a bead arrangement spaced to optimize Bragg scattering near 100 kHz. Thus the 100-kHz plots did not necessarily show the greatest transmission when the grazing angle was near 18°. The reason for this will be made clearer as the Bragg predictions are investigated.

C. Assessment of the Bragg scattering predictions

In the following, the predictions of enhanced shallow-grazing-angle transmission based on Bragg theory are compared with the tank and simulated data. First, the frequency, f_d, associated with the prediction of the maximum transmitted energy is tested by spectral analysis. Following this, the predicted propagation direction of the constant amplitude fronts, γ, and the apparent sound speed, c_{ca}, are tested by beamforming techniques.

1. Frequency

The peak of the shallow-grazing-angle transmission spectrum associated with Bragg scattering predictions is tested by processing tank signals recorded by the hydrophone located 11.7 cm below the interface. This hydrophone is chosen for the spectral analysis because it is located in the same vertical plane as the main response axis of the source; thus, issues arising from possible off-axis pulse distortions are avoided. As done for the waterfall plots, only scattering data for beads arranged in a square array are used. For this data, the source beam was set at an 18° grazing angle, which would optimize Bragg scattering at $f_d=102$ kHz for the given bead spacing above the hydrophone. Also, the 6.35-mm bead deployment was used to optimize the scattering strength. A spectral response amplitude, $R(f)$, defined as

$$R(f)=\left|\frac{g_{\mathrm{gly}}(f)}{g_{\mathrm{oil}}(f)}\right| \tag{6}$$

forms the basis of the analysis. Here $g_{\mathrm{gly}}(f)$ is the Fourier transform of signals recorded by the hydrophone in the glycerin and $g_{\mathrm{oil}}(f)$ is the Fourier transform of a reference signal recorded in the vegetable oil at a distance of 1 m from the source. Tank signals resulting from transmitting 0.1-ms pulses centered at 100, 115, 130, 145, and 200 kHz through the array are analyzed by performing fast Fourier transforms over a time window from 0.83 ms to 1.6 ms. This window is dominated by scattering from the beads and excludes any contribution due to the refracted wave associated with the source's side lobe. The transformed signals are overlapped in the frequency domain to span a broad frequency range.

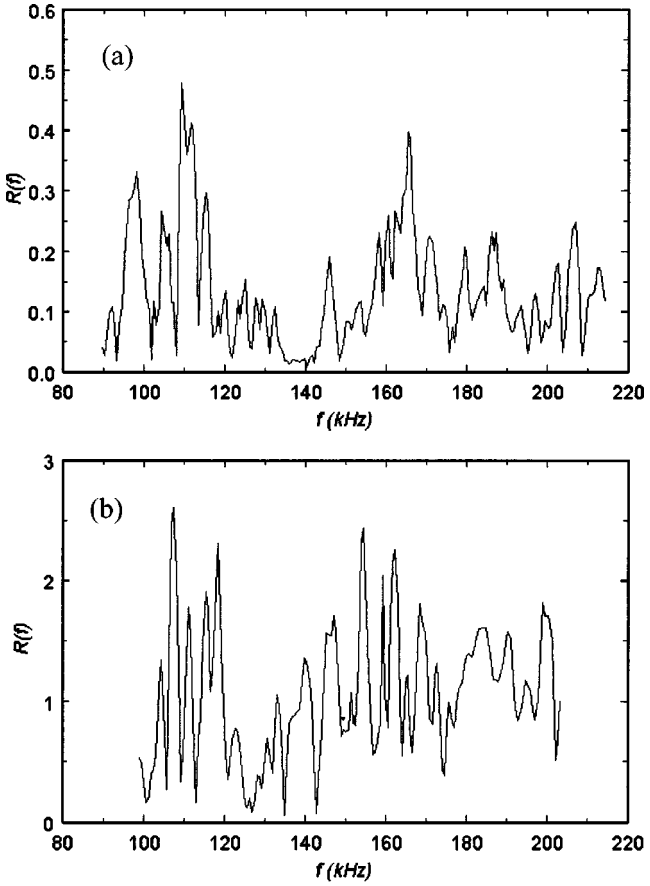


FIG. 9. Spectral amplitude of (a) tank data and (b) simulated tank data.

As an aid for interpretation of the results, simulated tank data are generated and processed in a similar fashion for comparison. To obtain the simulated data, the model described in Sec. III was configured to mimic the required experimental geometry, source characteristics, and fluid properties as specified in Table I. In addition, an angular scattering pattern was assigned to the beads in the model by amplitude weighting the signals scattered by the beads over the scattering angle ξ specified in Eq. (2). This is necessary because the scattering pattern of the 6.35-mm beads is not very omnidirectional at the frequencies considered. Although the actual scattering pattern associated with these beads is not known, it is conjectured that the pattern associated with the free-field condition (in vegetable oil) will be a reasonable approximation and this is used. However, because the free-field scattering pattern of polystyrene spheres changes with frequency, the transmission of pulses centered at closer frequency intervals than measured in the tank had to be simulated in order to obtain good overlap of the transformed signals across a broad frequency range. Thus to create the modeled spectral response, transmission was calculated for 0.1-ms pulses centered at 2.5-kHz intervals between 100 and 202.5 kHz.

Two spectral amplitude plots, obtained from (a) measured tank data and (b) simulated tank data, are given in Fig. 9. Note that the difference in overall amplitude is not significant here because the model used merely assumes the beads are perfect point scatterers. Aside from modulating their scattering pattern, no attempt is made to account for their intrinsic scattering strength since we are primarily interested in the relative distribution of energy from the array as a function of parameters such as frequency and receiver coordinates. Other differences in detail may be attributed to several possibilities; e.g., the bead array could not be deployed as perfectly square as in the model, the model does not account for refraction of the bead scattering pattern at the oil/glycerin interface, and there may be effects due to dispersion and multiple scattering (between the beads and between the beads and the oil/glycerin interface) which the model ignores.

In any case, Fig. 9 indicates that the model predictions are in qualitative agreement with the measured results. Both plots show a highly variable spectrum with no dominant frequency band. In particular, neither plot shows a definite maximum in amplitude at 102 kHz as expected based on Bragg scattering arguments. The reason for this is likely due to near-field effects. In the near field of an illuminated patch of roughness, Eq. (3) allows for a richer spectrum because reception will include signals scattered from a range of β angles. The beads floating at the oil/glycerin interface act as the radiating source for the signals recorded by the hydrophone in the glycerin. For the hydrophone to be in the far field of this effective radiating source, it should be located a distance below the oil/glycerin interface that is greater than $A_{\text{beads}}/2\lambda_{\text{gly}}$. Here, A_{beads} is the area of the beaded interface illuminated by the incident acoustic beam, and λ_{gly} is the wavelength in glycerin. For the parameters used in the tank measurements, the far field approaches 10 m at a frequency of 100 kHz.

The effect of signal reception in the near field on the spectral amplitude is demonstrated in Fig. 10. The spectra given are generated from simulated data in which 0.1-ms plane-wave pulses are incident on the oil/glycerin interface at an 18° grazing angle. To span the band shown, incident pulses centered on 40, 50, 60, 70, 80, 90, 100, 115, 130, 145, 160, 180, and 200 kHz are used. A receiver is placed in the same vertical plane as the source and directly beneath an array of 15 by 15 scatterers. The scatterers are deployed on the interface in a regular lattice with a spacing of 1.5 cm. Their scattering pattern is assumed to be omnidirectional and the glycerin is assumed to be a nonattenuating fluid. Figures 10(a), (b), and (c) correspond to the receiver placed 11.7, 46.8, and 93.6 cm below the oil/glycerin interface, respectively. Note that at 100 kHz, the far field of the 15 by 15 array in the glycerin is calculated to be 1.33 m. The basic results may be summarized as follows. At the shallowest receiver depth, the spectral amplitude oscillates with frequency, and there is no clearly dominant maximum. For the 46.8-cm-deep receiver, a high amplitude region containing two peaks appears to be forming in the 85–120 kHz frequency band. However, at the deepest receiver, this high amplitude region is already well formed with a single peak at the predicted frequency of 102 kHz.

Figure 10 demonstrates the need to be in the far field of the illuminated roughness to observe the frequency maximum caused by Bragg scattering. In fact, further simulations show the spectrum in Fig. 10(c) is little changed even if the 225 scatterers in the ''roughness'' are randomly but uni-

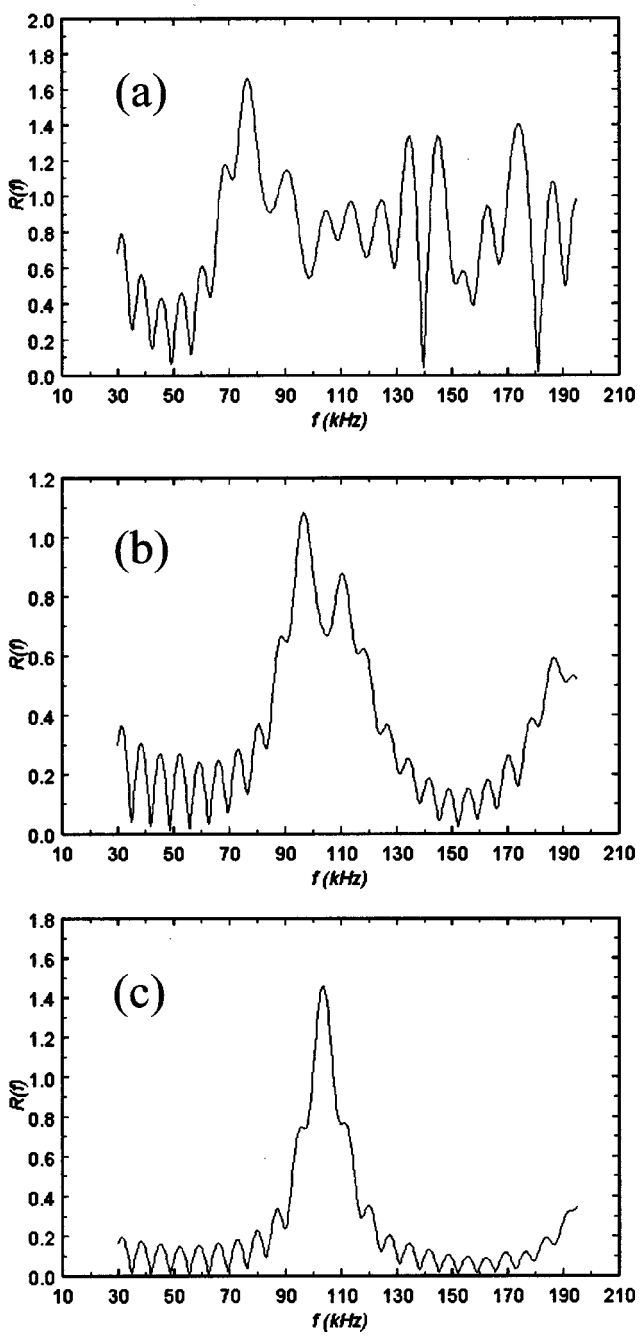


FIG. 10. Spectral amplitude of simulated data with receiver (a) 11.7 cm, (b) 46.8 cm, and (c) 93.6 cm below the fluid–fluid interface.

formly distributed in the same area as the regular lattice. However, our model indicates this observation may still be frustrated in realistic environments such as sandy bottoms because the sediment can attenuate the scattered acoustic field. Figure 11 presents another spectral amplitude plot obtained using simulated data. Transmission is modeled as for Fig. 10(c) except a glycerin attenuation level of $0.953\ \mathrm{dB}/\lambda_{\mathrm{gly}}$ (which is comparable to sandy ocean bottoms)[4,16,17] is used in the simulation. With attenuation included, incident pulses spaced at closer center-frequency intervals were needed to improve the overlap of signal spectral bands spanning the frequency range plotted. Thus tapered 0.1-ms pulses centered at 2.5-kHz intervals from 32.5 kHz to 205 kHz were used and the superposed spectral bands were smoothed into a

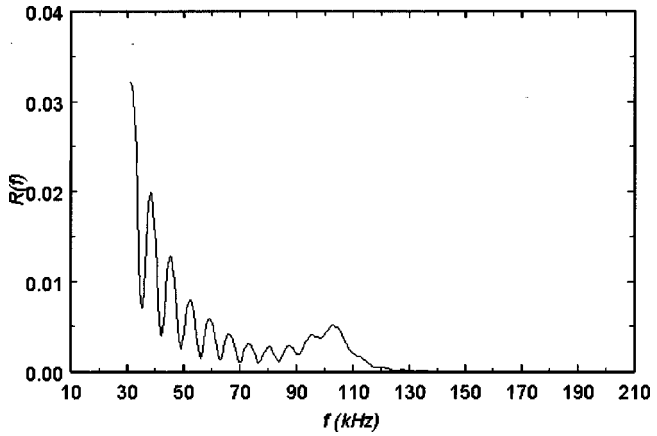


FIG. 11. Spectral amplitude of simulated data illustrating impact of attenuation.

single curve by using an averaging routine. The plot exhibits a general decay with frequency so that the highest displayed amplitude occurs at the low end of the plotted frequency range. Per predictions, a local maximum in amplitude appears at about 102 kHz, but the curve is skewed and not symmetric about this frequency. Furthermore, the level of this local maximum is almost 200 times lower than that shown in Fig. 10(c). If noise levels exceed the signal level at the predicted frequency, this local maximum can be obscured relative to less attenuated minor peaks at lower frequency.

2. Apparent sound speed and depression angle

The apparent sound speed, c_{ca}, and direction of constant amplitude wave fronts (i.e., depression angle), γ, scattered by the beads into the glycerin layer are obtained by processing the signals collected by the hydrophone array using techniques analogous to those reported by Chotiros and collaborators.[4,14] These techniques are similar to those used in time-delayed beamformers. A search through combinations of sound speed and depression angle is used to find a pair that results in high correlation between signals recorded at several hydrophones, indicating detection of a signal propagated through the receive array with the associated speed and angle. High correlation is displayed as a peak in a map spanning sound speed and depression angle space that is referred to as an ambiguity plot. In separate field experiments, Chotiros and collaborators used two different processing techniques to generate ambiguity plots from data collected with buried hydrophone arrays. In one experiment, they employed coherent processing methods.[14] In the other experiment,[4] these methods were not useful because uncertainties in the locations of the buried array elements were comparable to the acoustic wavelength for the range in frequencies covered. Therefore, in that case, ambiguity plots were generated based on incoherent (or intensity) processing methods. Our tank and simulated data will be processed and displayed using both these techniques. The processing steps employed to generate both coherent and incoherent ambiguity plots are described in the Appendix.

Figure 12 presents typical ambiguity plots obtained by coherently processing full waveform tank data taken from the set of waveforms acquired for 30 different bead configu-

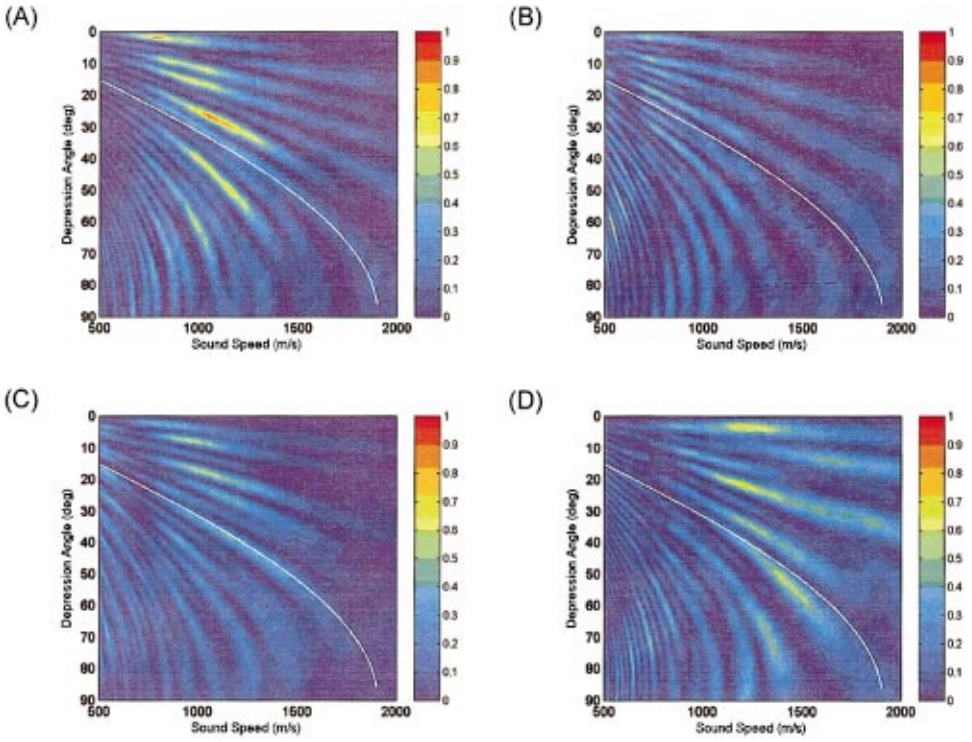


FIG. 12. Ambiguity plots generated by coherently processing the tank data for (a) first, (b) tenth, (c) twentieth, and (d) thirtieth configuration of the location of the beads.

rations. In each instance, the frequency is 100 kHz and the 6.35-mm-diam beads are deployed at the interface. Data were processed over a time interval that was clearly dominated by scattering from the beads and in which refracted signals and reverberation from tank hardware were low or excluded. Therefore, intense peaks in these plots are the result of signals scattered by the beads only. Ambiguity plots for the first, tenth, twentieth, and thirtieth bead configurations are shown. The first bead configuration corresponds to the beads arranged in a regular lattice and the maximum amplitude in all plots is scaled by the maximum exhibited for this configuration. Also shown in each plot is a curved white line that has been proposed as a bound for scattering effects. Chotiros *et al.*[14] claimed that, in observations of enhanced acoustic bottom penetration at grazing angles below critical, peaks caused by rough surface scattering will occur above this bound. The sound speed and depression angle coordinates $(c_{\mathrm{ca}}, \gamma)$ of this line are determined by the following two equations:

$$c_{\mathrm{ca}} = \frac{c_f c_h}{\sqrt{c_f^2 + c_h^2}} \tag{7}$$

and

$$\gamma = \tan^{-1}(c_h / c_f). \tag{8}$$

Here c_f is the fast compressional wave speed and c_h is the horizontal phase speed, which is treated as a variable. In subcritical penetration data acquired in a silty, sandy bottom and processed coherently, Chotiros *et al.* observed peaks in ambiguity plots that were located at speed and angle coordinates below this line. They attributed these peaks to a propagating Biot slow wave.

In each plot of Fig. 12, multiple peaks are seen and these peaks are located at different coordinates for the different plots. In some instances, these peaks appear above the white line; in others, they are below. Similar results are observed in ambiguity plots generated from data associated with the 26

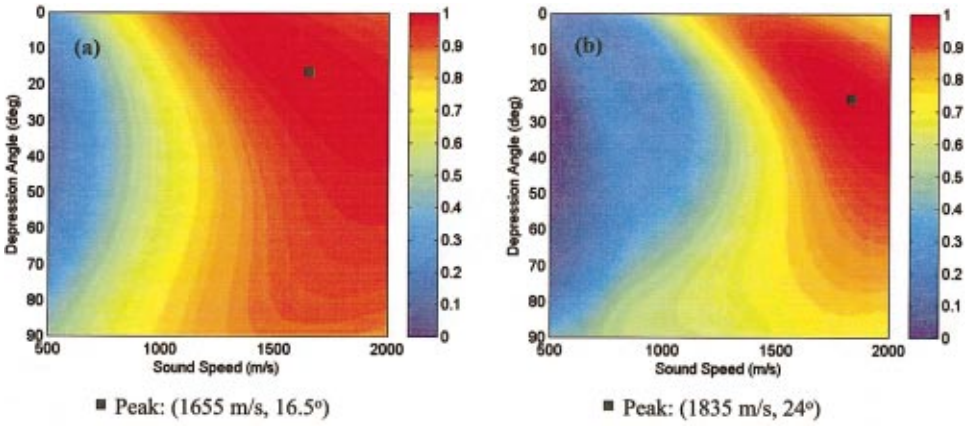


FIG. 13. Ambiguity plots corresponding to (a) tank and (b) simulated data generated by incoherent processing techniques. $f = 100$ kHz.

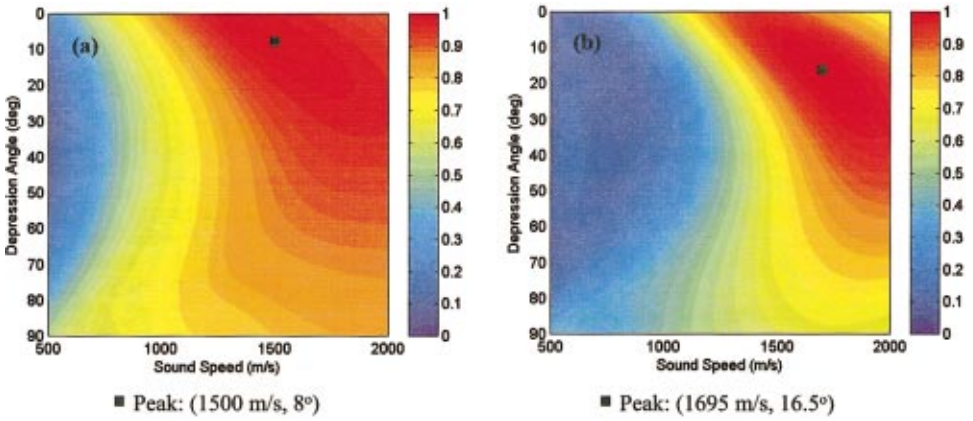


FIG. 14. Ambiguity plots corresponding to (a) tank and (b) simulated data generated by incoherent processing techniques. $f=200$ kHz.

remaining bead configurations. Clearly, the location of the beads on the interface impacts the position of the peaks appearing in the ambiguity plots. As a result, two important observations are made. First, peaks associated with roughness scattered signals can not necessarily be excluded from the region below the white line. To do so would require moving the receiver array far enough from the area of roughness being illuminated by the acoustic beam to limit the range of scattering angles β (see Fig. 5) accessible to the array. This is not likely to be the case when an acoustic beam is scanned over a buried receiver array at a shallow grazing angle as in our tank experiment or most field experiments. The acoustic beam intercepts an elongated area, allowing reception of scattered signals from over a wide range of angles. Therefore, Eqs. (7) and (8) cannot be used as suggested by Chotiros *et al.* to rigorously separate enhanced bottom transmission caused by roughness scattering at the surface from that caused by other possible transmission mechanisms. This point was also made by Thorsos *et al.*[8] based on simulations but Fig. 12 demonstrates the failure explicitly through experiment.

The second important observation made from Fig. 12 concerns the high variability in peak positions as the bead configuration is changed. This is also observed in ambiguity plots generated incoherently from the intensity envelopes of the full waveform tank data. The variability is a consequence of placing the hydrophone array in the near field of the illuminated rough area. Signals scattered from the beads interfere to produce signals traveling in directions that depend on the group configuration. In the near field, these signals can be received from many different directions and Eq. (4) will predict a different apparent direction for each. However, in regard to the average scattered intensity, it has been suggested[10,11] that the dominant values for c_{ca} and γ should still coincide with, or at least come close to, Bragg predictions as stated for a downward ($\beta=90°$) scattered field; i.e., $c_{ca}=1193$ m/s and $\gamma=39°$. This is tested in the following by averaging the intensity envelopes recorded by each hydrophone over an ensemble of bead configurations. These averaged intensity envelopes are then used to generate the ambiguity plots.

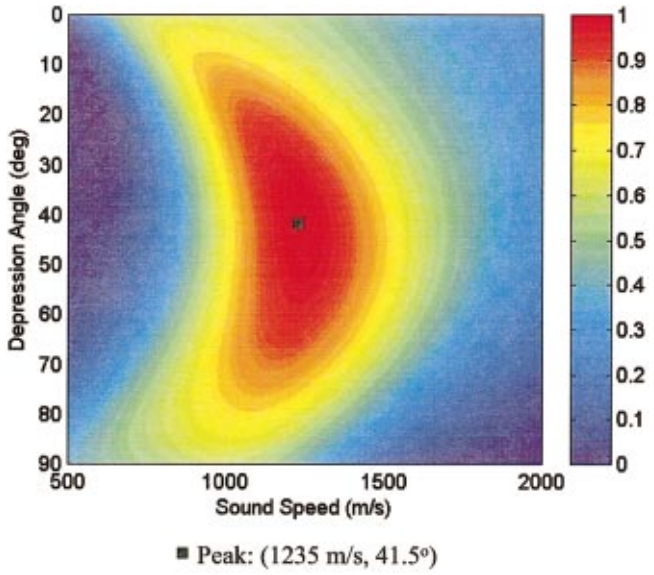


FIG. 15. Ambiguity plot generated using intensity envelopes obtained from simulated data in which the hydrophone array is located in the far field region of interface scatterers. $f=100$ kHz.

Figures 13 and 14 show ambiguity plots for squared signals (intensities) averaged over 96 bead configurations at 100 kHz and over 30 bead configurations at 200 kHz, respectively. Figures 13(a) and 14(a) refer to the tank data while Figs. 13(b) and 14(b) correspond to predictions of our model. Here, the difference in ensemble size arises because the 200-kHz plot of tank data was processed from the usual set of 30 waveforms acquired for 30 different bead configurations. However, the 100-kHz plot of tank data was processed from the extra intensity histories output from the Phillips PM 3394 oscilloscope after averaging 96 squared signals from the multiplier circuits. To create the simulated data, our model was set up to mimic the experiment, employing the same number of scatterers confined in the same size area of the interface and assuming uniform distribution of the beads so that their positions could be set by a random number generator. Intensities averaged over 100 and 30 configurations were used to obtain Figs. 13(b) and 14(b), respectively.

In each figure, a broad, high amplitude region appears between depression angles of about 0° and 45° and sound speeds of approximately 1200 and 2000 m/s. It is noted that, for each frequency, a significant portion of the high amplitude regions seen in the tank and simulated data overlap, indicating fair agreement between the two. Also, the sound speed associated with the peak is consistently less than 1905

TABLE II. Coordinates of source and buried hydrophones used in simulation of field experiment.

Coord. (m)	Source	Hydro.1	Hydro.2	Hydro.3	Hydro.4	Hydro.5	Hydro.6
X	−20.0	0.0	0.0	0.0	0.0	0.0	0.95
Y	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Z	4.5	−0.05	−0.18	−0.35	−0.57	−0.78	−0.42

m/s, which is the actual propagating speed of scattered signals in the glycerin. However, at each frequency, both the tank data and the simulated data yield significantly higher sound speeds and shallower depression angles than the 1193 m/s sound speed and 39° depression angle predicted by Bragg theory for reception of a predominantly downward (β=90°) scattered field. The simulated data are, comparatively, farther off in sound speed and closer in depression angle but the modest mismatch with the measured data probably arises because the model does not include the effects of refraction at the oil/glycerin interface on the bead scattering pattern. Including refraction effects would tend to shift the peak produced by simulated data toward shallower depression angles because scattering would be biased toward shallower angles. Insufficient ensemble averaging could also be a factor in the 200-kHz comparisons since simulations including larger ensembles do produce small shifts in the peak position. Nevertheless, in contrast to previous expectations, a significant discrepancy remains between the peak position in both measured and simulated ensemble-averaged ambiguity plots and the Bragg predictions, which appears to worsen with frequency. Additional simulations indicate that both the scattering pattern of the scatterers and their proximity to the hydrophone array affect the position of the peak in the ambiguity plot with the bead scattering pattern impacting this position the most. The significance of these two effects can be demonstrated with our model by relaxing both restrictions.

Figure 15 shows an ambiguity plot generated using intensity envelopes obtained from simulated data in which the hydrophone array is located in the far field region of an array of omnidirectional scatterers. In this case, the intensity envelopes are unaveraged; i.e., derived from a single scatterer configuration. A 100-kHz plane wave is directed at an 18° grazing angle. An array of 15 by 15 scatterers, separated by 1.5 cm, float at the oil/glycerin interface in a square lattice arrangement directly over a hydrophone array. (The scattering array starts at an x-coordinate of +3.75 cm from that of the interface hydrophone.) Relative receiver positions in the hydrophone array are the same as used in our tank measurement except that the hydrophone array is moved farther down in the glycerin layer by 81.9 cm. A peak in the ambiguity plot occurs at sound speed–depression angle coordinates of (1235 m/s, 41.5°), which compare favorably to that of 1193 m/s and 39° based on Bragg scattering predictions. It is noteworthy that no ensemble averaging was required to obtain this result and the choice of using a regular array or distributing the scatterers uniformly in the same area led to the same result in simulation. This analysis as well as the previous spectral amplitude comparisons demonstrate validity of the predictions based on Bragg theory for a hydrophone array placed in a nonattenuating medium and in the far field region of illuminated interface scatterers.

As a final note, we point out that, while medium attenuation may limit our ability to observe the frequency f_d predicted from Bragg arguments to dominate the spectrum scattered by roughness, it does not affect agreement with the predicted apparent sound speed and depression angle. When the glycerin layer is modeled as an attenuating liquid in the scattering configuration used to obtain Fig. 15, the resulting simulations yield ambiguity plots essentially identical to Fig. 15. There are two reasons for this. First, with the exception of a reduction in amplitude, the waveform recorded by a particular hydrophone for the attenuated case is similar to that of the nonattenuated case. Second, in the processing technique used to generate an ambiguity plot, the signals recorded by each hydrophone are normalized to offset signal loss for the deeper hydrophones (see the Appendix).

V. MODEL PREDICTIONS COMPARED TO FIELD EXPERIMENTS

In most field experiments, shallowly buried hydrophones are used to assess bottom transmission. Assuming scattering from a rough interface is the dominant mechanism for subcritical penetration, such shallow burial depths may result in near-field reception if the surface patch illuminated by the source is sufficiently large. By processing tank and simulated tank data in the previous section, near-field effects were shown to impact the validity of predictions based on Bragg scattering. As a result, the maximum spectral amplitude of the scattered signals can deviate from the predicted frequency and the peak in ambiguity plots generated by incoherent processing methods can be broadened and shifted. In this section, Eqs. (1) and (2) are used to demonstrate the extent to which the effects seen in our tank can appear in field experiments by simulating results of a subcritical transmission experiment reported by Chotiros.[4] Simulated data are first spectrally processed to assess the extent to which the buried hydrophone array is in the near field of the acoustically illuminated part of the interface and then processed to generate an ambiguity plot which is compared with those reported in the literature.

Our model is configured according to Chotiros' field experiment. A 20-kHz conical acoustic beam, with a 10° 3-dB-down beam width, is incident on a water-sediment interface at a 12.7° grazing angle. A six-element hydrophone array is buried in the sediment under the illuminated area. Table II lists the coordinates of the source and buried hydrophones.[8] In the simulation, the sediment is assumed to be an attenuating liquid with an attenuation of $0.865\,\mathrm{dB}/\lambda_{\mathrm{sed}}$, where λ_{sed} is the wavelength of the acoustic beam in the sediment. The water column and sediment have a sound speed of 1536 m/s and 1729 m/s, respectively. It is further assumed that dis-

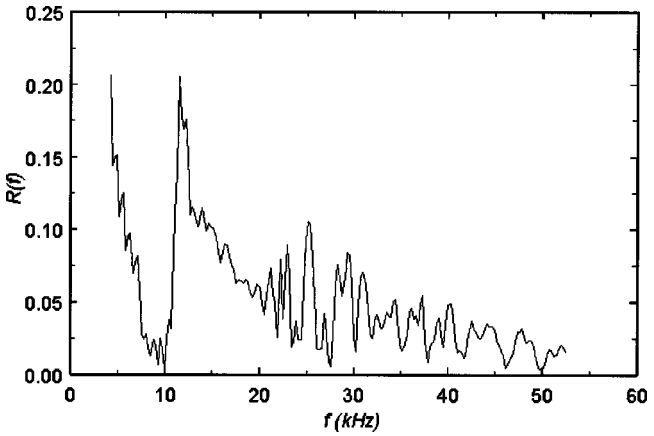


FIG. 16. Spectral response plot corresponding to simulation of the field experiment performed by Chotiros.

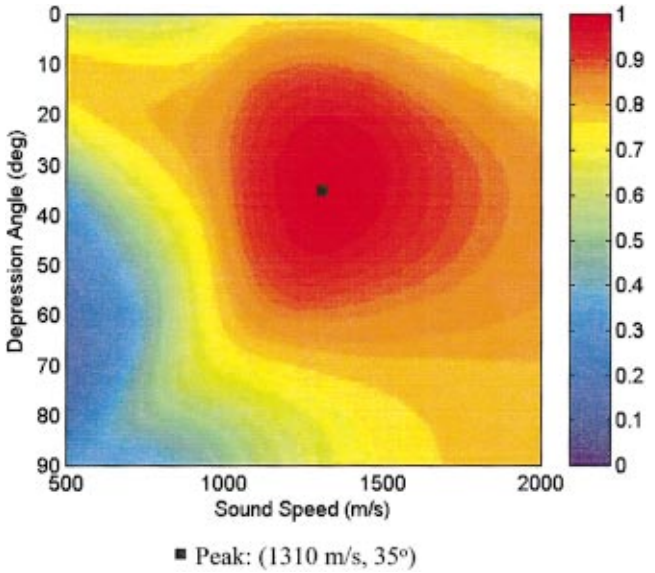


FIG. 17. Ambiguity plot generated using intensity envelopes obtained from simulating the field experiment performed by Chotiros. $f=20$ kHz.

crete, omnidirectionally scattering, identical, point scatterers are sufficient to represent the roughness that exists on the ocean bottom, using an average separation based on the predictions of Bragg theory to produce a downward scattered beam at 20 kHz. Thus a square array of 70 (x-direction) by 50 (y-direction) scatterers, separated by 7.87 cm, is initially placed on the fluid–fluid interface, centered above the hydrophone array. The size of the scatterer array corresponds to an area of 5.5 m by 3.9 m, which is comparable to the area on the interface illuminated by the incident beam (as defined by the 3-dB-down beam width). Inputting the appropriate parameters in Eqs. (4) and (5) (assuming $\beta=90°$), Bragg theory predicts an apparent sound speed of 1164 m/s and depression angle of 42° for the downward scattered field.

Figure 16 is a spectral response plot obtained from the simulated signals received by the hydrophone located 78 cm below the water-sediment interface. The frequency range shown is spanned by overlaying responses computed for incident pulses centered 5 kHz apart from 5 to 50 kHz. This plot does not show a maximum amplitude at the frequency predicted by Bragg theory. As observed in comparisons with the tank measurements, this implies that the buried hydrophone is in the near field of the scatterer array. Although wave attenuation mechanisms in the bottom were expected to reduce the effective area of the array audible to the hydrophone by reducing the signals scattered by the outer portion of the illuminated area, this reduction is insufficient to place the hydrophone in the far field. If the entire 5.5 m by 3.9 m scattering area produces signals accessible to the hydrophone, the $A_{\text{scatter}}/2\lambda_{\text{sed}}$ criterion requires it to be buried approximately 125 m below the water-sediment interface to be in the far field at 20 kHz. Additional simulations indicate that scattering arrays as small as 15 by 15 (corresponding to an area of about 1.2 m by 1.2 m) will still not exhibit a maximum at the frequency predicted according to Bragg theory. For the deepest hydrophone listed in Table II to be in the far field, scatterers on the interface should be illuminated by a beam projected on an area smaller than 35 cm by 35 cm. Therefore, even if interface scattering is the dominant mechanism for subcritical penetration in Chotiros' measurement, the results of our model suggest that the spectral content of the field scattered through the roughness will not be in accord with expectations based on heuristic Bragg scattering arguments; i.e., Eq. (3) with $\beta=90°$. Similar caveats can be made regarding the expected depression angle of the constant amplitude fronts, γ, and their apparent speed, c_{ca}.

Figure 17 is an ambiguity plot generated using intensity envelopes obtained from further simulations of the field experiment. The intensity envelopes are obtained by averaging squared signals from over 100 different scatterer configurations. These configurations retain the same number of scatterers in the regular array confined to an area of approximately 5.5 m by 3.9 m, but now they are uniformly distributed along the x- and y-directions with positions determined by a random number generator. The ambiguity plot exhibits a peak at a 35° apparent depression angle and 1310-m/s sound speed, which lies within a broad high amplitude region spanning 20° to 50° and 1100 to 1500 m/s. While close, this peak position is noticeably shifted from the 42° and 1164-m/s predictions based on Bragg scattering as well as the observed values of 40° and 1200 m/s previously reported.[4] However, the present results do compare quite well with the 34° and 1300-m/s values reported by Thorsos *et al.*[8] and Moe[10] who modeled Chotiros' experiment with a water/fluid-sediment interface characterized by small-scale Gaussian roughness and calculated the average transmitted intensity with perturbation theory.

Although limited, the agreement between ambiguity plots produced from field data and their model has been used by Thorsos *et al.* and Moe to argue in favor of roughness scattering as the mechanism responsible for observations of enhanced acoustic transmission at shallow grazing angles. Discrepancies were generally attributed to not being able to ensemble average over many roughness configurations in the field experiments. Our simulations bear out the consequences of not averaging. Simulations of Chotiros' field experiment with propagation through individual uniformly distributed scatterer configurations produced ambiguity plots displaying highly variable peak positions in sound speed–depression

angle space. Even when the scatterer distribution is not randomized, but arranged as a square lattice, the peak in the ambiguity plot varies quite a bit around the average position upon shifting the position of the hydrophone array by a fraction of a wavelength. As noted in the comparisons with tank data and in agreement with the simulated spectral response in Fig. 16, this variability is a result of signal reception in the near field of the illuminated roughness patch. Both the variability with scatterer configuration and the deviation from Bragg theory disappear in simulations of the far field reception. Therefore, the agreement between sound speed–depression angle values obtained from available field data and those predicted by Bragg theory is likely to be accidental if roughness scattering is the primary mechanism for shallow-angle transmission. This could explain the difficulties sometimes experienced in obtaining consistent results from bottom transmission measurements in the field.[18]

However, we also note it is not clear that the deviation of the peak in the ensemble-averaged ambiguity plot from the position predicted by Bragg theory will remain generally small in reality. Our experience with tank data suggests that, if the scattering by the roughness features is not omnidirectional, an ambiguity plot based on an ensemble-average of recorded field data can exhibit significant deviation in peak position from the Bragg prediction for downward scattered field components. Furthermore, at high enough frequencies, this can cause the peak to shift as a function of frequency. Therefore, one cannot conclude that agreement with Fig. 17 or with the plots of Thorsos *et al.* and Moe would result upon ensemble averaging Chotiros' field data. Models assuming small-scale roughness such as those used by Thorsos *et al.* and Moe may not be sufficient to achieve agreement. In this regard, we merely suggest that the roughness level corresponding to using the 6.35-mm-diam beads in our tank experiments at 100 kHz is reasonable when scaled to retain the same roughness size to wavelength ratio at 20 kHz.

VI. SUMMARY

A set of tank experiments was performed to test Bragg scattering arguments for enhanced transmission into a bottom layer at subcritical grazing angles. In these experiments, two immiscible fluids, vegetable oil floating on glycerin, formed an interface where polystyrene beads of three sizes (6.35, 4.76, and 2.38 mm diam) were floated to simulate roughness. An array of hydrophones placed in the bottom layer (glycerin) was used to measure the acoustic levels transmitted across the interface. This array was also employed as a beamformer to determine the apparent depression angle and sound speed of the scattered signals. Data were acquired in the frequency range of 100–200 kHz both with and without beads deployed and for various configurations in which the locations of the beads floating on the interface were varied. In addition, a simulation of the tank measurements was performed to provide a qualitative but realistic representation of data, accounting for most tank-related artifacts including those due to the dimensions of the roughened area, the proximity of the source, the incident pulse shape, reverberation from tank walls, directional scattering by the beads, and the location of the receive hydrophones.

Results of the measurements demonstrated that scattering by interface roughness is a viable mechanism for enhanced penetration at subcritical grazing angles. In general, signal levels scattered by the beads into the glycerin layer increased with increasing bead diameter. The largest beads exhibited signal levels exceeding the background noise floor of the no-bead case by almost 40 dB, whereas the signal levels scattered by the smallest beads were close to the background noise floor. Furthermore, the tank and simulated data were spectrally processed to test the frequency maximum f_d predicted in the spectrum of the scattered field. When the beads were modeled with a realistic scattering pattern, the spectrum obtained from simulation was in good qualitative agreement with that obtained from measurements. However, neither the spectrally processed tank nor simulated data exhibit a dominant maximum at f_d. With additional simulations it was determined that simple Bragg frequency predictions break down for reception of the scattered field in the near field of the acoustically illuminated area. Recovery of the Bragg prediction was verified by simulating reception in the far field.

The tank and simulated data were also beamformed to test Bragg predictions for the apparent sound speed c_{ca} and the depression angle γ of the constant amplitude fronts. When using single-ping, full-waveform signals acquired in the tank, the beamformed results exhibited c_{ca} and γ combinations that were not well localized and were dependent on the bead positions on the oil/glycerin interface. As with the frequency comparisons, this was a consequence of placing the hydrophones in the near field of the bead array. Nevertheless, previous work[10,11] suggested that the apparent sound speed and depression angle of the average field propagation should come close to Bragg predictions for a downward scattered field. To characterize the average field propagation, the intensity envelope of the scattered signals received at each hydrophone was ensemble averaged over various bead configurations. The resulting beamformed intensity-enveloped tank data showed that roughness scattering transmits a field that, on average, still propagates at a significantly shallower γ and higher c_{ca} than predicted from Bragg arguments. A simulation of the tank data, incorporating a scattering pattern to the beads, yielded qualitative agreement with the measured results. Additional simulations were then used to show that these discrepancies with heuristic Bragg predictions can result when the roughness scatters a nonomnidirectional field and when reception is in the near field of the illuminated area of roughness.

As mentioned above, enhanced transmission levels were easily observed in our tank with the two larger diameter beads, but little transmission was observed with the smallest beads. The ineffectiveness of the roughness level represented by the small beads is in contrast to the transmission reported by Chotiros and collaborator[2,4] in their tank experiment. But, despite this discrepancy, it is likely that at least some of the transmission observed in field experiments is caused by interface roughness. When our simulation was configured to

examine transmission effects in Chotiros' field experiment, the results suggested roughness scattering to be a feasible mechanism for explaining the enhanced bottom transmission observed at shallow grazing angles, but care needs to be used in interpreting the results. In this case, our simulations indicate that agreement with transmission properties predicted on the basis of heuristic Bragg arguments for downward scattered field components are not guaranteed. Unaveraged, near-field reception of the transmitted field can result in deviations from expected values of f_d, γ, and c_{ca}. Furthermore, ensemble-averaged predictions of the near-field reception can also be flawed if the roughness features are large enough to scatter in a nonomnidirectional pattern.

ACKNOWLEDGMENTS

The authors gratefully acknowledge support from several Office of Naval Research programs: the Ocean Engineering and Marine Systems Program, the Ocean Acoustics Program, and the In-house Laboratory Independent Research Program administered through the Coastal Systems Station. The authors also wish to thank Dr. D. G. Todoroff who helped to formulate the concept of this experiment, R. Ball and E. Kloess for their support in the experimental portion of this effort, and Dr. J. Stroud for reviewing this manuscript.

APPENDIX: PROCESSING METHODS USED TO GENERATE AMBIGUITY PLOTS

Chotiros and collaborators employed both incoherent[4] and coherent[14] processing techniques to generate ambiguity plots. In both methods, one hydrophone in the array is designated as a reference; and the signals from each of the nonreference hydrophones are processed and normalized to offset for signal loss in the deeper hydrophones. The reference hydrophone provides a reference time, t_{ref}, which corresponds to the time when the signal passes this hydrophone. The reference hydrophone is also used to provide a spatial reference, $\mathbf{r}_{\mathrm{ref}}$, with respect to the other hydrophones in the array. The arrival time of an acoustic signal, traveling at a sound speed c in the direction given by the unit vector $\mathbf{n}$, and recorded by the kth hydrophone in the array that is positioned at $\mathbf{r}_k$, is given by

$$t_k = t_{\mathrm{ref}} + \mathbf{n}\cdot(\mathbf{r}_k - \mathbf{r}_{\mathrm{ref}})/c. \tag{A1}$$

In the processing it is noted that the vectors in Eq. (A1) are two-dimensional projections onto the vertical plane containing the source beam axis because the wave front illuminating the roughness above the hydrophone array is assumed to be planar. This simplification was tested using simulations of data collected on and off the plane and found to be reasonable.

For a particular sound speed and depression angle pair, the processed signals obtained by each nonreference hydrophone are time-shifted with respect to the reference time with the aid of Eq. (A1), and the resultant time-shifted signals are summed. This procedure is repeated until the entire sound speed and depression angle space is covered. For the entire set of sound speed and depression angle coordinates, the amplitude associated with a particular sound speed–depression angle pair is assigned a color with red corresponding to the highest amplitude level and blue representing the lowest level.

In this work, no significant advantage was gained by using either the summed signals or the decibel equivalent of the summed signals to isolate the sound speed–depression angle peaks in ambiguity plots. As a result, we chose to generate all ambiguity plots displayed in this paper using the summed signals and not their decibel equivalent.

For all examples shown in this paper, the hydrophone located in the bottom fluid layer that is closest to the roughened fluid–fluid interface is used as the reference hydrophone. In our tank experiment, simulations of our tank experiment, and simulations showing far field effects, the reference hydrophone is positioned in the glycerin 4.0 cm (tank and simulated tank data) and 85.9 cm (far field effects) below the oil/glycerin interface. For the frequency range used (100–200 kHz), these distances are below the evanescent zone; and the only signals of appreciable amplitude that each hydrophone in the array records are those which are scattered into the glycerin by the beads. This insured that the ambiguity plot illustrates the apparent sound speed and depression angle due solely to the energy scattered into the glycerin by the beads. In our simulations of Chotiros' field experiment,[4] the hydrophone located 5 cm below the roughened water-sediment interface is chosen as the reference hydrophone to mimic the reported work even though this is in the evanescent zone at 20 kHz and evanescent waves are not accounted for in our simulation.

For the ambiguity plots generated by incoherent processing methods, the signals are processed as follows. First the intensity envelope of the signals recorded by each nonreference hydrophone is cross-correlated with the intensity envelope obtained by the reference hydrophone. Next, these cross-correlated signals are enveloped using a 25-point box car filter run in both directions on the signal trace and then squared. This procedure is somewhat similar to that used by Chotiros who processed his data with a spike function to localize the signals at a higher resolution than at the 10-kHz sample rate at which his data were recorded.[4] Our tank and simulated data were recorded at a 5-MHz sample rate which provided the resolution required by correlation techniques to adequately localize the signals obtained with our hydrophone array.

For ambiguity plots generated by coherent processing methods, the recorded signals are processed similar to that of Chotiros *et al.*[14] The waveforms collected by each of the nonreference hydrophones are Hilbert transformed to produce real and imaginary time domain signals. For a particular sound speed and depression angle pair, these complex signals are time-shifted as described above and then summed (real with real and imaginary with imaginary). The square of the absolute value of the summed complex signals is the amplitude observed in these types of ambiguity plots.

[1] N. P. Chotiros, ''High frequency acoustic bottom penetration: Theory and experiment,'' in *Proceedings of Oceans '89* (IEEE Publications No. 89CH2780-5) (IEEE, New York, 1987), Vol. 3.
[2] F. A. Boyle and N. P. Chotiros, ''Experimental detection of a slow acous-

tic wave in sediment at shallow grazing angles,'' J. Acoust. Soc. Am. **91**, 2615–2619 (1992).
[3] J. L. Lopes, ''Observation of anomalous acoustic penetration into sediment at shallow grazing angles,'' J. Acoust. Soc. Am. **99**, 2473–2474(A) (1996).
[4] N. P. Chotiros, ''Biot model of sound propagation in water-saturated sand,'' J. Acoust. Soc. Am. **97**, 199–214 (1995).
[5] H. J. Simpson and B. H. Houston, ''A synthetic array measurement of acoustical waves propagating into a water-saturated sandy bottom for a smoothed and a roughened interface,'' J. Acoust. Soc. Am. **107**, 2329–2336 (2000).
[6] J. L. Lopes, I. C. Paustian, R. Marciniak, and D. L. Folds, ''Dual-frequency acoustic lens sonar system development for the detection of both buried objects and objects proud of the bottom,'' Proceedings of SPIE AeroSense99, Orlando, FL, April 1999.
[7] A. Maguer, W. L. J. Fox, H. Schmidt, E. Pouliquen, and E. Bovio, ''Mechanisms for subcritical penetration into a sandy bottom; Experimental and modeling results,'' J. Acoust. Soc. Am. **107**, 1215–1225 (2000).
[8] E. I. Thorsos, D. R. Jackson, and K. L. Williams, ''Modeling of subcritical penetration into sediments due to interface roughness,'' J. Acoust. Soc. Am. **107**, 263–277 (2000).
[9] E. I. Thorsos, D. R. Jackson, J. E. Moe, and K. W. Williams, ''Modeling of subcritical penetration into sediments due to interface roughness,'' *SACLANTCEN Conference Proceedings Series CP-45: High Frequency Acoustics in Shallow Water*, Lerici, Italy, 30 June–4 July 1997.
[10] J. E. Moe, ''Near and far field acoustic scattering through and from two dimensional fluid–fluid rough interfaces,'' Applied Physics Laboratory, Seattle, WA, Tech. Rep. APL-UW TR 9606, October 1996.
[11] J. E. Moe and D. R. Jackson, ''Near-field scattering through and from a two-dimensional fluid–fluid rough interface,'' J. Acoust. Soc. Am. **103**, 275–287 (1998).
[12] D. Tang, ''Small scale volumetric inhomogeneities of shallow water sediments: Measurements and discussion,'' *SACLANTCEN Conference Proceedings Series CP-45: High Frequency Acoustics in Shallow Water*, Lerici, Italy, 30 June–4 July 1997.
[13] D. Tang, ''Geoacoustic parameters for high-frequency bottom scattering,'' J. Acoust. Soc. Am. **104**, 1813(A) (1998).
[14] N. P. Chotiros, A. M. Mautner, A. Lovik, A. Kristensen, and O. Bergem, ''Acoustic penetration of a silty sand sediment in the 1–10 kHz band,'' IEEE J. Ocean Eng. **22**, 604–615 (1997).
[15] See, for example, L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York, 1982), Sec. 8.8.
[16] E. L. Hamilton, ''Geoacoustic modeling of the sea floor,'' J. Acoust. Soc. Am. **68**, 1313–1340 (1980).
[17] R. Lim, J. L. Lopes, R. H. Hackman, and D. G. Todoroff, ''Scattering by objects buried in underwater sediments: Theory and experiment,'' J. Acoust. Soc. Am. **93**, 1762–1783 (1993).
[18] N. P. Chotiros, ''Analysis of Sea Test Data I,'' Technical Report No. ARL-TR-95-11, Applied Research Laboratory, University of Texas at Austin, April 1995.

Modeling of bottom backscattering from three-dimensional volume inhomogeneities and comparisons with experimental data

Dan Li and George V. Frisk
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Dajun Tang
Applied Physics Laboratory, University of Washington, Seattle, Washington 98105

(Received 8 December 1998; revised 2 November 2000; accepted 15 December 2000)

In this paper, backscattering from 3D volume inhomogeneities in the seabed is modeled and the results compared with experimental data at 250–650 Hz. The experiment was part of the Acoustic Reverberation Special Research Program (ARSRP) and the data were obtained in a sediment pond on the western flank of the Mid-Atlantic Ridge. A volume scattering model based on first-order perturbation theory is developed incorporating contributions from both sound speed and density fluctuations. With the propagators, i.e., the Green's functions, handled accurately through numerical wave number integration and random fluctuations generated effectively by a new scheme modified from the spectral method, the model is capable of simulating monostatic, backscattered fields in the frequency domain as well as in the time domain owing to 3D volumetric sediment inhomogeneities. The model compares favorably and consistently with the ARSRP backscattering data over the entire frequency band, with the fluctuations of sound speed and density in two irregular sediment layers, identified from the data analysis, described by a power-law type of power spectrum. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1348300]

PACS numbers: 43.30.Gv, 43.30.Ft, 43.30.Hw [DLB]

I. INTRODUCTION

A considerable amount of effort has been expended on the study of acoustic scattering from the ocean bottom, which is critical to various sonar operations. Rough seafloor surfaces, exposed rocks on the seafloor, and large topographical features such as sea mountains and ridges have historically been considered the primary contributors to bottom scattering. In recent years, however, increasing evidence has shown that inhomogeneities within sediments can be a dominant factor in total bottom scattering, especially when the seafloor is relatively flat and/or the bottom attenuation is small. In his early work on the measurement of the bottom backscattering strength, Merklinger[1] speculated that subbottom inhomogeneities may be the primary contributors to the backscattered field. At high frequencies, Jackson *et al.*[2–4] and Lyons *et al.*[5] successfully developed bottom volume scattering models and compared the models to field data. Tang *et al.*'s[6] analysis of high-frequency scattering data showed that the gas voids in the sediment are probable significant scatterers.

As part of the Office of Naval Research Bottom/Subbottom Acoustic Reverberation Special Research Program (ARSRP), a backscattering experiment was conducted over a sediment pond on the western flank of the Mid-Atlantic Ridge in July 1993. Analysis of experimental data revealed that sediment inhomogeneities within two irregular layers beneath the water/sediment interface were the major contributors to backscattering at oblique angles.[7,8]

In order to achieve a good understanding of the scattering mechanism, the objective of this work is to develop a backscattering model that can incorporate layers of 3D bottom volumetric inhomogeneities in sediments and to compare the results with the ARSRP backscattering data. In the model, the incident field will be obtained by a numerical wave number integration technique using OASES.[9] OASES is a general purpose computer code for modeling seismo-acoustic propagation in horizontally stratified waveguide. The importance of having a good propagation model such as OASES has been discussed in Refs. 10, 11. This approach enables the model to deal with volume scattering by scatterers embedded in a stratified background medium. A first-order perturbation method[4,12–14] is applied to determine the scattered field, where sound speed and density fluctuations in the fluid sediment are considered to be the causes of scattering. Rough surface scattering is not included here since field survey reveals that the seafloor roughness is small compared to the wavelength concerned. The ARSRP data analysis also demonstrates the insignificance of the water/sediment interface scattering in the experimental area.[7,8] While the necessity to include poro-elastic effects in a sediment scattering model remains an open question,[4] a fluid model is assumed appropriate for the soft sediments and low frequencies under consideration. The volume inhomogeneities, i.e., the sound speed and density variations, are assumed to be statistically homogeneous and can be generated by a spectral method.[15] The random field is assumed to be correlated in all three dimensions and defined by a power spectrum. The computation to generate the above random field are greatly reduced after taking advantage of the fact that the source and all the hydrophones on the receiving array are omnidirectional and are co-located along a vertical line. The model is then used to interpret the backscattering phenomena that have been ob-

served in the ARSRP sediment scattering experiment.

This paper is structured as follows: In Sec. II, we present the volume scattering model based on a first-order perturbation theory. The description of the random sound speed and density fluctuations, as well as the scheme to generate 3D azimuthally-summed random fields, are introduced. Section III is devoted to the comparisons between model predictions and experimental data.

II. FORMULATIONS FOR MODELING BOTTOM VOLUME BACKSCATTERING

In this section we derive the formula for modeling volume scattering due to sound speed and density fluctuations. It is similar to the approaches by Morse and Ingard,[12] Chernov,[13] Hines,[14] and Jackson *et al.*[4] Also included are the correlation function to describe random sound speed and density fluctuations, and the description of a scheme to generate 3D random fields using the spectral method.

A. Volume scattering theory

The wave equation in an inhomogeneous medium[16] can be written as

$$\rho(\mathbf{r})\nabla\cdot\left(\frac{1}{\rho(\mathbf{r})}\nabla p(\mathbf{r},t)\right)-\frac{1}{c^2(\mathbf{r})}\frac{\partial^2 p(\mathbf{r},t)}{\partial t^2}=f(\mathbf{r},t), \quad (1)$$

where $p(\mathbf{r},t)$ represents the pressure, $\rho(\mathbf{r})$ represents the density, $c(\mathbf{r})$ represents the sound speed, and $f(\mathbf{r},t)$ represents the source term as a function of space $\mathbf{r}$ and time t. Assuming a point source and harmonic time dependence $\exp(i\omega t)$, where ω is the radial frequency, the pressure field obeys the inhomogeneous Helmholtz equation:[13]

$$[\nabla^2+k^2(\mathbf{r})]P(\mathbf{r})-\frac{\nabla\rho(\mathbf{r})}{\rho(\mathbf{r})}\cdot\nabla P(\mathbf{r})=S_\omega\delta(\mathbf{r}-\mathbf{r}_s), \quad (2)$$

where $k(\mathbf{r})$ is the medium wave number

$$k(\mathbf{r})=\frac{\omega}{c(\mathbf{r})}, \quad (3)$$

and S_ω is the source strength, while

$$f(\mathbf{r},t)=F(\mathbf{r},\omega)e^{i\omega t}, \quad p(\mathbf{r},t)=P(\mathbf{r},\omega)e^{i\omega t}.$$

We assume that the background density is constant and the background sound speed varies only with depth z. Assume small variations of both density and sound speed, we write

$$\rho(\mathbf{r})=\rho_0+\delta\rho\mathbf{r}), \quad c(\mathbf{r})=c_0(z)+\delta c(\mathbf{r}), \quad (4)$$

where $\delta\rho(\mathbf{r})\ll\rho_0$ and $\delta c(\mathbf{r})\ll c_0(z)$, and they both have zero means.

If we expand the Helmholtz equation against the small quantities δc and $\delta\rho$ up to first-order terms. Equation (2) becomes

$$\left[\nabla^2+k_0^2(z)-\frac{2\,\delta c(\mathbf{r})k_0^2(z)}{c_0(z)}\right]P(\mathbf{r})-\frac{\nabla(\,\delta\rho(\mathbf{r}))}{\rho_0}\cdot\nabla P(\mathbf{r})=S_\omega\delta(\mathbf{r}-\mathbf{r}_s). \quad (5)$$

We also express the pressure field $P(\mathbf{r})$ as the summation of the zeroth-order field $P_0(\mathbf{r})$ and the first-order scattered field $P_s(\mathbf{r})$:

$$P(\mathbf{r})=P_0(\mathbf{r})+P_s(\mathbf{r}), \quad (6)$$

where the zeroth-order field $P_0(r)$ satisfies

$$[\nabla^2+k_0^2(z)]P_0(\mathbf{r})=S_\omega\delta(\mathbf{r}-\mathbf{r}_s). \quad (7)$$

Then the scattered field satisfies

$$[\nabla^2+k_0^2(z)]P_s(\mathbf{r})=\frac{2\,\delta c(\mathbf{r})k_0^2(z)}{c_0(z)}P_0(\mathbf{r})+\frac{\nabla(\,\delta\rho(\mathbf{r}))}{\rho_0}\cdot\nabla P_0(\mathbf{r}). \quad (8)$$

Similar to the approaches by Hines[14] and Jackson *et al.*,[4] δc and $\delta\rho$ are assumed correlated. If we use the notation

$$\epsilon(\mathbf{r})\equiv\frac{\delta c}{c_0}, \quad (9)$$

then we assume

$$\frac{\delta\rho}{\rho_0}=\beta\epsilon(\mathbf{r}), \quad (10)$$

where β is the ratio between the density and sound speed variations. Then the scattered field equation becomes

$$[\nabla^2+k_0^2(z)]P_s(\mathbf{r})=2k_0^2(z)\epsilon(\mathbf{r})P_0(\mathbf{r})+\beta\nabla\epsilon(\mathbf{r})\cdot\nabla P_0(\mathbf{r}). \quad (11)$$

Applying Green's Theorem to the equation above, we have

$$P_s(\mathbf{R_r})=\int_v[2k_0^2(z')\epsilon(\mathbf{r}')P_0(\mathbf{r}')+\beta\nabla\epsilon(\mathbf{r}')\cdot\nabla P_0(\mathbf{r}')]G(\mathbf{R_r},\mathbf{r}')d\mathbf{r}', \quad (12)$$

where $\mathbf{R_r}$ stands for the receiver position and $\mathbf{r}'$ stands for the scatterer position and the corresponding Green's function can be obtained from

$$[\nabla^2+k_0^2(\mathbf{r})]G(\mathbf{r},\mathbf{r_s})=\delta(\mathbf{r}-\mathbf{r}_s), \quad (13)$$

and the boundary conditions, e.g., continuity of normal particle velocity and pressure across each layer interface for a fluid medium. After integration by parts, the scattered field can be further expressed as

$$P_s(\mathbf{R_r})=\int_v\epsilon(\mathbf{r}')[(2+\beta)k_0^2(z')P_0(\mathbf{r}')G(\mathbf{R_r},\mathbf{r}')-\beta(\nabla P_0(\mathbf{r}')\cdot\nabla G(\mathbf{R_r},\mathbf{r}'))]d\mathbf{r}'. \quad (14)$$

The scattered intensity I_s is

$$I_s = \langle |P_s(\mathbf{R_r})|^2 \rangle$$
$$= \int_{v_1}\int_{v_2} \langle \epsilon(\mathbf{r_1})\epsilon^*(\mathbf{r_2})\rangle [(2+\beta)k_0^2(z_1)P_0(\mathbf{r_1})G(\mathbf{R_r},\mathbf{r_1}) - \beta(\nabla P_0(\mathbf{r_1})\cdot\nabla G(\mathbf{R_r},\mathbf{r_1}))] \times [(2+\beta)k_0^2(z_2)P_0^*(\mathbf{r_2})G^*(\mathbf{R_r},\mathbf{r_2}) - \beta(\nabla P_0^*(\mathbf{r_2}) \cdot \nabla G^*(\mathbf{R_r},\mathbf{r_2}))] d\mathbf{r_1} d\mathbf{r_2}. \quad (15)$$

Here (*) stands for the complex conjugate operation.

In the above equation, $\langle \epsilon(\mathbf{r_1})\epsilon^*(\mathbf{r_2})\rangle$ is the correlation function of the sound speed fluctuation. In this first-order scattering model, we note that the scattered intensity is determined solely by the second-order statistics of the random sound speed and density variations.

The technique of Fourier synthesis can be applied if the scattered field as a function of time needs to be calculated,

$$y(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} P_s(\mathbf{R_r},\omega)S(\omega)\exp(i\omega t)d\omega, \quad (16)$$

where $S(\omega)$ is the spectrum of the source signal.

The validity of the above volume scattering formula is not restricted to layered fluid media. As long as the scatterers are located in a fluid layer, the bottom can be a fluid, fluid-elastic, or elastic layered medium. The random scattering layer is not necessarily immediately beneath the water/sediment interface. The attenuation of both compressional waves and shear waves and the background sound speed gradient are easily incorporated into the model because they are included in the calculation of the Green's function by OASES. This theory will be suitable for studying scattering by a sediment bottom, where interface roughness does not dominate the scattering process. This is exactly the situation that we encountered in the ARSRP scattering experiment.

B. Description of random sound speed and density fluctuations

The second-order statistics of sound speed and density variations are of great importance to volume scattering modeling, as can be seen in the previous subsection. However, it is difficult to have direct measurement of such sediment properties, let alone a good model of them. Traditionally, Gaussian and exponential correlation functions have often been used because of their mathematical simplicity.[13,14,17] Recently, correlation function with a power-law type power spectrum has been shown to be more closely related to real sediments.[4,18,19] Model/data comparisons in the next section will reveal that the best fit is obtained by a power-law spectrum. For conciseness, description of only the power-law distribution will be given in the following.

In the last subsection, we defined two random variables, the relative sound speed variation ϵ and the density variation. Since they are assumed to be fully correlated, only the statistics of ϵ will be needed. In addition to the zero mean assumption regarding ϵ, we require that the sound speed fluctuation be a homogeneous random variable. In other words, the process is spatially wide-sense stationary, by which we mean that the covariance of the random variable is invariant with respect to spatial translation,

$$\langle \epsilon(\mathbf{r_1})\epsilon(\mathbf{r_2})\rangle = C(\mathbf{r_1}-\mathbf{r_2}). \quad (17)$$

One may argue that stationarity can never be achieved in a real environment. Nevertheless, from a practical point of view, it is a reasonable assumption if we confine ourselves to a small enough region so that the change across the whole region is little. A stationary random process can be characterized in the wave number domain by its power spectrum

$$W(\mathbf{k}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} C(\mathbf{r})\exp(-i\mathbf{k}\cdot r)d\mathbf{r}, \quad (18)$$

where $\mathbf{k}=(k_x,k_y,k_z)$ is the wave number vector.

A power-law type of power spectrum has been used to describe turbulence (Tatarski[20]), to characterize heterogeneity in the earth's crust (Frankel and Clayton[19]), and to model the seafloor morphology (Goff and Jordan[21]). In this work, we will follow Goff and Jordan's development and propose a power-law type of power spectral density function with a cutoff wave number. The power spectrum for a two-dimensional random field is

$$W(k_x,k_y) = \begin{cases} \dfrac{4\pi\nu\sigma^2 l_x l_y}{(1+k_x^2l_x^2+k_y^2l_y^2)^{\nu+1}} & \nu \neq 0 \\ \dfrac{4\pi\sigma^2 l_x l_y}{(1+k_x^2l_x^2+k_y^2l_y^2)} & \nu = 0 \end{cases}. \quad (19)$$

An asymptotic roll-off rate $-2(\nu+1)$ of the spectrum is related to the fractal (Hausdorff) dimension D_{fractal} of such a stochastic process as:

$$D_{\text{fractal}} = 3-\nu. \quad (20)$$

Therefore, decreasing the parameter ν would increase the randomness, which is the case for the three-dimensional case as well. In the special case of $\nu=1/2$, the power spectrum reduces to the isotropic exponential power spectrum when $l_x=l_y$. For a correlated three-dimensional case, the power spectrum is

$$W(k_x,k_y,k_z) = \frac{8\pi\sqrt{\pi}l_x l_y l_z\sigma^2\Gamma(\nu+3/2)}{\Gamma(\nu)(1+k_x^2l_x^2+k_y^2l_y^2+k_z^2l_z^2)^{\nu+3/2}}, \quad (21)$$

where Γ represents the Gamma function. The roll-off rate now is $-2(\nu+3/2)$ and the related fractal (Hausdorff) dimension is

$$D_{\text{fractal}} = 4-\nu. \quad (22)$$

C. Generation of 3D random field: The spectral method

As derived in subsection A, scattered fields can be obtained from Eq. (14), applying first-order perturbation theory. In a backscattering scenario, where the source and receiver are on the same vertical axis, only the sound speed fluctuation ϵ in the above equation is dependent on azimuth ϕ if the integral is carried out in cylindrical coordinates. Consequently,

$$\eta(r,z)=\int_0^{2\pi}\epsilon(r,\phi,z)\,d\phi, \tag{23}$$

can be generated to reduce a 3D problem to a 2D one.

The spectral method for generating an azimuthally summed 3D random field $\eta(r,z)$ will be presented in this subsection. For the sake of simplicity, generation of azimuthally summed 2D random fields will be studied first. In other words, the dimension z (corresponding to the vertical direction) is dropped for the time being. However, it can be added to the formulas in a straightforward manner.

In Cartesian coordinates, the power spectral density W can be obtained as

$$W(k_x,k_y)=\lim_{L\to\infty}\frac{1}{L^2}\left\langle\left|\int_{-L/2}^{L/2}\int_{-L/2}^{L/2}\epsilon(x,y)e^{-i(k_xx+k_yy)}\,dx\,dy\right|^2\right\rangle, \tag{24}$$

where k_x, k_y are the wave numbers in the x, y directions, L^2 is the patch size assuming a square area, $\langle\cdot\rangle$ stands for the ensemble average operation and $|\cdot|$ indicates the modulus operation. The power spectral density should be normalized such that

$$\frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}W(k_x,k_y)dk_x\,dk_y=\sigma^2, \tag{25}$$

where σ is the standard deviation of the sound speed perturbation. Equation (24) can be proved to satisfy the relation[22]

$$W(k_x,k_y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}C(\xi_x,\xi_y)e^{-i(k_x\xi_x+k_y\xi_y)}\,d\xi_x\,d\xi_y, \tag{26}$$

where $C(\xi_x,\xi_y)$ is the correlation function and ξ_x, ξ_y are the lag distances between two points in the x and y directions. The consistency between Eq. (24) and Eq. (26) has an important implication. It guarantees that if the power spectral density of the generated random field, estimated through Eq. (24), converges to the desired one, the generated random field would have the desired correlation function according to Eq. (26), and therefore the desired second-order statistics.

According to Eq. (24), a random field can be generated by

$$\epsilon(x,y)=\frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sqrt{W(k_x,k_y)L^2}R_n\times(k_x,k_y)e^{i(k_xx+k_yy)}\,dk_x\,dk_y. \tag{27}$$

Here the random number R_n fits the distribution $R_n(k_x,k_y)\sim[N(0,1)+iN(0,1)]/\sqrt{2}$, where $N(0,1)$ stands for a normal distribution with zero mean and standard deviation 1. In order for $\epsilon(x,y)$ to be real, $R_n(k_x,k_y)$ has to be conjugate symmetric between the first and the third quadrants as well as the second and the fourth quadrants. If the following coordinate transformations

$$\begin{cases}k_x=k_r\cos\theta\\ k_y=k_r\sin\theta\end{cases}\quad\text{and}\quad\begin{cases}x=r\cos\phi\\ y=r\sin\phi,\end{cases} \tag{28}$$

are performed, Eq. (27) becomes

$$\epsilon(r,\phi)=\frac{1}{(2\pi)^2}\int_0^{\infty}\int_0^{2\pi}\sqrt{W(k_r,\theta)L^2}R_n\times(k_r,\theta)e^{ik_rr\cos(\theta-\phi)}k_r\,dk_r\,d\theta, \tag{29}$$

and the equation for $\eta(r)$ is

$$\int_0^{2\pi}\epsilon(r,\phi)d\phi=\frac{1}{2\pi}\int_0^{\infty}\int_0^{2\pi}\sqrt{W(k_r,\theta)L^2}R_n\times(k_r,\theta)k_rJ_0(k_rr)dk_r\,d\theta. \tag{30}$$

We point out that $\eta(r)$ is a *nonstationary* process. Here $d\phi=2\pi/N(r)$, J_0 is the zeroth-order Bessel function, and $N(r)$ is the number of scatterers around the circle with radius r. Assuming constant density for the scatterers, $N(r)$ will increase proportionally to r.

Therefore, the variance of $\eta(r)$ will decrease accordingly, which makes $\eta(r)$ a *nonstationary* process. Collecting the items on the right-hand side which depend on θ, we have

$$F(k_r)=\frac{1}{2\pi}\int_0^{2\pi}\sqrt{W(k_r,\theta)L^2}R_n(k_r,\theta)d\theta. \tag{31}$$

Changing the continuous integral to a discrete sum and using the relation

$$\frac{(2\pi)^2}{L^2}=\Delta k_x\Delta k_y=k_r\Delta k_r\Delta\theta\quad(k_r\neq0),$$

where $\Delta(\cdot)$ represents the sampling interval, Eq. (31) becomes

$$F(k_r)=\sum_{i=1}^{M(k_r)}\sqrt{\frac{W(k_r,\theta_i)\Delta\theta}{k_r\Delta k_r}}R_n(k_r,\theta_i), \tag{32}$$

where $\theta_i=i\Delta\theta$ and $M(k_r)$ is the number of samples around a circle of radius k_r. The value of F at $k_r=0$ can be arbitrary as long as it is finite because it will be multiplied by k_r, which is 0 here, in Eq. (30). $F(k_r)$ can be treated as a new random variable with mean 0 and variance

$$\sum_{i=1}^{M(k_r)}\frac{W(k_r,\theta_i)\Delta\theta}{k_r\Delta k_r}\quad(k_r\neq0),$$

since $R_n(k_r,\theta_i)$ are independent of each other and each has the variance of 1. Therefore, $F(k_r)$ can be replaced by

$$\sqrt{\frac{\int_0^{2\pi}W(k_r,\theta)d\theta}{k_r\,dk_r}}R_n(k_r)\quad(k_r\neq0),$$

where $R_n(k_r)\sim N(0,1)$. Notice that here $R_n(k_r)$ is a real number due to the fact that the imaginary parts of $R_n(k_r,\theta)$ have been canceled because of the earlier mentioned property of conjugate symmetry. Equation (30) now becomes

$$\eta(r)=\sum_{i=1}^{\infty}\sqrt{\frac{\int_0^{2\pi}W(k_{r_i},\theta)d\theta}{k_{r_i}\Delta k_r}}R_n(k_{r_i})k_{r_i}J_0(k_{r_i},r)\Delta k_r, \tag{33}$$

where $k_{r_i}=i\Delta k_r$.

As mentioned above, the correct second-order statistics of the generated $\eta(r)$ should be guaranteed. From Eq. (26), the correlation function of ϵ in Cartesian coordinates can be written as

$$\langle \epsilon(x_1,y_1)\epsilon^*(x_2,y_2)\rangle = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W(k_x,k_y)e^{i[k_x(x_2-x_1)+k_y(y_2-y_1)]} \times dk_x\,dk_y. \quad (34)$$

If the coordinate system is again changed to polar coordinates, we obtain

$$\langle \epsilon(r_1,\phi_1)\epsilon^*(r_2,\phi_2)\rangle = \frac{1}{(2\pi)^2}\int_0^{\infty}\int_0^{2\pi} W(k_r,\theta)e^{ik_r[r_1\cos(\theta-\phi_1)-r_2\cos(\theta-\phi_2)]}k_r \times dk_r\,d\theta. \quad (35)$$

Therefore, the correlation function of $\eta(r)$ is

$$\langle \eta(r_1)\eta^*(r_2)\rangle = \int_0^{2\pi}\int_0^{2\pi}\langle \epsilon(r_1,\phi_1)\epsilon^*(r_2,\phi_2)\rangle d\phi_1\,d\phi_2 = \int_0^{\infty}\left[\int_0^{2\pi} W(k_r,\theta)d\theta\right]k_r J_0(k_r r_1)J_0 \times (k_r r_2)dk_r. \quad (36)$$

It can be seen that $\eta(r)$ generated by Eq. (33) satisfies the above equation.

Let us denote

$$W_s = \int_0^{2\pi} W(k_r,\theta)d\theta,$$

which can be considered the *radial power spectral density*. It is instructive to compute the spectral estimates $\hat{W}_s$ from the random realizations of sound speed perturbations generated by Eq. (33) for comparison with W_s. In continuum notation, we calculate

$$\hat{W}_s = \frac{2\pi k_r}{L}\left\langle \left| \int_0^{L/2} w(r)\,\eta(r) r J_0(k_r r)dr \right|^2 \right\rangle, \quad (37)$$

where $w(r)$ is a real non-negative weighting or window function normalized such that

$$\frac{2}{L}\int_0^{L/2} w^2(r)dr = 1. \quad (38)$$

The weighting function is chosen to improve the spectral estimation by reducing an effect known as spectral leakage. So if the estimated $\hat{W}$ agrees with the desired W_s, the generated $\eta(r)$ can be considered satisfactory. Equation (37) has been used in numerical experiments[10] as the criterion to determine the acceptability of the generated random field.

Equation (33) can be easily extended to the generation of a 3D azimuthally-summed random field in cylindrical coordinates:

$$\eta(r,z) = \frac{1}{2\pi}\sum_{m=0}^{\infty}\sum_{n=-\infty}^{\infty}\sqrt{\frac{D\int_0^{2\pi}W(k_{r_m},k_{z_n},\theta)d\theta}{k_{r_m}\Delta k_r}} \times R_n(k_{r_m},k_{z_n})k_{r_m}J_0(k_{r_m}r)e^{ik_{z_n}z}\Delta k_r\Delta k_z, \quad (39)$$

where $k_{r_m}=m\Delta k_r$, $k_{z_n}=n\Delta k_z$, and $R_n(k_r,k_z)$ has to be conjugate symmetric in the z direction in order for $\eta(r,z)$ to be real. The layer thickness is denoted by D. This process can be viewed as passing the random series $\eta(r)$ generated according to Eq. (33) through a filter in the z direction. As for the power spectral density $W(k_r,k_z,\theta)$ used in the generation, it can be in any form, including the result of a direct measurement in an experiment.

If the area within a few wavelengths distance from the origin is not a concern, Eq. (39) can be evaluated using a Fourier transform instead of a Hankel transform in the r direction and a Fourier transform in the z direction, which is essentially the same as the Fast-Field-Program (FFP) approach introduced by DiNapoli and Deavenport.[23] First the Bessel function $J_0(k_r r)$ is expressed in terms of Hankel functions,

$$J_0(k_r r) = \tfrac{1}{2}[H_0^{(1)}(k_r r) + H_0^{(2)}(k_r r)], \quad (40)$$

where $H_0^{(1)}$ and $H_0^{(2)}$ are the zeroth-order Hankel functions of the first and second kind, respectively. Next, $H_0^{(1)}(k_r r)$ and $H_0^{(2)}(k_r r)$ are replaced by their asymptotic forms[24]

$$\lim_{k_r r\to\infty} H_0^{(1)}(k_r r) = \sqrt{\frac{2}{\pi k_r r}}e^{i(k_r r-\pi/4)}, \qquad \lim_{k_r r\to\infty} H_0^{(2)}(k_r r) = \sqrt{\frac{2}{\pi k_r r}}e^{-i(k_r r-\pi/4)}, \quad (41)$$

to arrive at the following expression,

$$\eta(r,z) = \frac{\Delta k_z}{2\pi}\sqrt{\frac{\Delta k_r}{2\pi r}} \times \sum_{m=0}^{\infty}\sum_{n=-\infty}^{\infty}\left[e^{-i(\pi/4)}\sqrt{D\int_0^{2\pi}W(k_{r_m},k_{z_n},\theta)d\theta} \times R_n(k_{r_m},k_{z_n})e^{i(k_{r_m}r+k_{z_n}z)} + e^{i(\pi/4)}\sqrt{D\int_0^{2\pi}W(k_{r_m},k_{z_n},\theta)d\theta} \times R_n(k_{r_m},k_{z_n})e^{-i(k_{r_m}r-k_{z_n}z)}\right]. \quad (42)$$

Numerical experiments[10] have shown that the random fields generated by the above scheme possess the desired second-order statistics, although we will not present the results here for the sake of conciseness.

III. MODEL/DATA COMPARISONS

A. ARSRP backscattering experiment scenario

On July 24–26, 1993, as part of the Acoustic Reverberation Special Research Program (ARSRP), a bottom backscattering experiment was conducted over the Site A sedi-

ment pond located at 26 °11′N, 46 °09′W in the vicinity of the Mid-Atlantic Ridge.[25] The water depth was about 4400 m for most of the sediment pond area and the sound speed profile was a typical linear deep ocean profile obtained from XBT measurements. The gradient was 0.0154/s for the water column close to the seafloor. The sound speed in the water at the water-sediment interface was extrapolated to be 1530 m/s. The top layer beneath the water-sediment interface was made of calcareous mud with a sound speed of 1510 m/s according to core data.[26] No direct measurements of the attenuation coefficient or the sound speed gradient were available in the sediment area where the experiment was conducted.

A chirp source (DTAGS[27]) and vertical line array were deep-towed 200 m to 400 m above the seafloor. The receiver was a 24 element vertical line array (VLA) hung beneath the source, with a spacing between adjacent hydrophone elements of 2 m. For areas where the backscattered data were collected, the sediment was at least 100 m thick. For a transmitted acoustic source level of approximately 200 dB *re*: 1 μPa @1 m and a source frequency of 250–650 Hz, the acoustic signal can penetrate to the rock basement beneath the sediment layer in spite of sediment attenuation and geometrical spreading. The duration of the source signal was 0.125 s. For each ping, the raw waveform was recorded on all hydrophones for 4.5 s starting 0.1 s before each chirp trigger with a sampling interval of 0.432 ms. The sampling interval was 0.432 ms. During the experiment, the ship moved from southwest toward northeast across the sediment pond area. More details about the experiment can be found in Refs. 10, 7, 8, 25.

B. Data analysis results

Previous analysis of the ARSRP experimental data revealed that the inhomogeneities within two irregular layers beneath the water/sediment interface were the primary contributors to oblique backscattering. Detailed description of the data processing techniques are given in Refs. 7, 8. Only the data analysis results are introduced in the following. Taking advantage of the vertical line array, endfire beamforming yields signals which are dominated by normal incidence returns. With the source/receiver moving across the sediment pond, sediment profiling can be obtained by aligning the normal incidence returns with respect to reflection arrivals at the water/sediment interface. After employing an edge detection algorithm,[28] the inferred sediment structure is shown in Fig. 1 at the east side of the sediment pond. Evidently, layering with gentle horizontal changes is the main characteristic of the sediment. However, some random features can be seen in two irregular layers beneath the water/sediment interface, from 0.02 s to 0.045 s and from 0.065 s to 0.11 s. A geological interpretation suggests that they are most likely to be chaotic bedding or turbidity layers.[29] What interests us is to model these inhomogeneities contribute to backscattering in oblique directions.

Previous analyses of part of the data[30,8] indicate that these irregular layers are the main scatterers. Before going into modeling in the next subsection, we elaborate further the point that inhomogeneous layers are the dominant scatterers.

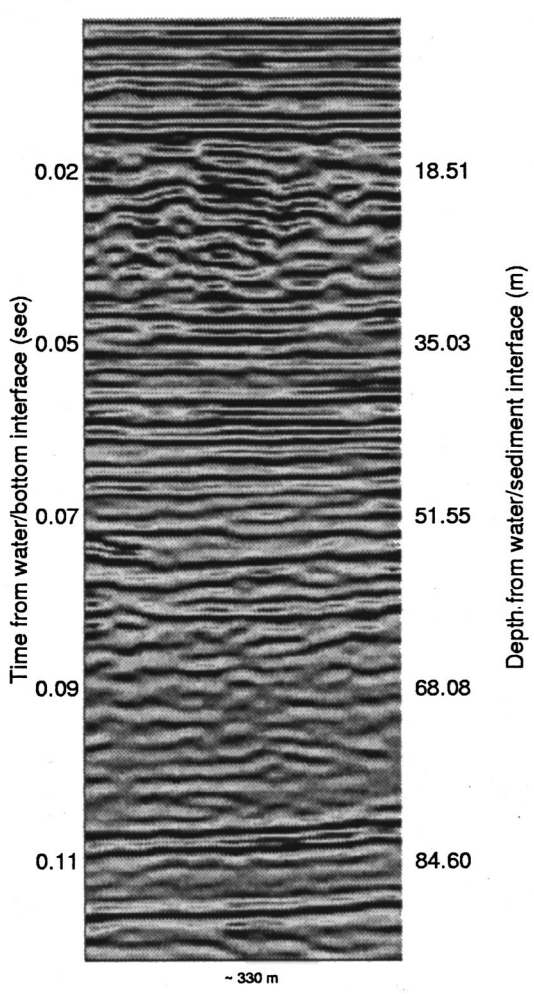


FIG. 1. The inferred sediment structure at the east side of the sediment pond.

The backscattered returns at oblique grazing angles are much weaker than the returns at normal and near normal incidence angles. Due to the sidelobe effect, the usual conventional beamforming cannot provide enough attenuation to the latter so as not to contaminate the oblique returns. Therefore, a hybrid method has been applied:[7,8] The returns in the normal and near normal incidence directions, obtained through endfire beamforming, are first subtracted from the total returns; the backscattered returns at oblique grazing angles can then be isolated by multiple constraints beamforming. Averaging the envelope of the beamformed results over eight consecutive pings and further smoothing the curve by a low-pass filter allows detection of any potential trends. The averaged envelope is shown in Fig. 2. Three groups of peaks appear in each look direction. Assuming that these peaks occur at horizontal interfaces and using a nominal sound speed of 1530 m/s, we obtain the best fit indicated by the solid lines. Here time 0 corresponds to the arrival time of the first water/sediment interface reflection. In contrast to the third group of peaks, which is associated with sediment/basement interface interaction according to travel time, the first two interfaces would reside at the locations of the abovementioned two irregular layers, where planar interfaces do not appear to exist. Consequently, the probable sources of the observed

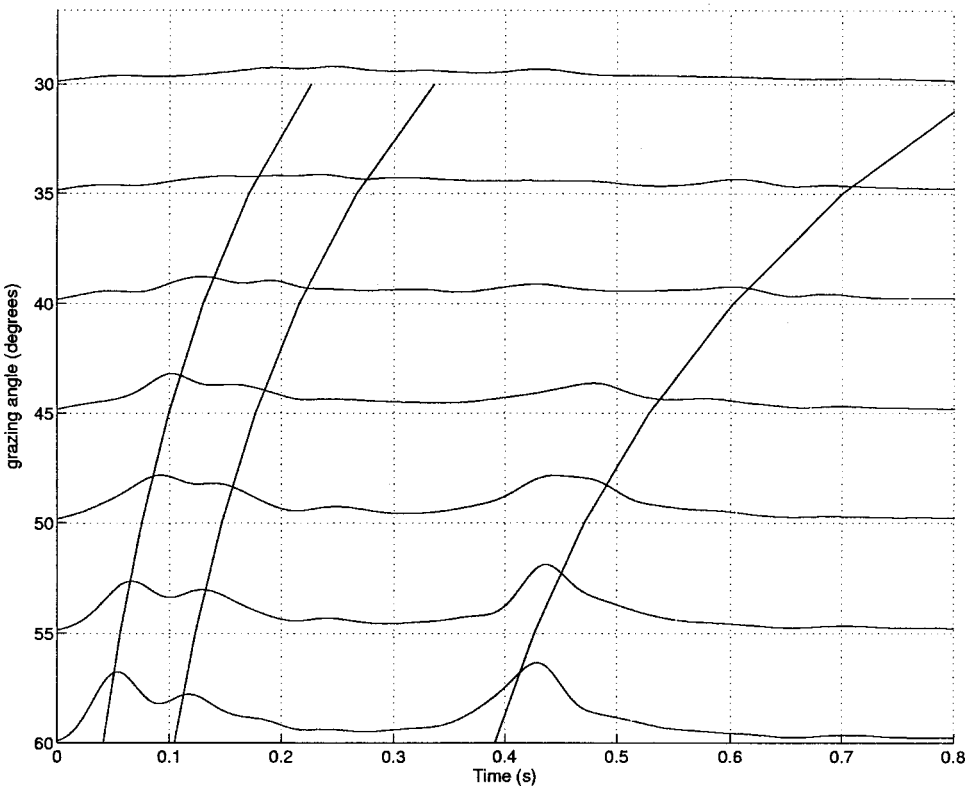


FIG. 2. The average envelope of oblique angle beamforming results over eight pings at the east side of the sediment pond (arbitrary units).

strong backscattering are the sediment inhomogeneities in the two irregular layers.

The same procedure is executed also for pings collected at the west side of the sediment pond.[10] Interestingly, the upper irregular layer is absent on this side of the sediment pond. Only one group of peaks, corresponding to the lower irregular layer, is observed instead of two groups of peaks in the two irregular layer case. This confirms our finding that the irregular layers beneath the water/sediment interface are the dominant contributors to backscattering in oblique directions.

Notice that the water/sediment interface contribution should have arrived at 0.02 s at 60 deg grazing angle in Fig. 2, the lack of which suggests the insignificance of backscattering at the water/sediment interface. This has been independently confirmed by Jeffrey and Ogden.[31]

C. The sediment background model

The sediment background model specifies the acoustic parameters governing the zeroth-order field. A simple fluid sediment background model is proposed to account for propagation to and from the scatterers, as shown in Fig. 3. All the parameters were carefully selected based on measurements or archival data. Table I lists the model parameters and their sources. Based on the arrival time of the sediment/basement interface returns, the average sound speed gradient in the sediment is estimated to be 1.12/s,[7] compared to a typical sound speed gradient of 1/s for the same type of sediment according to Hamilton.[32] From the amplitudes of the normal-incidence reflections at the water/sediment interface, together with the arrival time, the attenuation is estimated as 0.1 dB/λ, consistent with values in the literature[33,34] for the calcareous mud type of sediment in this area. The density of the sediment is chosen to be 1500 kg/m^3, well within the range for this type of sediment. The sound speed is assumed to be linear from the water/sediment interface down to 100 m. A false interface and an isovelocity bottom is added at 100 m beneath the water/sediment interface. As a result, the possible returns from beneath the false interface are ignored. Figure 2 shows that the only strong returns after the second group of peaks are due to the interactions at the sediment/basement interface. However, they would arrive later than backscattered returns at 30 deg grazing angle, which is the smallest angle at which backscattered returns can still be distinguished from the background noise. Therefore, a simple time-gating will eliminate their effect on model/data comparisons. The positions of the two irregular layers can be determined from the travel time obtained from bottom profiling shown in Fig. 1 together with the above proposed sound speed profile. A simple calculation indicates that the first irregular layer is located from $H_1 = 15.85$ m to $H_2 = 31.42$ m beneath the water/sediment interface, while the second one is from $H_3 = 56.42$ m to $H_4 = 85.7$ m. Notice that these numbers are different from those marked on the side of the bottom profiling plot where a nominal sound speed of 1530 m/s was used. The water/sediment interface was found to be flat as mentioned in the previous section and verified by the data analysis in Refs. 8, 31, which vindicates the choice of a range-independent water/sediment interface with no roughness in the model. All the above parameters are input to OASES so that the Green's function can be evaluated.

D. Estimation of the backscattering strength

To characterize the backscattering process quantitatively, we use the backscattering strength, recognizing the

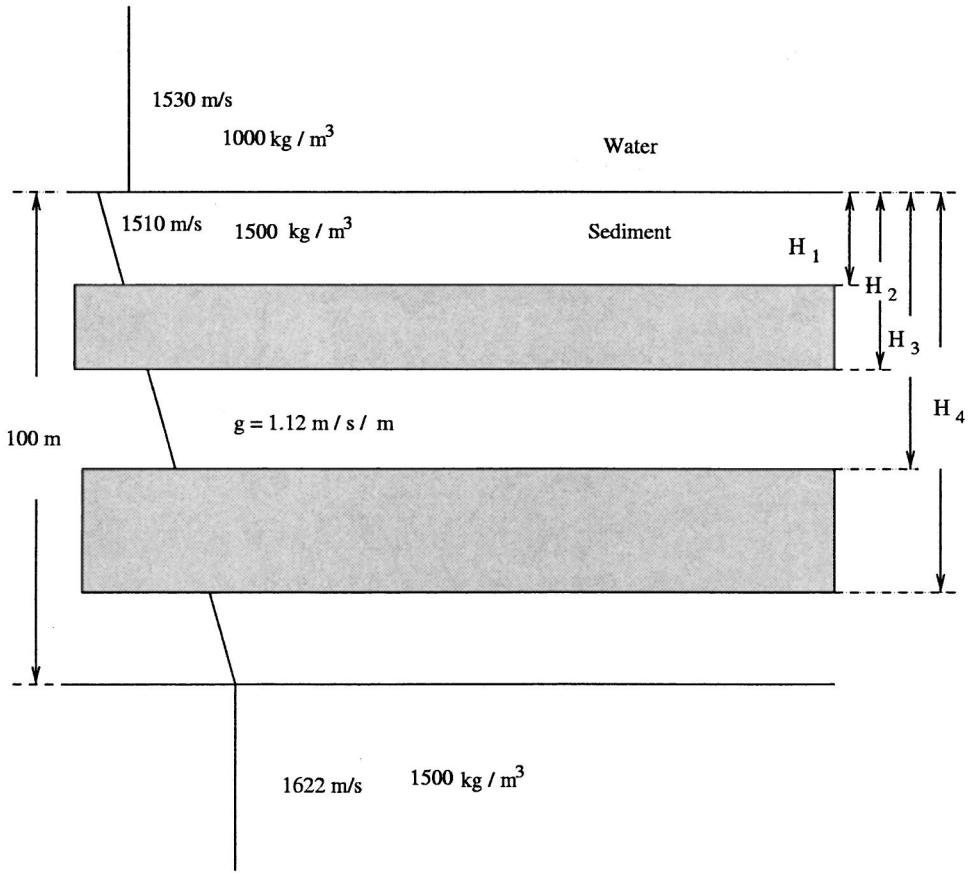


FIG. 3. Fluid bottom reverberation scenario.

need to be cautious in using the backscattering strength to describe a volume scattering process.[11] This quantity is appropriate when multi-path effects are insignificant, which is the case for backscattering from the two irregular layers found in the ARSRP experiment, even though reflections at layer interfaces and the sediment/basement interface complicate the picture. Here are the reasons: First, thanks to the vertical line array and beamforming, the oblique backscattered returns can be discriminated from reflections at layer interfaces, which makes the angular dependence of the estimated backscattering strength trustworthy; second, the multipath effect is not significant in an environment with a small sound speed gradient in the bottom and a lower sound speed at the top of the sediment compared to that in the water right above it; third, the interest is confined to the angular regime between 30 and 60 deg grazing angles, where the sediment/basement reflections arrive late as discussed before and can be filtered out by time-gating.

TABLE I. The parameter values and their sources for the sediment model.

Parameter	Value	Source
Sound speed in the water at the water/sediment interface	1530 m/s	Extrapolation of the CTD measurement
Sound speed in the sediment at the water/sediment interface	1510 m/s	Acoustic Lance Data
Sediment sound speed gradient	1.12/s	Estimated from the arrival time of the sediment/basement interface returns
Sediment attenuation	0.1 dB/λ	Estimated from the amplitudes of the sediment/basement interface reflections
Sediment density	1500 kg/m^3	Predicted based on sediment type and Hamilton results

The estimation of the backscattering strength from data is described in Ref. 8. The concentration here is on how to arrive at the frequency dependence of the backscattering strength. Let's start from

$$\mathrm{BSS} = 10 \log_{10} \frac{I_s R^2}{I_{\mathrm{in}} A}, \tag{43}$$

where I_s is the intensity of the backscattered signals at the reference receiver, I_{in} is the intensity of the incident sound wave a unit distance away from the scattering volume, R is the distance from the reference receiver to the top of the scattering volume, and A is the top area of the scattering volume.

A reference receiver is needed because we now have a vertical line array instead of a single receiver. In this data analysis, the middle hydrophone of the array (the twelfth from the top) is chosen as the reference receiver. Also, the directivity aspect of the array and subsequent beamforming makes it difficult to determine the insonified area precisely. Here, a simple case is considered where the returns are from the maximum response axis (MRA) of the array, i.e., the returned signals suffer little attenuation passing through the beamformer. As a result, A is estimated by the 3 dB area

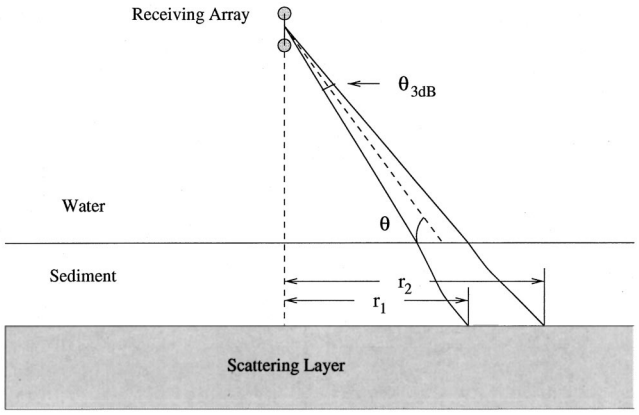


FIG. 4. Schematic illustration of the scattering strength estimation.

corresponding to the annulus where the array response is within 3 dB compared to the MRA. By applying ray tracing, A can be calculated by

$$A=\pi(r_2^2-r_1^2), \tag{44}$$

where r_1 and r_2 are the horizontal ranges between the reference receiver and the radial bounds of the 3 dB area (annulus) at the top of the scattering layer, as shown in Fig. 4. Bear in mind that the 3 dB beamwidth $\theta_{3\ \mathrm{dB}}$ and therefore the 3 dB insonified area will change with frequency. The lower the frequency, the wider the beam, the larger the insonified area. In order to get the backscattering strength for each frequency, we need to estimate $I_{\mathrm{in}}(f)$ and $I_s(f)$, where f stands for the frequency. To calculate $I_{\mathrm{in}}(f)$, the source signal recorded at the hydrophone is first Fourier transformed. The amplitude squared of the source spectrum is then adjusted for the spreading loss since $I_{\mathrm{in}}(f)$ should be measured a unit distance away from the scattering volume. Meanwhile, the calculation of $I_s(f)$ would involve one more procedure, which is to determine the exact piece of beamformed data at each angle that corresponds to the backscattered returns from each scattering layer. It can be expressed as follows,

$$I_s(f)=\langle|\mathrm{FFT}\{y(t)\}|^2\rangle, \quad T_1\leqslant t\leqslant T_2, \tag{45}$$

where $y(t)$ are the scattered returns, $\mathrm{FFT}\{\cdot\}$ stands for the Fourier transform, $|\cdot|$ stands for modulus, $\langle\cdot\rangle$ stands for ensemble average over different pings, and T_1 and T_2 are the travel times corresponding to arrivals from the top and the bottom of a scattering layer. Note that $y(t)$ along with T_1 and T_2 are different for each scattering angle. The backscattering strength versus grazing angle for each frequency can then be evaluated according to Eq. (43).

The two irregular scattering layers are treated separately. When computing the incident field on the lower irregular layer, the effect of the upper irregular layer on the transmitted field beneath it is neglected. This approach is consistent with the Born approximation employed in our scattering model since scattering from volume inhomogeneities is assumed to be weak. Taking into consideration the attenuation in the sediment, an amount of $2\alpha R_1$ will be added to the above backscattering strength, where α is the attenuation coefficient and R_1 is the ray-path length in the sediment.

Figures 5 and 6 show the estimated backscattering strength versus grazing angle and frequency. Notice that the grazing angle here and from now on is the angle measured at the water/sediment interface instead of at the top of the irregular region. However, due to the fact that the sound speed gradient is small, the largest difference, appearing at 30 deg, is 1.4 deg for the upper layer and 3.4 deg for the lower layer, values which are smaller than the beamwidth. Therefore, the

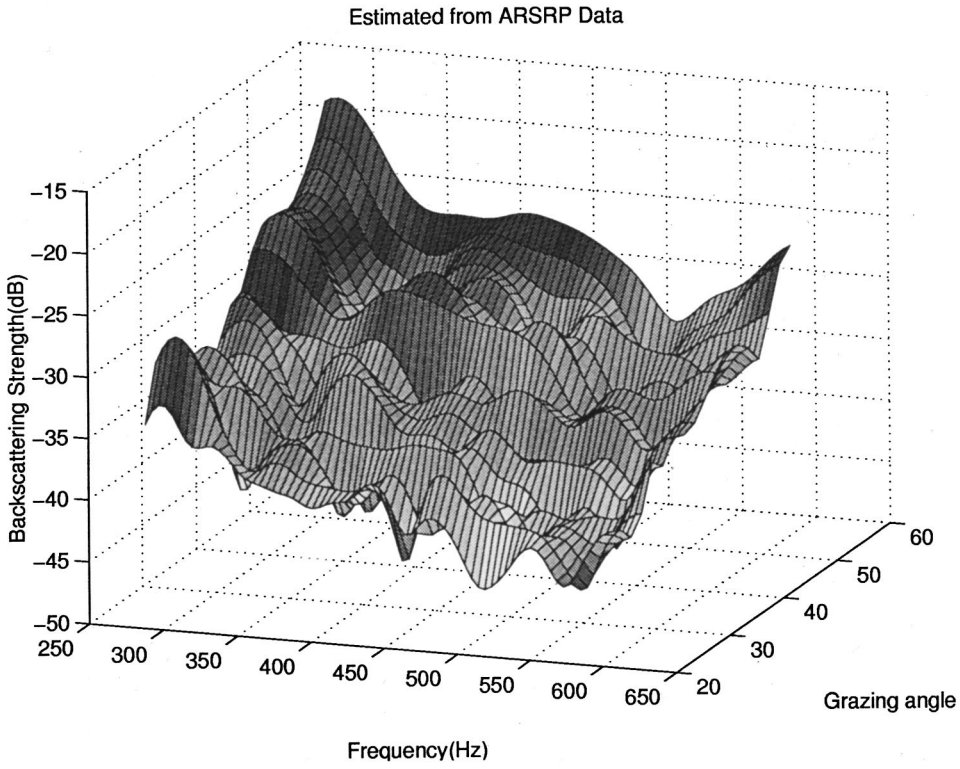


FIG. 5. The estimated backscattering strength for the upper irregular region at the east side of the sediment pond.

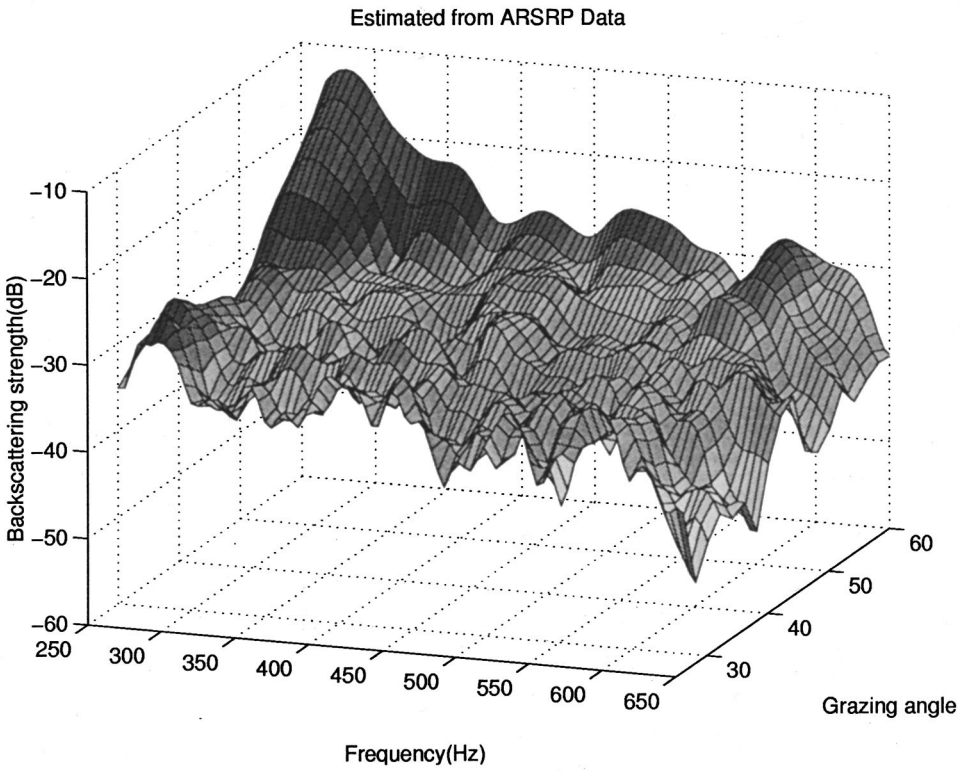


FIG. 6. The estimated backscattering strength for the lower irregular region at the east side of the sediment pond.

angular dependence obtained can be considered as reliable. It can be seen that, in general, the backscattering strength decreases with an increase of the frequency and/or a decrease of the scattering grazing angle. The widely used Lambert's law states that the backscattering strength is a function of $\sin^2\theta$ and the constant Mackenzie coefficient. It would fail to produce the observed frequency dependence of the backscattering strength.

In Fig. 6, there is a peak standing out at low frequencies and high grazing angles. The peak is only confined to grazing angles larger than 50 deg and frequencies less than 350 Hz, which looks artificial. It may be related to the aforementioned subtraction procedure applied to alleviate the strong normal and near normal incidence reflections. Since the subtraction can hardly be perfect in a real environment, the residue of the subtraction process would manifest itself primarily at large grazing angles. To make matters worse, the tilting of the layers beneath the lower irregular layer, as can be seen around 0.11 s in Fig. 1, may affect the subtraction process as well. Incidentally, the returns from the above tilting layers would have about the same arrival time as the scattered returns from the lower irregular layers at about a 60 deg grazing angle. In addition, there is a problem related to the beamformer used to obtain the oblique backscattered returns. The beam width of the main lobe is larger at lower frequencies and for look directions closer to normal incidence, which increases the chance of picking up those residues in the normal and near normal incidence directions. These may help explain the appearance of the above peak. To avoid this artificial peak, we will concentrate on the backscattering strength at frequencies larger than 300 Hz in the model/data comparisons.

Notice that the estimation of the backscattering strength from the data is related to the sediment background model, especially the attenuation coefficient. Because the scattered returns from the two irregular layers are identified by their travel times, errors in the sound speed profile would have little impact on the backscattering estimation. However, the attenuation coefficient would determine the compensation level of the backscattering strength at different scattering angles, therefore influencing the angular dependence of the backscattering strength as well as the level. In this study, we obtain the best estimate of the attenuation coefficient from the normal incidence returns as 0.1 dB/λ. According to the archival data compiled by Kibblewhite,[33] the attenuation coefficient for silt and clay sediment is between 0.01 dB/λ to 0.12 dB/λ. The example shown in Fig. 7 illustrates the sensitivity of the estimated backscattering strength at 450 Hz to the attenuation coefficient.

E. Model and data comparisons

Backscattering strength versus grazing angle is chosen to be compared between model simulation and experimental data for a selected set of frequencies over the signal spectrum. The model will simulate the scattered field at each hydrophone on the receiving array for selected frequencies using Eq. (14). Subsequently, the beamforming operation, the same as that used in experimental data analysis, will yield the backscattering intensity at different grazing angles for each frequency. The backscattering strength can then be estimated using Eq. (43). The entire model/data comparison process is summarized in Fig. 8.

Model/data comparisons will be carried out for the two irregular layers at the east side of the sediment pond only. We choose to compare at six frequencies: 300 Hz, 350 Hz,

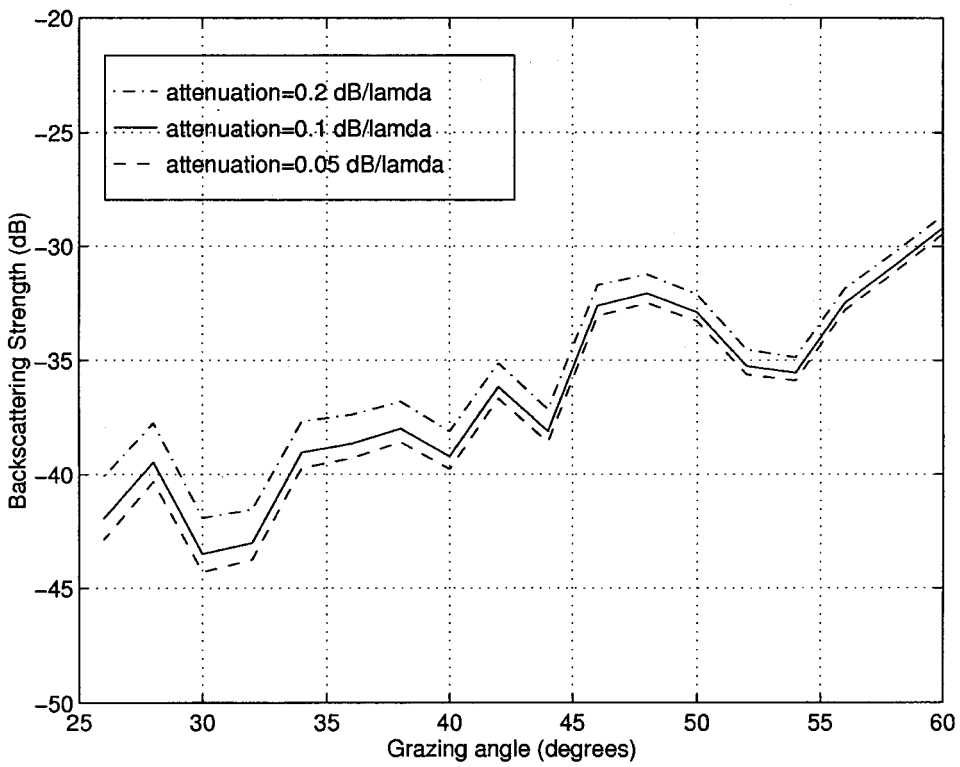


FIG. 7. The sensitivity of the estimated backscattering strength to the attenuation coefficient at 450 Hz for the upper irregular layer at the east side of the sediment pond.

400 Hz, 450 Hz, 500 Hz, and 550 Hz. As for the power spectrum and characteristic length scale to describe the inhomogeneous sound speed and density fluctuations, we elect to test the Gaussian distribution and the power law distribution. There is no evidence suggesting horizontal anisotropy in the ARSRP Site A sediment pond area. So we will simply let the horizontal correlation length $l_x = l_y$ in the model. Meanwhile, there is no measurement of the correlation length in this sediment area. Sediment profiling, shown in Fig. 1, suggests that the random field is much better correlated horizontally than vertically. This is consistent with the geoacoustic model proposed by Lysanov,[35] in which he suggested large-scale in the horizontal plane and small-scale in depth for random inhomogeneities in the sediment. The above sediment profiling covered a horizontal distance of about 330 m. A rough estimate indicates that the horizontal correlation length is on the order of 10 m, recognizing that this estimate is just the starting point of our parameter search. In the meantime, the standard deviation σ of the sound speed fluctuation can range from 1.5% to 8% and β, the ratio between density and sound speed fluctuations, between 1 and 10 according to Yamamoto.[36] Hines chooses β to be 2 in his model.[14] The density fluctuations, incorporated in the model by β, act like a dipole source with its major axis in the backscattering direction. Therefore, total backscattering will be enhanced. Also in Yamamoto's estimation, both sound speed and density fluctuations are characterized by a power-law distribution with ν [cf. Eq. (21)] between 0.4 and 0.65.

Comparisons are conducted by examinating the backscattering strength curves for both the model and data across all selected frequencies. As shown in Fig. 9, the best fit for the upper irregular layer is with a power law distribution. The parameters for the model are correlation lengths $l_x = l_y = 20\,\text{m}$, $l_x = 0.7\,\text{m}$, the standard deviation of sound speed fluctuations $\sigma = 2.8\%$, $\beta = 3$, and $\nu = 0.5$. The model predicted curves are ensemble averages of 200 realizations; therefore the standard deviations, shown as error bars, are small. Since l_x and l_z are different, even with $\nu = 0.5$, the power-law spectrum will not be equivalent to an exponential distribution. In general, the model curves agree with the data consistently. Notice that in the data curves, there is a small bump at about a 47 deg grazing angle for all the selected frequencies except 500 Hz. No specific cause has been found for this and the model cannot predict it, either. As for the parallel shift between the model and data curves at 300 Hz, it might be partly due to the ambient noise found to be peaking between 250 and 300 Hz.

Figure 10 shows the model/data comparison for the lower irregular layer at the east side of the sediment pond. The best fit again belongs to random sound speed fluctuations with a power law type of power spectrum. The horizontal correlation lengths are $l_x = l_y = 20\,\text{m}$, the vertical correlation length is 0.8 m, the standard deviation $\sigma = 2\%$, $\beta = 3$, and $\nu = 0.5$. Here 200 realizations are used to obtain the ensemble average. The matches are again satisfactory.

IV. CONCLUSIONS

Based on perturbation theory and the Born approximation, a volume scattering model has been developed which is capable of simulating the backscattered field and time series due to 3D volumetric bottom inhomogeneities, e.g., sound speed and density fluctuations. Using OASES as the propagation model in this work, we then have the ability to model volume scattering from a bottom with complicated sound

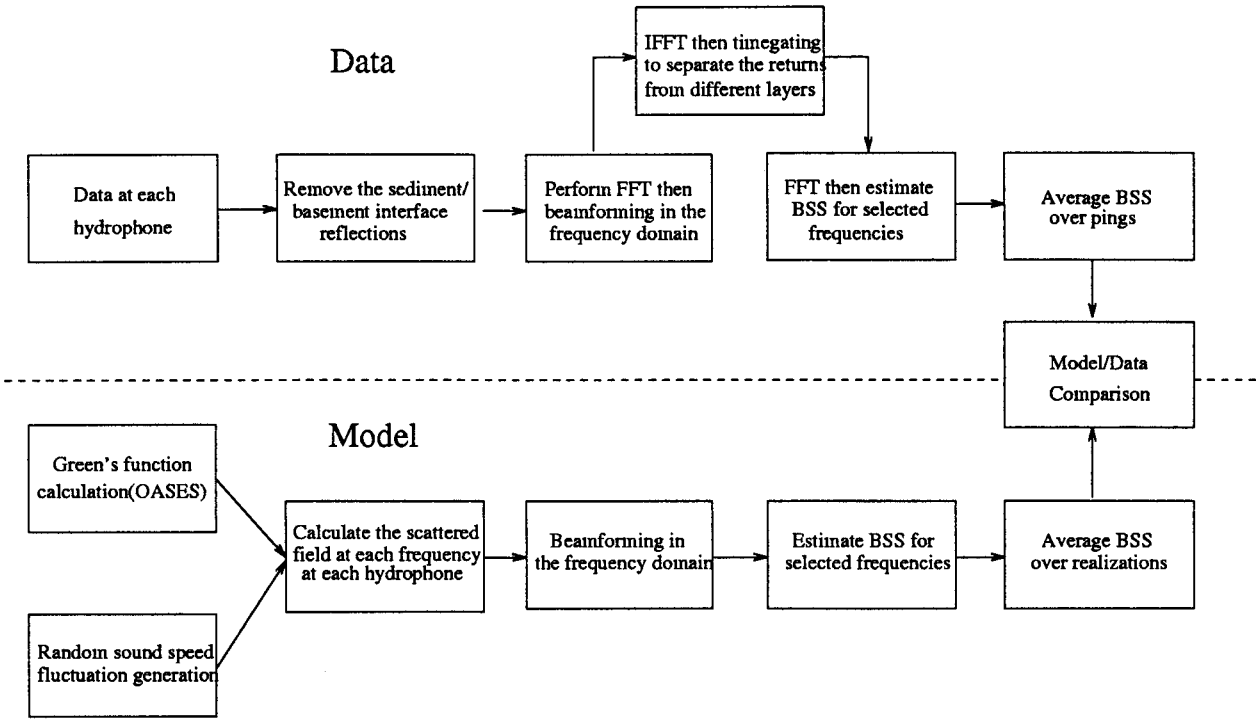


FIG. 8. The flow chart of the model/data comparison process.

speed structures such as sound speed gradients, multilayering, and even an elastic basement as long as the scatterers are located in a fluid layer.

While simulating scattering from full 3D volumetric inhomogeneities is still very computationally expensive, we have taken advantage of the monostatic backscattering configuration in this work, so that the azimuthally summed 3D random field can be generated with much relaxed computational requirements. It enables us to get a grip on the full scope of monostatic volume backscattering in a real environment.

In the model and ARSRP backscattering data comparison, a power-law distribution of random sound speed/density variations was found to fit the data very well in the frequency range of 250–650 Hz. Parameter wise, on the one hand, the horizontal correlation length is much larger than the vertical correlation length, which is consistent with some of the geophysical models of the sediment (so-called ''pancake'' mod-

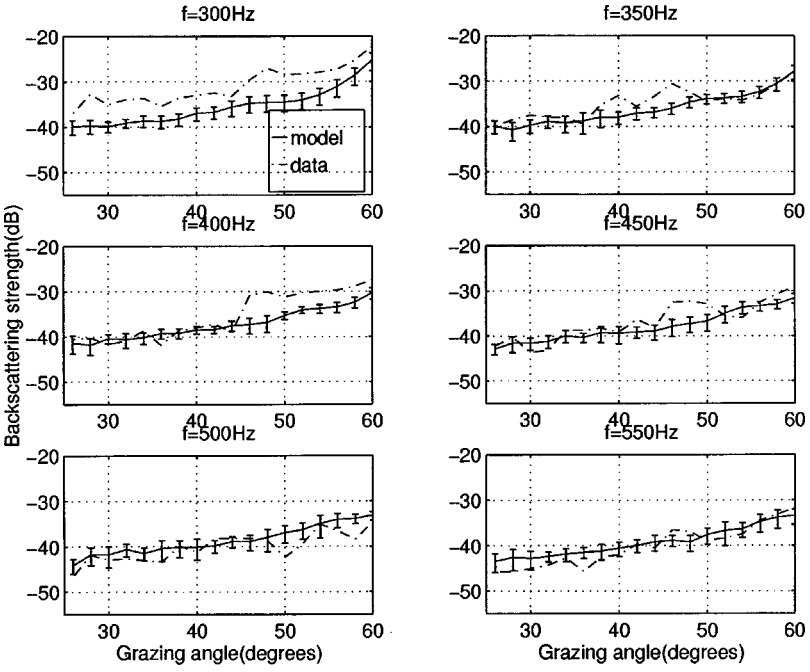


FIG. 9. Model/data comparison for the upper irregular layer at the east side of the sediment pond. Sound speed fluctuations are described by a power-law distribution with $l_x = l_y = 20$ m, $l_z = 0.7$ m, $\sigma = 2.8\%$, $\beta = 3$, and $\nu = 0.5$. The error bars show the standard deviation of the simulated backscattering strength.

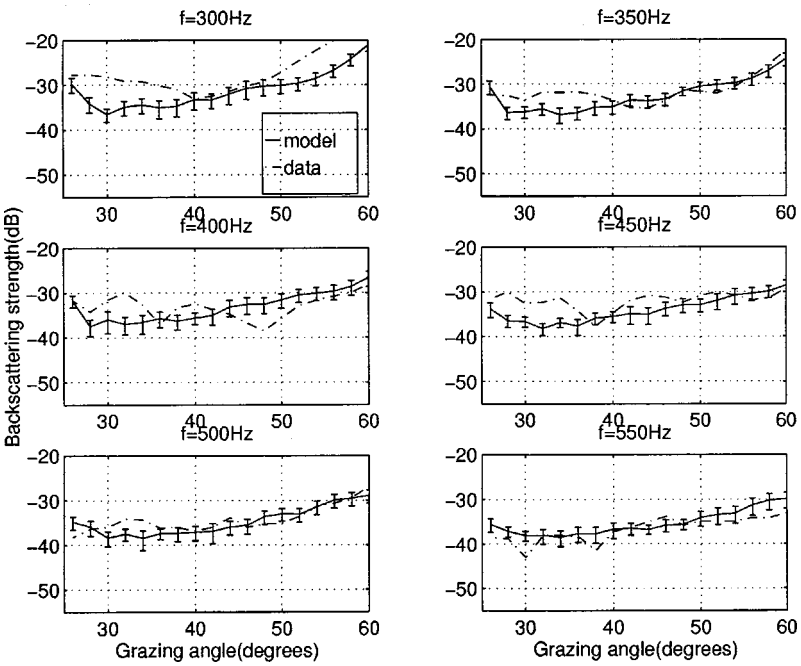


FIG. 10. Model/data comparison for the lower irregular layer at the east side of the sediment pond. Sound speed fluctuations are described by a power-law distribution with $l_x = l_y = 20$ m, $l_z = 0.8$ m, $\sigma = 2\%$, $\beta = 3$, and $\nu = 0.5$. The error bars show the standard deviation of the simulated backscattering strength.

els). The angular dependence is mainly determined by the decay rate of the corresponding power spectrum versus wave number, i.e., the fractal dimension, which is directly related to the sound speed/density fluctuations at small scales. Overall, this first-order volume scattering model has been shown to be capable of matching the ARSRP experimental data, and the power-law distribution was found to describe the sound speed and density fluctuations in the sediment satisfactorily.

[1] H. M. Merklinger, ''Bottom reverberation measured with explosive charges fired deep in the ocean,'' J. Acoust. Soc. Am. **44**, 508–513 (1968).

[2] D. R. Jackson, D. P. Winebrenner, and A. Ishimaru, ''Application of the composite roughness model to high-frequency bottom backscattering,'' J. Acoust. Soc. Am. **79**, 1410–1422 (1986).

[3] D. R. Jackson and K. B. Briggs, ''High-frequency bottom backscattering: Roughness versus sediment volume scattering,'' J. Acoust. Soc. Am. **92**, 962–977 (1992).

[4] D. R. Jackson, K. B. Briggs, K. L. Williams, and M. D. Richardson, ''Tests of models for high-frequency sea-floor backscatter,'' IEEE J. Ocean Eng. **21**, 458–470 (1996).

[5] A. P. Lyons, A. L. Anderson, and F. S. Dwan, ''Acoustic scattering from the seafloor: Modeling and data comparison,'' J. Acoust. Soc. Am. **95**, 2441–2451 (1994).

[6] D. Tang, G. Jin, D. R. Jackson, and K. L. Williams, ''Analyses of high-frequency bottom and subbottom backscattering for two distinct shallow water environments,'' J. Acoust. Soc. Am. **96**, 2930–2936 (1994).

[7] D. Tang, G. V. Frisk, C. J. Sellers, and D. Li, ''Low-frequency acoustic backscattering by volumetric inhomogeneities in deep-ocean sediments,'' J. Acoust. Soc. Am. **98**, 508–516 (1995).

[8] D. Li, ''Low-frequency bottom backscattering data analysis using multiple constraints beamforming,'' Master's thesis, MIT/WHOI Joint Program, Cambridge, MA and Woods Hole, MA, 1995.

[9] H. Schmidt, SAFARI: Seismo-acoustic fast field algorithm for range independent environments. User's Guide. Rep. SR-113, 1988. SACLANT Undersea Research Centre, La Spezia, Italy.

[10] D. Li, ''Modeling of Monostatic Bottom Backscattering from Three-Dimensional Volume Inhomogeneities and Comparisons with Experimental Data,'' Ph.D. thesis, MIT/WHOI Joint Program, Cambridge, MA and Woods Hole, MA, 1997.

[11] D. Li, D. Tang, and G. V. Frisk, ''Evaluation of sound propagation models used in bottom volume scattering studies,'' J. Acoust. Soc. Am. **108**, 2039–2052 (2000).

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

[13] L. A. Chernov, *Wave Propagation in a Random Medium* (McGraw-Hill, New York, 1975), translated from Russian by R. A. Silverman.

[14] P. C. Hines, ''Theoretical model of acoustic backscatter from a smooth seabed,'' J. Acoust. Soc. Am. **88**, 324–334 (1990).

[15] M. B. Priestly, *Spectral Analysis and Time Series* (Academic, New York, 1981).

[16] G. V. Frisk, *Ocean and Seabed Acoustics: A Theory of Wave Propagation* (Prentice-Hall, Englewood Cliffs, NJ, 1994).

[17] D. Tang, ''Acoustic Wave Scattering from a Random Ocean Bottom,'' Ph.D. thesis, MIT/WHOI Joint Program, Cambridge, MA and Woods Hole, MA, 1991.

[18] T. Yamamoto, ''Acoustic scattering in the ocean from velocity and density fluctuations in the sediments,'' J. Acoust. Soc. Am. **99**, 866–879 (1996).

[19] A. Frankel and R. W. Clayton, ''Finite difference simulations of seismic scattering: Implications for the propagation of short-period seismic waves in the crust and models of crustal heterogeneity,'' J. Geophys. Res. **91**, 6465–6489 (1986).

[20] V. I. Tatarski, *Wave Propagation in a Turbulent Medium* (McGraw-Hill, New York, 1961).

[21] J. A. Goff and T. H. Jordan, ''Stochastic modeling of seafloor morphology: Inversion of Sea Beam data for second-order statistics,'' J. Geophys. Res. **93**, 13589–13608 (1988).

[22] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures* (Wiley-Interscience, New York, 1971).

[23] F. R. DiNapoli and R. L. Deavenport, ''Theoretical and numerical Green's function solution in a plane layered medium,'' J. Acoust. Soc. Am. **67**, 92–105 (1980).

[24] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables* (Dover, New York, 1972).

[25] D. Tang, W. C. Burgess, S. D. Rajan, C. J. Sellers, G. V. Frisk, and D. Li, ''A summary of the direct-path scattering experiments in the ARSRP sediment pond and some preliminary results,'' in Fall ARSRP Research Symposium, December 1993, San Diego.

[26] S. Fu, R. Wilkens, and L. Frazer, ''Acoustic lance: New *in situ* seafloor velocity profiles,'' J. Acoust. Soc. Am. **99**, 234–242 (1996).

[27] J. F. Gettrust, J. H. Ross, and M. M. Rowe, ''Development of a low frequency, deep tow geoacoustics system,'' Sea Technology, September 1991.

[28] J. S. Lim, *Two-Dimensional Signal and Image Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1990).

[29] B. E. Tucholke, private communication.
[30] D. Tang, ''Shallow water reverberation due to sediment volume inhomogeneities,'' IEEE J. Ocean Eng. (submitted for publication).
[31] J. Jeffrey and P. M. Ogden, ''Bottom scattering strength measured in deep and shallow water using deep towed acoustics/geophysics system,'' NRL/FR/7140-97-9842, 1997, Naval Research Lab.
[32] E. L. Hamilton, ''Sound velocity gradients in marine sediments,'' J. Acoust. Soc. Am. **65**, 909–922 (1979).
[33] A. C. Kibblewhite, ''Attenuation of sound in marine sediments: A review with emphasis on new low-frequency data,'' J. Acoust. Soc. Am. **86**, 716–738 (1989).
[34] E. L. Hamilton, ''Compressional wave attenuation in marine sediments,'' Geophysics **37**, 620–646 (1972).
[35] Yu. P. Lysanov, ''Geoacoustic model of the upper sedimentary layer in shallow seas,'' Dokl. Akad. Nauk SSSR **251**, 714 (1980).
[36] T. Yamamoto, ''Velocity variabilities and other physical properties of marine sediments measured by crosswell acoustic tomography,'' J. Acoust. Soc. Am. **98**, 2235–2248 (1995).

The explicit secular equation for surface acoustic waves in monoclinic elastic crystals

Michel Destrade[a)]
Mathematics, Texas A&M University, College Station, Texas 77843-3368

(Received 11 December 2000; accepted for publication 25 January 2001)

The secular equation for surface acoustic waves propagating on a monoclinic elastic half-space is derived in a direct manner, using the method of first integrals. Although the motion is at first assumed to correspond to generalized plane strain, the analysis shows that only two components of the mechanical displacement and of the tractions on planes parallel to the free surface are nonzero. Using the Stroh formalism, a system of two second order differential equations is found for the remaining tractions. The secular equation is then obtained as a quartic for the squared wave speed. This explicit equation is consistent with that found in the orthorhombic case. The speed of subsonic surface waves is then computed for 12 specific monoclinic crystals. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1356703]

PACS numbers: 43.35.Pt [ANN]

I. INTRODUCTION

The modern theory of surface acoustic waves in anisotropic media owes most of its results to the pioneering works of A. N. Stroh. Although his two seminal articles[1,2] went largely unnoticed for a long time, their theoretical implications were far reaching, as many came to realize since their publication. Among others, Currie,[3] Barnett and Lothe,[4] Chadwick and Smith[5] were able to use his ''sextic formalism'' to address many problems, such as the existence of a single real secular equation for the wave speed, the existence of a limiting velocity (the smallest velocity of body wave solutions) which defines ''subsonic'' and ''supersonic'' ranges for the speeds, or numerical schemes to compute the polarization vectors and the speed of the surface wave. A comprehensive review of these topics can be found in a textbook by Ting.[6]

However precise numerical procedures might be, there is still progress to be made in the search for secular equations in analytic form. So far, explicit expressions have remained few. The secular equation for surface waves in orthorhombic crystals was established by Sveklo[7] as early as 1948 and later, Royer and Dieulesaint[8] proved that it could account for 16 different crystal configurations, such as tetragonal, hexagonal, or cubic. For monoclinic media, Chadwick and Wilson[9] devised a procedure to derive the secular equation, which is given as ''explicit, [...] *apart* from the solution of [a] bicubic equation.'' The object of this paper is to derive *one* expression for the secular equation which is *fully explicit*, when the surface wave propagates in monoclinic crystals.

A classical approach to the problem of surface waves in anisotropic crystals is to consider that a wave propagates with speed v in the direction x_1 of a material axis (on the free plane surface) of the material, and is attenuated along another material axis x_2, orthogonal to the free surface, so that the mechanical displacement $\mathbf{u}$ is written as $\mathbf{u}=\mathbf{u}(x_1+px_2-vt)$, where p is unknown. Then, assuming a complex exponential form for the displacement, the equations of motion are written in the absence of body forces and solved for p. Finally, the boundary conditions yield the secular equation for v. The principal mathematical difficulty arising from this procedure is that the equations of motion yield a sextic (generalized plane strain) or a quartic (plane strain) for p which in general are impractical to solve analytically, or even, as a numerical scheme suggests in the sextic case, are actually insoluble analytically (in the sense of Galois).[10]

In 1994, Mozhaev[11] proposed ''some new ideas in the theory of surface acoustic waves.'' He introduced a novel method based on first integrals[12] of the displacement components, which bypasses the sextic (or quartic) equation for p and yields directly the secular equation. He successfully applied this method to the case of orthorhombic materials. In the present paper, generalized plane strain surface waves in a monoclinic crystal with plane of symmetry at $x_3=0$ are examined. The method of first integrals is adapted in order to be applied to the tractions components on the planes $x_3=\text{const}$, rather than to the displacement components. This switch presents several advantages. First, the equations of motion, the boundary conditions, and eventually the secular equation itself, are expressed directly in terms of the usual elastic stiffnesses. Second, it makes it apparent that one of the traction components is zero and thus that, in this paper's context, generalized plane strain leads to plane stress. Third, the boundary conditions are written in a direct and natural manner, because they correspond to the vanishing of the tractions on the free surface and at infinite distance from this surface. Finally, this procedure can easily accommodate an internal constraint, such as incompressibility[13,14] (the secular equation for surface waves in incompressible monoclinic linearly elastic materials is obtained elsewhere).

The plan of the paper is the following. After a brief review of the basic equations describing motion in linearly elastic monoclinic materials (Sec. II), the equations of motion are written down in Sec. III for a surface acoustic wave with three displacement components which depend on two

a)Electronic mail: destrade@math.tamu.edu

coordinates, that in the direction of propagation and that in the direction normal to the free surface (generalized plane strain). Then in Sec. IV, it is seen that one of the traction components is identically zero (plane stress), and that consequently, so is one of the displacement components (plane strain). For the remaining two traction components, coupled equations of motion and the boundary conditions are derived in Sec. V. Finally in Sec. VI, the method of first integrals is applied and the secular equation for acoustic surface waves in monoclinic elastic materials is derived explicitly. As a check, the subcase of orthorhombic materials is treated, and numerical results obtained by Chadwick and Wilson[9] for some monoclinic materials are recovered.

II. PRELIMINARIES

First, the governing equations for a monoclinic elastic material are recalled. The material axes of the media are denoted by x_1, x_2, and x_3, and the plane $x_3=0$ is assumed to be a plane of material symmetry. For such a material, the relationship between the nominal stress $\boldsymbol{\sigma}$ and the strain $\boldsymbol{\epsilon}$ is given by[15]

$$\begin{bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{23} \\ \sigma_{31} \\ \sigma_{12} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & 0 & 0 & c_{16} \\ c_{12} & c_{22} & c_{23} & 0 & 0 & c_{26} \\ c_{13} & c_{23} & c_{33} & 0 & 0 & c_{36} \\ 0 & 0 & 0 & c_{44} & c_{45} & 0 \\ 0 & 0 & 0 & c_{45} & c_{55} & 0 \\ c_{16} & c_{26} & c_{36} & 0 & 0 & c_{66} \end{bmatrix} \begin{bmatrix} \epsilon_{11} \\ \epsilon_{22} \\ \epsilon_{33} \\ 2\epsilon_{23} \\ 2\epsilon_{31} \\ 2\epsilon_{12} \end{bmatrix}, \tag{1}$$

where c's denote the elastic stiffnesses, and the strain components ϵ's are related to the displacement components u_1, u_2, u_3 through

$$2\epsilon_{ij}=(u_{i,j}+u_{j,i}) \quad (i,j=1,2,3). \tag{2}$$

The equations of motion, written in the absence of body forces, are

$$\sigma_{ij,j}=\rho u_{i,tt} \quad (i=1,2,3), \tag{3}$$

where ρ is the mass density of the material, and the comma denotes differentiation.

Finally, the 6×6 matrix $\mathbf{c}$ given in Eq. (1) must be positive definite in order for the strain-energy function density to be positive.

III. SURFACE WAVES

Now the propagation of a surface wave on a semi-infinite body of monoclinic media is modeled. In the same manner as Mozhaev,[11] the amplitude of the associated displacement is assumed to be varying sinusoidally with time in the direction of propagation x_1, while its variation in the direction x_2, orthogonal to the free surface, is not stated explicitly. Thus calling v the speed of the wave, and k the associated wave number, the displacement components are written in the form

$$u_j(x_1,x_2,x_3,t)=U_j(x_2)e^{ik(x_1-vt)} \quad (j=1,2,3), \tag{4}$$

where the U's depend on x_2 only. For these waves, the planes of constant phase are orthogonal to the x_1-axis, and the planes of constant amplitude are orthogonal to the x_2-axis.

The stress-strain relations (1) reduce to

$$\begin{aligned} t_{11}&=ic_{11}U_1+c_{12}U_2'+c_{16}(U_1'+iU_2),\\ t_{22}&=ic_{12}U_1+c_{22}U_2'+c_{26}(U_1'+iU_2),\\ t_{33}&=ic_{13}U_1+c_{23}U_2'+c_{36}(U_1'+iU_2),\\ t_{32}&=c_{44}U_3'+ic_{45}U_3, \quad t_{13}=c_{45}U_3'+ic_{55}U_3,\\ t_{12}&=ic_{16}U_1+c_{26}U_2'+c_{66}(U_1'+iU_2), \end{aligned} \tag{5}$$

where the prime denotes differentiation with respect to kx_2, and the t's are defined by

$$\sigma_{ij}(x_1,x_2,x_3,t)=kt_{ij}(x_2)e^{ik(x_1-vt)} \quad (i,j=1,2,3). \tag{6}$$

The boundary conditions of the problem (surface $x_2=0$ free of tractions, vanishing displacement as x_2 tends to infinity) are

$$t_{i2}(0)=0, \quad U_i(\infty)=0 \quad (i=1,2,3). \tag{7}$$

Finally, the equations of motion (3) reduce to

$$\begin{aligned} &it_{11}+t_{12}'=-\rho v^2U_1, \quad it_{12}+t_{22}'=-\rho v^2U_2,\\ &it_{13}+t_{32}'=-\rho v^2U_3. \end{aligned} \tag{8}$$

At this point, a sextic formalism could be developed for the three displacement components U_1, U_2, U_3, and the three traction components t_{12}, t_{22}, t_{32}. However, it turns out that one of these traction components is identically zero, as is now proved.

IV. PLANE STRESS

It is known (see the Appendix of Stroh's 1962 paper,[2] and also Ting's book,[6] p. 66) that for a two-dimensional deformation of a monoclinic crystal with axis of symmetry at $x_3=0$, the displacements u_1 and u_2 are decoupled from u_3. Taking $u_3=0$ for surface waves, it follows from the stress–strain relationships (5) that $t_{13}=t_{32}=0$. Here, an alternative proof of this result is presented.

Using Eqs. $(5)_4$, $(8)_3$, and $(5)_5$, two first order differential equations for t_{32} and U_3 are found as

$$t_{32}=ic_{45}U_3+c_{44}U_3', \quad t_{32}'=(c_{55}-\rho v^2)U_3-ic_{45}U_3'. \tag{9}$$

These equations may be inverted to give U_3 and U_3' as

$$\begin{aligned} (c_{44}c_{55}-c_{45}^2-c_{44}\rho v^2)U_3&=ic_{45}t_{32}+c_{44}t_{32}',\\ (c_{44}c_{55}-c_{45}^2-c_{44}\rho v^2)U_3'&=(c_{55}-\rho v^2)t_{32}-ic_{45}t_{32}'. \end{aligned} \tag{10}$$

Differentiation of Eq. $(10)_1$ and comparison with Eq. $(10)_2$ yields the following second order differential equation for t_{32}:

$$c_{44}t_{32}''+2ic_{45}t_{32}'-(c_{55}-\rho v^2)t_{32}=0. \tag{11}$$

The boundary conditions (7) and Eq. $(5)_4$ imply that the stress component t_{32} must satisfy $t_{32}(0)=t_{32}(\infty)=0$. The only solution of this boundary value problem for the differ-

ential equation (11) is the trivial one. Consequently,

$$t_{32}(x_2)=0 \quad \text{for all} \quad x_2, \tag{12}$$

and so it is proved that, as far as the propagation of surface acoustic waves in monoclinic crystals with plane of symmetry at $x_3=0$ is concerned, generalized plane strain leads to plane stress.

It is also worth noting that by Eq. $(10)_1$, plane stress leads in turn to plane strain which, as an assumption, was not needed a priori. This result was obtained by Stroh[2] in a different manner: ''[when] there is a reflection plane normal to the x_3 axis, [...] there is no coupling of the displacement u_3 with u_1 and u_2; any two dimensional problem reduces to one of plane strain ($u_3=0$) and one of anti-plane strain ($u_1=u_2=0$).''

Now the equations of motion can be written for the remaining displacements and traction components.

V. EQUATIONS OF MOTION

Here, the equations of motion are derived, first as a system of four first order differential equations for the nonzero components of mechanical displacement and tractions, and then as a system of two second order differential equations for the tractions.

The stress–strain relations (5) and the equations of motion (8) lead to a system of differential equations for the displacement components U_1, U_2, and for the traction components t_1, t_2, defined by

$$t_1=t_{12}, \quad t_2=t_{22}. \tag{13}$$

This system is as follows

$$\begin{bmatrix}\mathbf{u}'\\ \mathbf{t}'\end{bmatrix}=\begin{bmatrix} i\mathbf{N}_1 & \mathbf{N}_2\\ -(\mathbf{N}_3+X\mathbf{1}) & i\mathbf{N}_1^T\end{bmatrix}\begin{bmatrix}\mathbf{u}\\ \mathbf{t}\end{bmatrix}, \tag{14}$$

where $\mathbf{u}=[U_1,U_2]^T$, $\mathbf{t}=[t_1,t_2]^T$, $X=\rho v^2$, and the 2×2 matrices $\mathbf{N}_1$, $\mathbf{N}_2$, and $\mathbf{N}_3$ are submatrices of the fundamental elasticity matrix $\mathbf{N}$, introduced by Ingebrigsten and Tonning.[16] Explicitly, $\mathbf{N}_1$, $\mathbf{N}_2$, $\mathbf{N}_3$ are given by[6]

$$\begin{aligned} -\mathbf{N}_1&=\begin{bmatrix} r_6 & 1\\ r_2 & 0\end{bmatrix}, \quad \mathbf{N}_2=\begin{bmatrix} s_{22} & -s_{26}\\ -s_{26} & s_{66}\end{bmatrix}=\mathbf{N}_2^T,\\ -\mathbf{N}_3&=\begin{bmatrix}\eta & 0\\ 0 & 0\end{bmatrix}=-\mathbf{N}_3^T,\end{aligned} \tag{15}$$

where the quantities r_2, r_6, s_{22}, s_{26}, s_{66}, and η are given in terms of the elastic stiffnesses as

$$\begin{aligned} &\Delta=\begin{vmatrix} c_{22} & c_{26}\\ c_{26} & c_{66}\end{vmatrix}=c_{22}c_{66}-c_{26}^2,\\ &r_6=\frac{1}{\Delta}(c_{22}c_{16}-c_{12}c_{26}), \quad r_2=\frac{1}{\Delta}(c_{12}c_{66}-c_{16}c_{26}),\\ &s_{ij}=\frac{1}{\Delta}c_{ij} \quad (i,j=2,6),\\ &\eta=\frac{1}{\Delta}\begin{vmatrix} c_{11} & c_{12} & c_{16}\\ c_{12} & c_{22} & c_{26}\\ c_{16} & c_{26} & c_{66}\end{vmatrix}\\ &\quad=c_{11}-\frac{c_{66}c_{12}^2+c_{22}c_{16}^2-2c_{12}c_{16}c_{26}}{c_{22}c_{66}-c_{26}^2}.\end{aligned} \tag{16}$$

Throughout the paper, it is assumed that the matrix $\mathbf{N}_3+X\mathbf{1}$ is not singular, which means that the surface wave propagates at a speed distinct from that given by $\rho v^2=\eta$. This assumption made, the second vector line of the system (14) yields

$$\mathbf{u}=i(\mathbf{N}_3+X\mathbf{1})^{-1}\mathbf{N}_1^T\mathbf{t}-(\mathbf{N}_3+X\mathbf{1})^{-1}\mathbf{t}'. \tag{17}$$

On the other hand, differentiation of the system (14) leads to

$$\begin{bmatrix}\mathbf{u}''\\ \mathbf{t}''\end{bmatrix}=\begin{bmatrix} -\mathbf{N}_1\mathbf{N}_1-\mathbf{N}_2(\mathbf{N}_3+X\mathbf{1}) & i(\mathbf{N}_1\mathbf{N}_2+\mathbf{N}_2\mathbf{N}_1^T)\\ -i[(\mathbf{N}_3+X\mathbf{1})\mathbf{N}_1+\mathbf{N}_1^T(\mathbf{N}_3+X\mathbf{1})] & -(\mathbf{N}_3+X\mathbf{1})\mathbf{N}_2-\mathbf{N}_1^T\mathbf{N}_1^T\end{bmatrix}\begin{bmatrix}\mathbf{u}\\ \mathbf{t}\end{bmatrix}. \tag{18}$$

Now the second vector line of this equation yields, using Eq. (17), a system of two second order differential equations for $\mathbf{t}$, written as

$$\hat{\alpha}_{ik}t_k''-i\hat{\beta}_{ik}t_k'-\hat{\gamma}_{ik}t_k=0, \tag{19}$$

where the symmetric 2×2 matrices $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\gamma}}$, are given by

$$\begin{aligned} \hat{\boldsymbol{\alpha}}&=-(\mathbf{N}_3+X\mathbf{1})^{-1},\\ \hat{\boldsymbol{\beta}}&=-\mathbf{N}_1(\mathbf{N}_3+X\mathbf{1})^{-1}-(\mathbf{N}_3+X\mathbf{1})^{-1}\mathbf{N}_1^T,\\ \hat{\boldsymbol{\gamma}}&=\mathbf{N}_2-\mathbf{N}_1(\mathbf{N}_3+X\mathbf{1})^{-1}\mathbf{N}_1^T,\end{aligned} \tag{20}$$

or, explicitly, by their components,

$$\begin{aligned} &\hat{\alpha}_{11}=\frac{1}{\eta-X}, \quad \hat{\alpha}_{12}=0, \quad \hat{\alpha}_{22}=-\frac{1}{X},\\ &\hat{\beta}_{11}=-\frac{2r_6}{\eta-X}, \quad \hat{\beta}_{12}=\frac{1}{X}-\frac{r_2}{\eta-X}, \quad \hat{\beta}_{22}=0,\\ &\hat{\gamma}_{11}=s_{22}+\frac{r_6^2}{\eta-X}-\frac{1}{X}, \quad \hat{\gamma}_{12}=\frac{r_2r_6}{\eta-x}-s_{26},\\ &\hat{\gamma}_{22}=\frac{r_2^2}{\eta-X}+s_{66}.\end{aligned} \tag{21}$$

The system (19) of second order differential equations for the traction components is more convenient to work with than the corresponding system for the displacement compo-

TABLE I. Values of the relevant elastic stiffnesses (GPa), density (kg m^{-3}), and surface wave speed (m s^{-1}) for 12 monoclinic crystals.

Material	c_{11}	c_{22}	c_{12}	c_{16}	c_{26}	c_{66}	ρ	v
aegirite-augite	216	156	66	19	25	46.5	3420	3382
augite	218	182	72	25	20	51.1	3320	3615
diallage	211	154	37	12	15	62.2	3300	4000
diopside	238	204	88	−34	−19	58.8	3310	3799
diphenyl	14.6	5.95	2.88	2.02	0.40	2.26	1114	1276
epidote	202	212	45	−14.3	0	43.2	3400	3409
gypsum	50.2	94.5	28.2	−7.5	−11.0	32.4	2310	3011
hornblende	192	116	61	10	4	31.8	3120	3049
microcline	122	66	26	−13	−3	23.8	2561	2816
oligoclase	124	81	54	−7	16	27.4	2638	2413
tartaric acid	46.5	93	36.7	−0.4	−12.0	8.20	1760	1756
tin fluoride	33.6	47.9	5.3	6.5	−5.1	12.9	4875	1339

nents, because the boundary conditions are simply written, using Eqs. (7), (5), and (13), as

$$t_i(0)=t_i(\infty)=0 \quad (i=1,2). \tag{22}$$

This claim is further justified in the next section, where the secular equation is quickly derived.

VI. SECULAR EQUATION

Now the method of first integrals is applied to the system (19). Mozhaev[11] defined the following inner product,

$$(f,\phi)=\int_0^\infty (f\overline{\phi}+\overline{f}\phi)dx_2, \tag{23}$$

and multiplying Eq. (19) by $i\overline{t_j}$ gives

$$\hat{\alpha}_{ik}D_{kj}+\hat{\beta}_{ik}E_{kj}+\hat{\gamma}_{ik}F_{kj}=0, \tag{24}$$

where the 2×2 matrices **D**, **E**, **F**, are defined by

$$D_{kj}=(it_k'',t_j), \quad E_{kj}=(t_k'',t_j), \quad F_{kj}=(t_k,it_j). \tag{25}$$

By writing down $F_{kj}+F_{jk}$, it is easy to check that the matrix **F** is antisymmetric. Integrating directly $E_{kj}+E_{jk}$, and integrating $D_{kj}+D_{jk}$ by parts, and using the boundary conditions Eq. (22), it is found that the matrices **E** and **D** are also antisymmetric. So **D**, **E**, and **F** may be written in the form

$$\mathbf{D}=\begin{bmatrix} 0 & D \\ -D & 0 \end{bmatrix} \quad \mathbf{E}=\begin{bmatrix} 0 & E \\ -E & 0 \end{bmatrix} \quad \mathbf{F}=\begin{bmatrix} 0 & F \\ -F & 0 \end{bmatrix}, \tag{26}$$

and Eq. (24) yields the following system of three linearly independent equations for the three unknowns D, E, F,

$$\begin{aligned} &\hat{\alpha}_{11}D+\hat{\beta}_{11}E+\hat{\gamma}_{11}F=0,\\ &\hat{\alpha}_{12}D+\hat{\beta}_{12}E+\hat{\gamma}_{12}F=0,\\ &\hat{\alpha}_{22}D+\hat{\beta}_{22}E+\hat{\gamma}_{22}F=0. \end{aligned} \tag{27}$$

This homogeneous linear algebraic system yields nontrivial solutions for D, E, and F, only when its determinant is zero, which, accounting for the fact that $\hat{\alpha}_{12}=\hat{\beta}_{22}=0$, is equivalent to $\hat{\beta}_{12}(\hat{\alpha}_{11}\hat{\gamma}_{22}-\hat{\alpha}_{22}\hat{\gamma}_{11})=-\hat{\alpha}_{22}\hat{\beta}_{11}\hat{\gamma}_{12}$, or equivalently, using the expressions Eqs. (21) and multiplying by $X^3(\eta-X)^3$,

$$\begin{aligned} &[\eta-(1+r_2)X]\{(\eta-X)[(\eta-X)(s_{22}X-1)+r_6^2X]\\ &\quad+X^2[(\eta-X)s_{66}+r_2^2]\}\\ &\quad=2r_6X^2(\eta-X)[(\eta-X)s_{26}-r_2r_6]. \end{aligned} \tag{28}$$

Hence *the secular equation is obtained explicitly as the quartic Eq. (28) in* $X=\rho v^2$, with coefficients expressed in terms of the elastic stiffnesses through Eqs. (16).

For consistency, the orthorhombic case, where $c_{16}=c_{26}=c_{45}=0$, is now considered. In this case, the coefficients Eq. (16) reduce to

$$\begin{aligned} &r_6=0, \quad r_2=\frac{c_{12}}{c_{22}}, \quad s_{22}=\frac{1}{c_{66}},\\ &s_{26}=0, \quad s_{66}=\frac{1}{c_{22}}, \quad \eta=c_{11}-\frac{c_{12}^2}{c_{22}}, \end{aligned} \tag{29}$$

and the right hand-side of Eq. (28) is zero, while the left hand-side yields the equation

$$\begin{aligned} &[\eta-(1+r_2)X]\{(\eta-X)^2(s_{22}X-1)\\ &\quad+X^2[(\eta-X)s_{66}+r_2^2]\}=0. \end{aligned} \tag{30}$$

The nullity of the first factor in this equation corresponds to $\hat{\beta}_{12}=0$. Because for the orthorhombic case, $\hat{\alpha}_{12}=\hat{\gamma}_{12}=\hat{\beta}_{11}=\hat{\beta}_{22}=0$ also, the equations of motion (19) then decouple into

$$\hat{\alpha}_{11}t_1''+\hat{\gamma}_{11}t_1=0, \quad \hat{\alpha}_{22}t_2''+\hat{\gamma}_{22}t_2=0, \tag{31}$$

whose solutions satisfying the boundary conditions Eqs. (22) are the trivial ones. The nullity of the second factor in Eq. (30) corresponds to the well-studied[6,7,17] secular equation for surface waves in orthorhombic crystals,

$$\begin{aligned} &\frac{c_{22}}{c_{11}}\left(\frac{c_{11}c_{22}-c_{12}^2}{c_{22}c_{66}}-\frac{\rho v^2}{c_{66}}\right)^2\left(1-\frac{\rho v^2}{c_{66}}\right)\\ &\quad-\left(\frac{\rho v^2}{c_{66}}\right)^2\left(1-\frac{\rho v^2}{c_{11}}\right)=0. \end{aligned} \tag{32}$$

Finally, concrete examples are given (see Table I). In each considered case, the secular equation (28) has either two or four positive real roots, out of which only one corresponds to a subsonic wave. The elimination of the other roots

is made by comparison with the speed of a homogeneous body wave propagating in the direction of the x_1 material axis. For this body wave, the functions $U_i(x_2)$, $t_i(x_2)$, $(i=1,2)$, are constant, and the equations of motion imply that the determinant of the 4×4 matrix in Eq. (14) is zero, condition from which the body wave speed can be found. Also, it is checked a posteriori that the value $X=\eta$ corresponds to the supersonic range, and so that the matrix $N_3+X\mathbf{1}$ is indeed invertible within the subsonic range. For instance, for tin fluoride, η is of the order of 3×10^7, the secular equation (28) has the roots 1339, 2350, 2513, and 3403, and the slowest body wave in the x_1 direction travels at 1504 m s^{-1}; hence a subsonic surface wave travels in tin fluoride at 1339 m s^{-1}.

Barnett, Chadwick, and Lothe,[18] and Chadwick and Willson[9] considered surface waves propagating in monoclinic materials, and computed the surface wave speed v in two steps, first by solving numerically a bicubic, then by substituting the result into another equation of which v is the only zero. These authors studied surface wave propagation for every value of the angle α between the reference plane and the plane of material symmetry. Numerical values for v are only given in the cases of aegirite-augite, diallage, gypsum, and microcline, and at $\alpha=0$, these results are in agreement with those presented in Table I. Sources of experimental data and extensive discussions on limiting speeds, existence of secluded supersonic surface waves, rotation of the reference plane with respect to the plane of material symmetry, etc., can be found in these articles and in references therein.

VII. CONCLUDING REMARKS

Surface wave motion in monoclinic crystals with plane of symmetry at $x_3=0$ turned out to correspond to plane strain and plane stress motion (Sec. IV). Thanks to this, the equations of of motion yielded a system of only two differential equations for the tractions (Sec. V). Once the method of first integrals was applied, a homogeneous system of three linearly independent equations for three unknowns was obtained (Sec. VI). Had the motion not corresponded to plane stress, then the same procedure would have given a system of 18 equations for 18 unknowns, when the equations of motion are written for the displacement components,[11] or a system of 9 equations for 9 unknowns, when the equations of motion are written for the traction components as in the present paper. However, these equations are not linearly independent, and the secular equation cannot be obtained in this manner. Hence, it ought to be stressed again that the method presented in the paper is not a general method for a surface wave traveling in arbitrary direction in an anisotropic crystal, but was limited to the study of a surface wave propagating in the x_1-direction of a monoclinic crystal with plane of symmetry at $x_3=0$, with attenuation in the x_2-direction.

Nevertheless, some plane strain problems remain open and it is hoped that the method exposed in this paper might help solve them analytically. Also, beyond mathematical satisfaction, the derivation of an explicit secular equation provides a basis for a possible nonlinear perturbative analysis.

ACKNOWLEDGMENTS

I thank Professor M. Hayes (University College, Dublin) for his decisive input in providing Ref. 11, Dr. A. Ibragimov (Texas A&M University, College Station) for translating the title of Ref. 7, and Dr. M. Ziane (Texas A&M University, College Station) for stimulating discussions. I also thank the referees for pointing out to Ref. 10 and to misprints and possible improvements on an earlier version of the paper.

[1] A. N. Stroh, ''Dislocations and cracks in anisotropic elasticity,'' Philos. Mag. **3**, 625–646 (1958).
[2] A. N. Stroh, ''Steady state problems in anisotropic elasticity,'' J. Math. Phys. **41**, 77–103 (1962).
[3] P. K. Currie, ''Rayleigh waves on elastic crystals,'' Q. J. Mech. Appl. Math. **27**, 489–496 (1974).
[4] D. M. Barnett and J. Lothe, ''Consideration of the existence of surface wave (Rayleigh wave) solutions in anisotropic elastic crystals,'' J. Phys. F: Met. Phys. **4**, 671–686 (1974).
[5] P. Chadwick and G. D. Smith, ''Foundations of the theory of surface waves in anisotropic elastic solids,'' Adv. Appl. Mech. **17**, 303–376 (1977).
[6] T. C. T. Ting, *Anisotropic Elasticity: Theory and Applications* (Oxford University Press, New York, 1996).
[7] V. A. Sveklo, ''Plane waves and Rayleigh waves in anisotropic media,'' (in Russian) Dokl. Akad. Nauk SSSR **59**, 871–874 (1948).
[8] D. Royer and E. Dieulesaint, ''Rayleigh wave velocity and displacement in orthorhombic, tetragonal, and cubic crystals,'' J. Acoust. Soc. Am. **76**, 1438–1444 (1984).
[9] P. Chadwick and N. J. Wilson, ''The behaviour of elastic surface waves polarized in a plane of material symmetry, II. Monoclinic media,'' Proc. R. Soc. London, Ser. A **438**, 207–223 (1992).
[10] A. K. Head, ''The Galois unsolvability of the sextic equation of anisotropic elasticity,'' J. Elast. **9**, 9–20 (1979).
[11] V. G. Mozhaev, ''Some new ideas in the theory of surface acoustic waves in anisotropic media,'' in *IUTAM Symposium on Anisotropy, Inhomogeneity and Nonlinearity in Solids*, edited by D. F. Parker and A. H. England (Kluwer, Holland, 1994), pp. 455–462.
[12] M. Y. Yu, ''Surface polaritons in nonlinear media,'' Phys. Rev. A **28**, 1855–1856 (1987).
[13] S. Nair and D. A. Sotiropoulos, ''Elastic waves in orthotropic incompressible materials and reflection from an interface,'' J. Acoust. Soc. Am. **102**, 102–109 (1997).
[14] D. A. Sotiropoulos and S. Nair, ''Elastic waves in monoclinic incompressible materials and reflection from an interface,'' J. Acoust. Soc. Am. **105**, 2981–2983 (1999).
[15] A. E. H. Love, *A Treatise on the Mathematical Theory of Elasticity* (Cambridge University Press, England, 1927).
[16] K. A. Ingebrigsten and A. Tonning, ''Elastic surface waves in crystal,'' Phys. Rev. **184**, 942–951 (1969).
[17] R. Stoneley, ''The propagation of surface waves in an elastic medium with orthorhombic symmetry,'' Geophys. J. **8**, 176–186 (1963).
[18] D. M. Barnett, P. Chadwick, and J. Lothe, ''The behaviour of elastic surface waves polarized in a plane of material symmetry. I. Appendum,'' Proc. R. Soc. London, Ser. A **433**, 699–710 (1991).

Finite-element modeling of lead magnesium niobate electrostrictive materials: Dynamic analysis

Jocelyne Coutte, Jean-Claude Debus, Bertrand Dubus, and Régis Bossut
Institut d'Electronique et de Microélectronique du Nord, U.M.R. C.N.R.S. 8520, Département I.S.E.N., 41 Boulevard Vauban, 59046 Lille Cedex, France

(Received 30 March 2000; accepted for publication 27 December 2000)

A finite-element model is proposed for the time-domain analysis of electrostrictive materials. Hom's material model, developed for lead magnesium niobate (PMN) ceramics, is used. It includes the quadratic dependence of strain with polarization, the saturation of polarization, assumes constant temperature, and excludes hysteresis. The theoretical formulation is justified by the principle of virtual works. The numerical model is obtained after discretization in space and time. The validation is performed by comparing numerical results with semianalytical results for an electrostrictive spherical shell subjected to a step in voltage or in charge. From these results, a method to compute the coupling coefficient of electrostrictive materials, based on Ikeda's definition, is proposed and applied to a bar with parallel electric field. © *2001 Acoustical Society of America.*
[DOI: 10.1121/1.1350452]

PACS numbers: 43.38.Ar, 43.38.Fx, 43.30.Yj [SLE]

I. INTRODUCTION

Since 1960, piezoelectric lead titanate zirconate (PZT) ceramics have been used in most electroacoustical transducers.[1] The current need for high power in sonar applications leads to the emergence of new transduction materials relying upon the magnetostrictive[2] or the electrostrictive effects.[3] In electrostrictive materials, the strain is proportional to the square of the polarization.[4,5] Compared to PZT ceramics, lead magnesium niobate ceramics ($Pb(Mg_{1/3}Nb_{2/3})O_3$) (PMN) exhibit two interesting properties:[6–9]

(1) the hysteresis in the electric field-polarization curve is reduced if the temperature is correctly chosen; and
(2) high strains ($\Delta l/l = 0.1\%$) can be obtained under moderate electric fields.

Handling the material nonlinearities, the temperature dependence and the application of the electrical polarization remain a challenge, both at the designing and technological levels, which limits the use of PMN in acoustic projectors. For transducer design, several works have provided semianalytical solutions: Sherman *et al.*[10] have developed a general approach to solve approximately nonlinear transducer equations by perturbation analysis. They consider six different types of transducers including the electrostrictive transducer and focus on the evaluation of the harmonic distortion. Piquette *et al.*[11] have derived physical equations for two geometries of electrostrictors (length expander bar with parallel field and thickness vibrator), to get a linearized equivalent circuit accounting for the effect of prestress and polarization. Their prediction includes coupling coefficient and harmonic distortion. To avoid the geometrical limitation of analytical models and describe an electrostrictive flextensional transducer, Hom *et al.*[12] have associated a one-dimensional description of the electrostrictive bar with an equivalent circuit representation of the shell extracted from finite-element modeling. Their results also include harmonic distortion prediction. Progress has also been made in numerical models. Debus *et al.*[13–15] have derived a general finite-element formulation for electrostrictive materials using the weighted-residual method and have applied it to the static analysis of a PMN bar. The resulting nonlinear equation was solved by direct method, implying an iterative scheme for the linear system resolution and an evaluation of electric displacement **D** at Gauss points. Ghandi *et al.*[16] have proposed an hybrid finite-element model by adding **D** as an internal variable. This method simplifies the introduction of material models where **D** is usually the independent variable. It also avoids the evaluation of **D** at Gauss points after resolution.

This paper presents the extension of the aforementioned finite-element formulation[15] to dynamic analysis. The constitutive laws used in the model are briefly summarized in Sec. II. Section III presents the theoretical formulation of the time-domain analysis of electrostrictive materials and its numerical implementation. Validation of the model on a PMN spherical shell is given in Sec. IV. Finally, the model is applied to the analysis of the coupling coefficient of a PMN bar with parallel electric field in Sec. V.

II. CONSTITUTIVE LAWS FOR ELECTROSTRICTORS

Hom's model[9,17] has been retained to describe the mechanical and dielectric behavior of electrostrictive ceramics. It includes the quadratic dependence of strain with electrical displacement at zero stress, assumes constant temperature, and excludes hysteresis.

Assuming stress and electric field as independent state variables, the constitutive equations of electrostriction are written in a condensed notation[15]

$$\begin{aligned} \mathbf{S} &= [s^D]\mathbf{T} + [g([Q],\mathbf{D})]^t\mathbf{D}, \\ \mathbf{E} &= -2[g([Q],\mathbf{D})]\mathbf{T} + [\beta^T(P_s,k,\mathbf{D})]\mathbf{D}, \end{aligned} \tag{1}$$

where **D** is the electric displacement vector, **E** is the electric field vector, **T** is the condensed stress tensor, and **S** is the condensed strain tensor. The superscript ^{t} denotes matrix transposition. $[s^D]$ is the elastic compliance tensor at constant excitation. The equivalent ''piezoelectric'' strain constant tensor $[g]$, is defined as

$$[g([Q],\mathbf{D})]=\begin{bmatrix} Q_{11}D_1 & Q_{12}D_1 & Q_{12}D_1 & 0 & (Q_{11}-Q_{12})D_3 & (Q_{11}-Q_{12})D_2 \\ Q_{12}D_2 & Q_{11}D_2 & Q_{12}D_2 & (Q_{11}-Q_{12})D_3 & 0 & (Q_{11}-Q_{12})D_1 \\ Q_{12}D_3 & Q_{12}D_3 & Q_{11}D_3 & (Q_{11}-Q_{12})D_2 & (Q_{11}-Q_{12})D_1 & 0 \end{bmatrix}, \tag{2}$$

where Q_{ij} are the electrostrictive constants of the condensed tensor. Defining a new material constant k as

$$k=\frac{1}{P_s(\beta_{33}^T)_{\mathbf{D}=0}}, \tag{3}$$

where P_s is the spontaneous polarization, the dielectric impermeability constants are written

$$\beta_{ij}^T(P_s,k,\mathbf{D})=\frac{\delta_{ij}}{k|\mathbf{D}|}\operatorname{arctanh}\left(\frac{|\mathbf{D}|}{\mathbf{P}_s}\right), \tag{4}$$

where δ_{ij} is equal to 1 if $i=j$ and to 0 if $i\neq j$. Other constitutive laws have been proposed[4,11,18] for electrostrictive ceramics, but will not be considered here. However, the formulation given in the next section is valid for any constitutive law and the numerical model can be easily modified.

III. FINITE-ELEMENT MODEL FOR DYNAMIC ANALYSIS OF ELECTROSTRICTIVE MATERIALS

A. Basic equations

Let us consider an electrostrictive domain Ω limited by a closed surface Γ pointing outwards. Newton's law for an elementary volume is written as

$$\frac{\partial T_{ij}}{\partial x_j}=\rho\frac{\partial^2 u_i}{\partial t^2}, \tag{5}$$

where T is the (noncondensed) second-rank stress tensor, **u** is the particle displacement, ρ the density of the electrostrictive material, **x** the spatial coordinate, and t the time. Gauss's theorem is written as

$$\frac{\partial D_i}{\partial x_i}=0. \tag{6}$$

Small strains and electrostatic theory are considered. Therefore,

$$S_{ij}=\frac{1}{2}\left(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}\right), \tag{7}$$

$$E_i=-\frac{\partial \varphi}{\partial x_i}, \tag{8}$$

where **S** is the (noncondensed) strain tensor, **E** the electric field, and φ the electrostatic scalar potential. For an electrostrictive material, the stress and the electrical displacement are obtained by inverting the set of equations (1)[15]

$$\begin{aligned} \mathbf{T}&=[c^E(k,P_s,[Q],\mathbf{D})]\mathbf{S}-[e(k,P_s,[Q],\mathbf{D})]^t\mathbf{E} \\ \mathbf{D}&=2[e(k,P_s,[Q],\mathbf{D})]\mathbf{S}+[\varepsilon^S(k,P_s,[Q],\mathbf{D})]\mathbf{E}, \end{aligned} \tag{9}$$

with

$$[\varepsilon^S(k,P_s,[Q],\mathbf{D})]=([\beta^T(k,P_s,\mathbf{D})]+2[g([Q],\mathbf{D})]\times[s^D]^{-1}[g([Q],\mathbf{D})]^t)^{-1}, \tag{10}$$

$$[e(k,P_s,[Q],\mathbf{D})]=[\varepsilon^s(k,P_s,[Q],\mathbf{D})][g([Q],\mathbf{D})]\times[s^D]^{-1}, \tag{11}$$

$$[c^E(k,P_s,[Q],\mathbf{D})]=[s^D]^{-1}-2[s^D]^{-1}[g([Q],\mathbf{D})]^t\times[\varepsilon^S(k,P_s,[Q],\mathbf{D})][g([Q],\mathbf{D})]\times[s^D]^{-1}. \tag{12}$$

It can be noted that Eqs. (10)–(12) are similar to usual relations between tensors in piezoelectricity.

B. Principle of virtual work

An arbitrary elementary generalized displacement vector $\delta\mathbf{w}$ is defined as

$$\delta\mathbf{w}=\begin{Bmatrix}\delta\mathbf{u}\\ \delta\varphi\end{Bmatrix}, \tag{13}$$

where $\delta\mathbf{u}$ and $\delta\varphi$ are arbitrary variations of the displacement field and of the electrical potential, respectively. Among all the admissible generalized displacements verifying $\delta\mathbf{w}=0$ at $t=t_1$ and $t=t_2$, the solution satisfies the equation[19]

$$\int_{t_1}^{t_2}\left\{\int\int\int_\Omega\left[\rho\left(\frac{\partial u_i}{\partial t}\right)\delta\left(\frac{\partial u_i}{\partial t}\right)-(T_{ij}\delta S_{ij}-D_i\delta E_i)\right]d\Omega+\int\int_\Gamma(\delta u_i T_{ij}n_j+\delta\varphi D_i n_i)d\Gamma\right\}dt=0, \tag{14}$$

where **n** is the unitary vector normal to Γ pointing outwards. The first integral corresponds to the variation of the Lagrangian function of the considered problem defined as the kinetic energy density minus the electric enthalpy. The surface integral contains two terms relative to the energy exchange through Γ. The first term corresponds to the work of applied forces and the second term to the energy exchanged with the external electrical generators. After integrating by parts the time integral on the kinetic energy and considering that $\delta\mathbf{u}=0$ at $t=t_1$ and $t=t_2$, Eq. (14) is written as

$$\int_{t_1}^{t_2}\left\{\int\int\int_\Omega\delta u_i\left(\rho\frac{\partial^2 u_i}{\partial t^2}\right)d\Omega+\int\int\int_\Omega(T_{ij}\delta S_{ij}-D_i\delta E_i)\times d\Omega-\int\int_\Gamma(\delta u_i T_{ij}n_j+\delta\varphi D_i n_i)d\Gamma\right\}dt=0. \tag{15}$$

As the choice of t_1 and t_2 is arbitrary, Eq. (15) is expressed

$$\int\int\int_{\Omega}\delta u_i\left(\rho\frac{\partial^2 u_i}{\partial t^2}\right)d\Omega+\int\int\int_{\Omega}(T_{ij}\delta S_{ij}-D_i\delta E_i)d\Omega$$

$$-\int\int_{\Gamma}(\delta u_i T_{ij}n_j+\delta\varphi D_i n_i)d\Gamma=0, \tag{16}$$

or in a condensed notation

$$\int\int\int_{\Omega}\delta\mathbf{w}^t[\rho^*]\frac{\partial^2\mathbf{w}}{\partial t^2}d\Omega+\int\int\int_{\Omega}\delta\boldsymbol{\Sigma}^t\boldsymbol{\Theta}d\Omega$$

$$-\int\int_{\Gamma}\delta\mathbf{w}^t\boldsymbol{\theta}d\Gamma=0. \tag{17}$$

$\boldsymbol{\Sigma}$ is the generalized strain vector, written as

$$\boldsymbol{\Sigma}=\begin{Bmatrix}\mathbf{S}\\-\mathbf{E}\end{Bmatrix}, \tag{18}$$

and $\boldsymbol{\theta}$ is the generalized force vector acting per unit area of the surface Γ

$$\boldsymbol{\theta}=\begin{Bmatrix}\mathbf{f}\\-q\end{Bmatrix}, \tag{19}$$

where $\mathbf{f}$ is the surface force acting on Γ

$$f_i=T_{ij}n_j, \tag{20}$$

and q is the charge density on Γ

$$q=-D_i n_i. \tag{21}$$

The density matrix $[\rho^*]$ is defined as

$$[\rho^*]=\begin{bmatrix}\rho&0\\0&0\end{bmatrix}. \tag{22}$$

C. Method for space discretization

The solution of Eq. (17) is based on the finite-element method and therefore requires a space–time discretization. The space domain Ω (and the surface Γ) are divided into subdomains called elements interconnected at some points called nodes. As the space boundary is not changing with time, the combined space–time problem is prismatic.[20] Therefore, inside element e, the physical variables (displacement $\mathbf{u}$, electrical potential φ) are partially discretized

$$\mathbf{u}^{(e)}=[N_u^{(e)}]\mathbf{U}^{(e)}, \tag{23}$$

$$\varphi^{(e)}=[N_\varphi^{(e)}]\boldsymbol{\Phi}^{(e)}. \tag{24}$$

$[N_u^{(e)}]=[N_u^{(e)}(\mathbf{x})]$ and $[N_\varphi^{(e)}]=[N_\varphi^{(e)}(\mathbf{x})]$ are spatial shape functions and $\mathbf{x}$ denotes the space coordinates. $\mathbf{U}^{(e)}=\mathbf{U}^{(e)}(t)$ and $\boldsymbol{\Phi}^{(e)}=\boldsymbol{\Phi}^{(e)}(t)$ are the vectors of the nodal values of the displacement and electric potential, respectively. The strain and the electrical field in each element are expressed as

$$\mathbf{S}^{(e)}=[B_u^{(e)}]\mathbf{U}^{(e)}, \tag{25}$$

$$\mathbf{E}^{(e)}=-[B_\varphi^{(e)}]\boldsymbol{\Phi}^{(e)}, \tag{26}$$

where $[B_u^{(e)}]=[B_u^{(e)}(\mathbf{x})]$ and $[B_\varphi^{(e)}]=[B_\varphi^{(e)}(\mathbf{x})]$ are matrices obtained from derivatives of the shape functions with respect to space coordinates. Inside element (e), the generalized stress vector is expressed using Eq. (9) as

$$\boldsymbol{\Theta}^{(e)}=\begin{Bmatrix}\mathbf{T}^{(e)}\\\mathbf{D}^{(e)}\end{Bmatrix}=\begin{bmatrix}[c^{E(e)}]&-[e^{(e)}]^t\\[2e^{(e)}]&[\epsilon^{S(e)}]\end{bmatrix}\times\begin{bmatrix}[B_u^{(e)}]&0\\0&-[B_\varphi^{(e)}]\end{bmatrix}\begin{Bmatrix}\mathbf{U}^{(e)}\\\boldsymbol{\Phi}^{(e)}\end{Bmatrix}. \tag{27}$$

After applying the finite-element discretization, Eq. (17) becomes

$$\sum_{e=1}^{m}\int\int\int_{\Omega^{(e)}}\delta\mathbf{w}^{(e)t}[\rho^*]\frac{\partial^2\mathbf{w}^{(e)}}{\partial t^2}d\Omega^{(e)}+\sum_{e=1}^{m}\int\int\int_{\Omega^{(e)}}\delta\boldsymbol{\Sigma}^{(e)t}\boldsymbol{\Theta}^{(e)}d\Omega^{(e)}-\sum_{e=1}^{n}\int\int_{\Gamma^{(e)}}\delta\mathbf{w}^{(e)t}\boldsymbol{\theta}^{(e)}d\Gamma^{(e)}=0, \tag{28}$$

where m is the number of elements in the domain Ω and n is the number of elements on the boundary Γ. Combining Eqs. (23) to (27) with Eq. (28), we obtain

$$\{\delta\mathbf{U}^t,\ \delta\boldsymbol{\Phi}^t\}\begin{bmatrix}[M]&[0]\\[0]&[0]\end{bmatrix}\begin{Bmatrix}\ddot{\mathbf{U}}\\\ddot{\boldsymbol{\Phi}}\end{Bmatrix}+\{\delta\mathbf{U}^t,\ \delta\boldsymbol{\Phi}^t\}\times\left(\begin{bmatrix}[K_{uu}]&[K_{u\varphi}]\\2[K_{u\varphi}]^t&[K_{\varphi\varphi}]\end{bmatrix}\begin{Bmatrix}\mathbf{U}\\\boldsymbol{\Phi}\end{Bmatrix}-\begin{Bmatrix}\mathbf{F}\\-\mathbf{Q}\end{Bmatrix}\right)=0. \tag{29}$$

$\mathbf{U}$ and $\boldsymbol{\Phi}$ are the vectors of the nodal values of the displacement and the electric potential, respectively. $\ddot{\mathbf{U}}$ (respectively, $\ddot{\boldsymbol{\Phi}}$) is the vector of the second time derivative of $\mathbf{U}$ (respectively, $\boldsymbol{\Phi}$). The elastic stiffness matrix $[K_{uu}]$, the electromechanical coupling matrix $[K_{u\varphi}]$, the dielectric stiffness matrix $[K_{\varphi\varphi}]$, and the mass matrix $[M]$ are defined as

$$[K_{uu}]=\sum_{e=1}^{m}\int\int\int_{\Omega^{(e)}}[B_u^{(e)}]^t[c^{E(e)}][B_u^{(e)}]d\Omega^{(e)}, \tag{30}$$

$$[K_{u\varphi}]=\sum_{e=1}^{m}\int\int\int_{\Omega^{(e)}}[B_u^{(e)}]^t[e^{(e)}][B_\varphi^{(e)}]d\Omega^{(e)}, \tag{31}$$

$$[K_{\varphi\varphi}]=-\sum_{e=1}^{m}\int\int\int_{\Omega^{(e)}}[B_\varphi^{(e)}]^t[\epsilon^{S(e)}][B_\varphi^{(e)}]d\Omega^{(e)}, \tag{32}$$

$$[M]=\rho\sum_{e=1}^{m}\int\int\int_{\Omega^{(e)}}[N_u^{(e)}]^t[N_u^{(e)}]d\Omega^{(e)}. \tag{33}$$

We define the vectors of the nodal values of the external force $\mathbf{F}$ and the electric charge $\mathbf{Q}$ as

$$\mathbf{F}=\sum_{e=1}^{n}\mathbf{F}^{(e)}=\sum_{e=1}^{n}\int\int_{\Gamma^{(e)}}[N_u^{(e)}]^t\mathbf{f}\,d\Gamma^{(e)}, \tag{34}$$

$$\mathbf{Q}=\sum_{e=1}^{n}\mathbf{Q}^{(e)}=-\sum_{e=1}^{n}\int\int_{\Gamma^{(e)}}[N_\varphi^{(e)}]q\,d\Gamma^{(e)}. \tag{35}$$

As Eq. (29) is valid for any value of the arbitrary displacement, we obtain

$$\begin{bmatrix} [M] & [0] \\ [0] & [0] \end{bmatrix} \begin{Bmatrix} \ddot{\mathbf{U}} \\ \ddot{\mathbf{\Phi}} \end{Bmatrix} + \begin{bmatrix} [K_{uu}] & [K_{u\varphi}] \\ 2[K_{u\varphi}]^t & [K_{\varphi\varphi}] \end{bmatrix} \begin{Bmatrix} \mathbf{U} \\ \mathbf{\Phi} \end{Bmatrix} = \begin{Bmatrix} \mathbf{F} \\ -\mathbf{Q} \end{Bmatrix}. \tag{36}$$

Equation (36) constitutes the finite-element set of nonlinear equations of an electrostrictive material for dynamic analysis. The physical tensors appearing in Eqs. (30)–(33) are functions of the electrical displacement [Eq. (9)] which is expressed in each element using $\mathbf{U}^{(e)}$ and $\mathbf{Q}^{(e)}$ [second line of Eq. (27)]. Equation (36) is modified by a condensation of the electrical potential nodal values to obtain an invertible ''mass matrix''

$$[M]\ddot{\mathbf{U}} + \{[K_{uu}] - 2[K_{u\varphi}][K_{\varphi\varphi}]^{-1}[K_{u\varphi}]^t\}\mathbf{U} = \mathbf{F} + [K_{u\varphi}][K_{\varphi\varphi}]^{-1}\mathbf{Q}. \tag{37}$$

$\mathbf{Q}$ is then obtained from the second line of Eq. (36). In the following, Eq. (37) is more generally written

$$[M]\ddot{\mathbf{U}} + [K]\mathbf{U} = \overline{\mathbf{F}}. \tag{38}$$

D. Method for time-domain integration and numerical implementation

The method for time discretization is the explicit central difference scheme which can be found in textbooks.[20] Details specific to electrostrictive materials analysis are described hereafter. For the whole time domain, discretized times $t_n = n\Delta t$ are defined where Δt is the time step. After time discretization, Eq. (38) and the second line of Eq. (36) are written

$$\frac{1}{\Delta t^2}[M]\mathbf{U}_{n+1} = \overline{\mathbf{F}}_n - [K_n]\mathbf{U}_n + \frac{1}{\Delta t^2}[M](2\mathbf{U}_n - \mathbf{U}_{n-1}), \tag{39}$$

$$\mathbf{\Phi}_n = -2[K_{\varphi\varphi_n}]^{-1}[K_{u\varphi_n}]^t\mathbf{U}_n - [K_{\varphi\varphi_n}]^{-1}\mathbf{Q}_n, \tag{40}$$

where the subscript n denotes the value at time t_n.

Two types of elements are developed in the ATILA[21] code to include the nonlinear time domain analysis: an eight-node quadrilateral and a six-node triangle for axisymmetrical analysis. These elements are isoparametric. The following procedure is used:

(1) Initialization of $\mathbf{U}_0, \mathbf{U}_{-1}$ (or $\dot{\mathbf{U}}_0$), and $\mathbf{D}_0$;
(2) Computation of [**M**]; and
(3) For each time step:

(i) Computation and assembling of the stiffness matrix $[K(\mathbf{D}_n)]$ and $\overline{\mathbf{F}}_n$;
(ii) Condensation of the electrical degrees of freedom;
(iii) Linear solving of Eq. (39) to get $\mathbf{U}_{n+1}$, evaluation of $\mathbf{\Phi}_{n+1}$ from Eq. (40); and
(iv) Computation of $\mathbf{D}_{n+1}$ at each Gauss point used in matrices integration as detailed in Ref. 15.

The initialization can be the solution of a previous static analysis[15] in order to include the effect of mechanical prestress and electrical dc polarization. For nonlinear problems, the proposed method reduces the memory required for the storage of the matrices and leads to a reasonable computation time, even if the number of time steps is large.[20] However, the stability is conditional and the time step Δt must be chosen small enough to avoid the possible divergences. In the following examples, we used $2.94\times10^{-2} < \omega_r\Delta t < 3.56\times10^{-2}$ for the sphere and $0.99\times10^{-2} < \omega_r\Delta t < 1.21\times10^{-2}$ for the bar, where ω_r is the fundamental frequency of vibration.

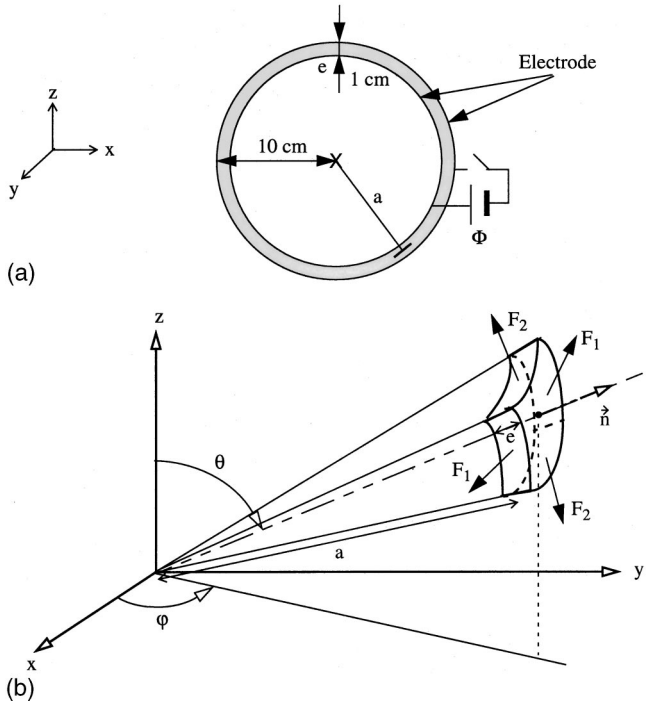


FIG. 1. (a) Geometry of the electrostrictive PMN spherical shell submitted to a voltage step. (b) Geometry of the elementary volume.

IV. VALIDATION: DYNAMIC ANALYSIS OF A PMN SPHERICAL SHELL

A. Description of the problem

The geometry of interest is an *in vacuo* PMN spherical shell radially polarized having an average radius of 100 mm and a thickness of 10 mm (Fig. 1). The dynamic excitation is provided electrically to the shell using electrodes located at inner and outer surfaces. Material constants are taken from Ref. 15 Sec. B 2. Two types of excitations are considered:

(1) a step in voltage where the voltage V applied to the electrodes is written
$$V(t) = V_{\text{init}} + V_{\text{step}}H(t); \quad \text{and} \tag{41}$$
(2) a step in charge where the charge on the electrodes is written
$$Q(t) = Q_{\text{init}} + Q_{\text{step}}H(t). \tag{42}$$

V_{init}, V_{step}, Q_{init}, and Q_{step} are the amplitudes of the initial voltage, the voltage step, the initial charge, and the charge step, respectively. H is the Heaviside step function. Initial velocity is set to zero. The lumped constant model (LC) used for validation is described in the Appendix.

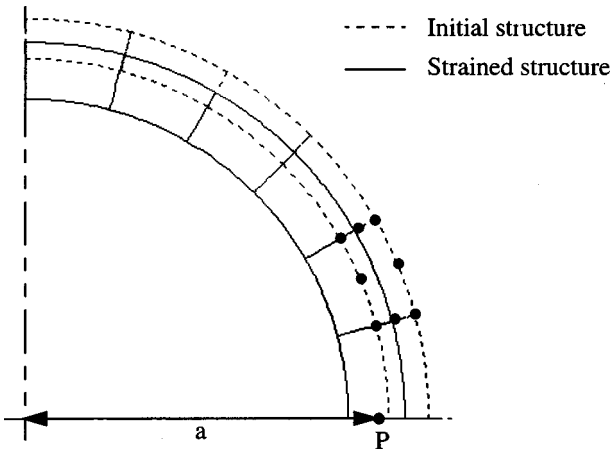


FIG. 2. PMN spherical shell submitted to an initial voltage and a voltage step. Full line: strained structure; dashed line: initial structure.

B. PMN spherical shell excited by a step in voltage

The response of the PMN spherical shell to a step in voltage is considered first. The finite-element mesh consists of six axisymmetrical electrostrictive (eight node quadrilateral) elements (Fig. 2). Figure 3 displays the displacement at point *P* (average radius) of the PMN shell versus time for the lumped constant model (LC), and finite-element computation including (FE2) or not including (FE1) saturation of polarization. The initial voltage is 2000 V and the voltage step is 2000 V. A radial vibration of the shell at constant *E* is observed (the corresponding natural frequency is denoted f_E). Displacement of sphere surface (not represented) displays an additional high-frequency vibration associated to wall thickness mode. The PMN shell is always in contraction. An excellent agreement is found between the FE1 and LC results. The average value of displacement (around 5.5 μm) is approximately 4 times larger than the initial displacement (around 1.4 μm). This result illustrates the quadratic dependence of strain with electric field when saturation is neglected. FE2 results display a distortion of the sinusoidal vibration at large displacements due to the polarization saturation.

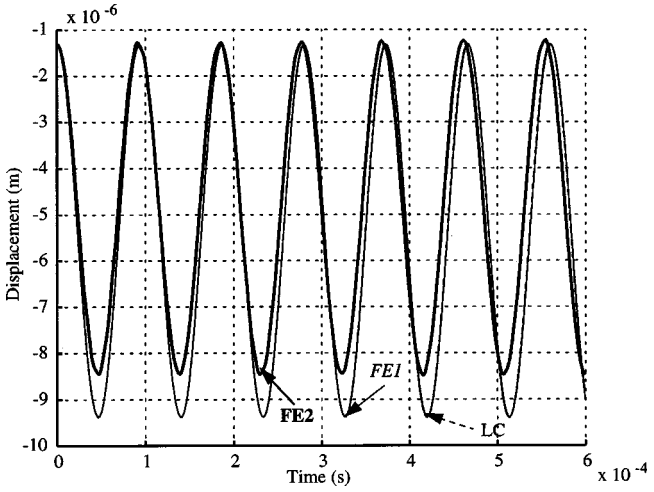


FIG. 3. Displacement at the point P of the PMN shell submitted to 2000 V voltage step (initial voltage of 2000 V) versus time. Thin line: finite element (model FE1); thick line: finite element (model FE2); dashed line: semianalytical model (LC).

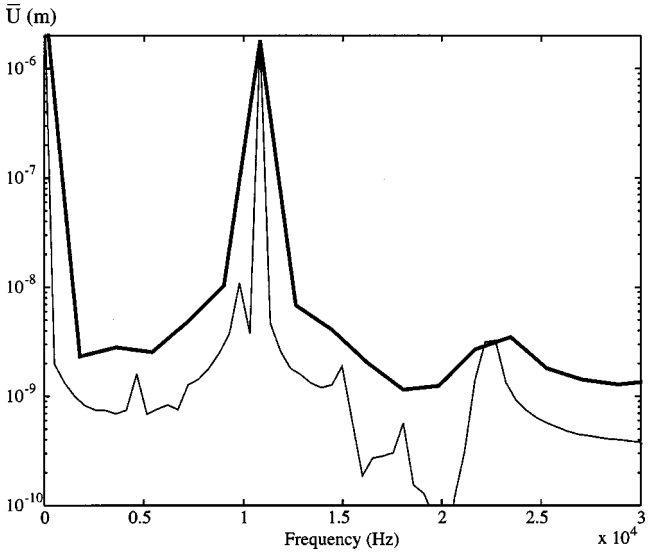


FIG. 4. DFT on the displacement of the PMN shell (model FE2). Time step is 5×10^{-7} s. Thick line: time window of 5.5×10^{-4} s. Thin line: time window of 1.65×10^{-3} s. $\bar{U}$ is the amplitude of displacement in the frequency domain.

A discrete Fourier transform (DFT) on displacement gives the natural frequency of vibration (Fig. 4). Frequencies computed for various initial voltages and various magnitudes of step are given in Table I. A decrease in the natural frequency with increasing step and initial voltage is observed. The sphere exhibits a behavior similar to Duffing's softening oscillator,[22] whose resonance frequency shifts down when the amplitude of vibration is increased. Increasing the time window of observation does not reveal significant energy transfer from fundamental mode toward other frequencies.

C. PMN spherical shell excited by a step in charge

The response of the PMN shell to a step in charge is analyzed for various initial charges and various charge steps. Figure 5 displays the displacement at point *P* of the shell versus time for LC and FE models for an initial charge and a charge step equivalent, respectively, to 2000 V. The natural radial vibration of the shell at constant *D* is observed (the

TABLE I. Natural frequency of vibration (in kHz) at constant *E* of the PMN spherical shell for various values of initial voltage and voltage step.

		Step (V)				
Initial (V)	Model	200	1000	2000	3000	4000
0	LC	11.31	11.26	11.14	10.96	10.71
	*FE*1	*11.31*	*11.26*	*11.14*	*10.96*	*10.73*
	FE2	**11.32**	**11.28**	**11.17**	**11.01**	**10.83**
2000	LC	11.14	10.96	10.71	10.42	10.08
	*FE*1	*11.14*	*10.96*	*10.73*	*10.44*	*10.08*
	FE2	**11.15**	**11.01**	**11.83**	**10.68**	**10.53**
4000	LC	10.66	10.44	10.08	9.73	9.35
	*FE*1	*10.68*	*10.44*	*10.08*	*9.73*	*9.35*
	FE2	**10.81**	**10.68**	**10.53**	**10.42**	**10.36**

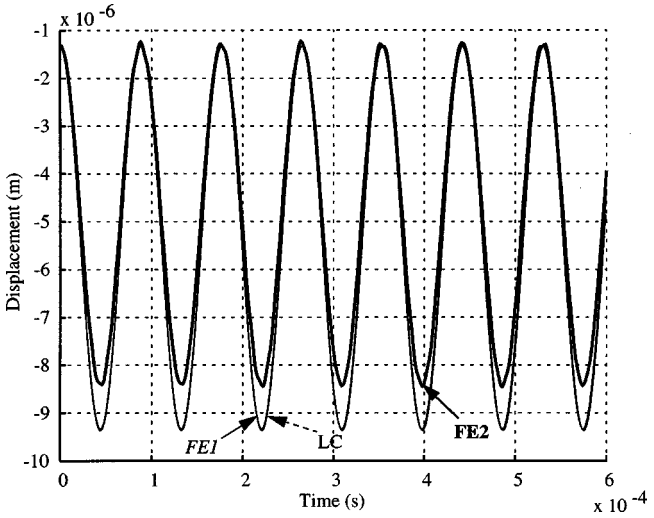


FIG. 5. Displacement at the point P of the PMN shell submitted to 5.47 mC charge step and an initial charge of 5.47 mC (equivalent to 2000 V) versus time. Thin line: finite element (model FE1); thick line: finite element (model FE2); dashed line: semianalytical model (LC).

corresponding natural frequency is denoted f_D). A good agreement is found between the different models. DFT on displacement gives the natural frequency of vibration directly. This frequency is identical for any initial value and any step and is equal to 11.31 kHz. These results can be analyzed using Eq. (A2). When D is prescribed to a constant value independent of time, the first line of Eq. (A2) turns into a linear stress–strain relationship independent of D around a static strain due to D. Therefore, the prescribed value of D affects the static value of the displacement and the amplitude of vibration [through the second line of Eq. (A2)] but it has no effect on the resonance frequency.

V. APPLICATION TO A PMN BAR

A. Description of the problem

The geometry of interest is an *in vacuo* PMN electrostrictive bar having a length of 10 mm and a cross area of 4 mm^2 (Fig. 6). Previous work done on this structure for static analysis include measurements made at NUWC New London (sample with square cross section)[23] and finite-element modeling performed at ISEN and NUWC/USRD Orlando (sample with circular crosssection).[13–15,24] The dynamic excitation is provided electrically to the bar using the electrodes located at both ends.

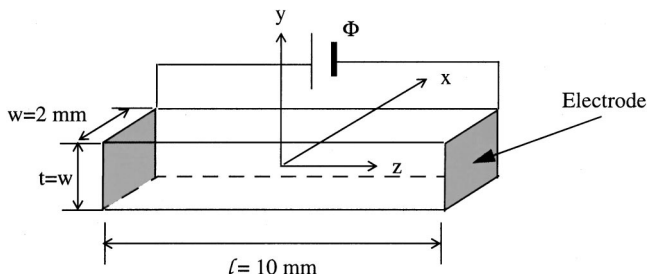


FIG. 6. Geometry of the electrostrictive PMN bar submitted to a voltage step.

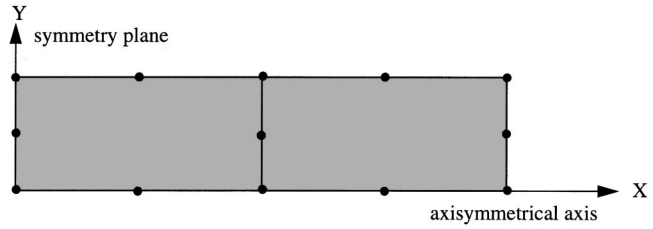


FIG. 7. Mesh of the PMN bar.

B. Results

The response of the PMN bar to a step in voltage is computed using Hom's model (FE2). The finite-element mesh consists of two axisymmetrical electrostrictive (eight node quadrilateral) elements (Fig. 7). Various initial voltages are considered. Figure 8 displays the displacement at the end of the bar versus time for the FE2 model for an initial voltage of 2000 V and a voltage step of 2000 V and for an initial charge and charge step equivalent to 2000 V. Natural longitudinal vibrations of the bar at constant E and constant D are obtained. The first mode, as well as higher-order modes, are seen in the figure. These high-frequency modes produce, at start, a faster response for the bar compared to the sphere. DFT on displacement gives the natural frequencies of vibration directly. Frequencies computed for values of initial voltage and magnitudes of steps are given in Table II. A decrease in the natural frequency with increasing step or initial voltage is observed. The frequency of natural vibration at constant D is found constant at 191.94 kHz.

C. About the coupling coefficient of electrostrictive materials

To evaluate the coupling coefficient of an electrostrictive bar, Hom *et al.*[17] have extended the method published in the IEEE standard on piezoelectricity.[25] Let us consider an idealized quasistatic electromechanical cycle where an initially mechanically and electrically free bar is connected to an external power source and deforms. The electrical energy

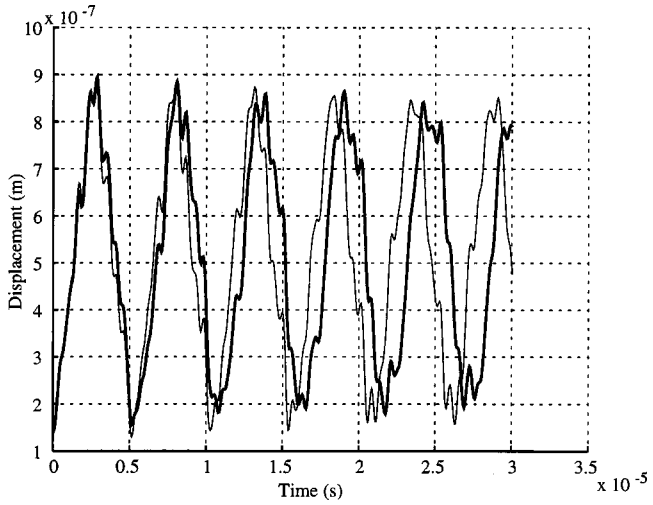


FIG. 8. Displacement at the end of the PMN bar versus time. FEM computation including polarization saturation. Thick line: initial voltage 2000 V, voltage step 2000 V. Thin line: initial charge 151 nC, charge step 151 nC.

TABLE II. Natural frequency of vibration (in kHz) at constant E of the PMN bar for various values of initial voltage and voltage step.

Initial (V)	Model	Step (V) 200	1000	2000	3000	4000
0	*FE1*	*191.57*	*190.84*	*189.03*	*185.53*	*181.16*
	FE2	**191.76**	**191.20**	**189.04**	**186.22**	**183.15**
2000	*FE1*	*188.32*	*184.84*	*181.16*	*175.75*	*169.78*
	FE2	**189.04**	**186.22**	**183.07**	**180.18**	**177.30**
4000	*FE1*	*181.16*	*175.75*	*169.78*	*163.67*	*156.99*
	FE2	**183.32**	**180.18**	**177.30**	**175.44**	**174.22**

stored in the bar at this stage is denoted W_1+W_2. Then, the strain is kept fixed and the electric field is removed. The remaining electrical energy is denoted W_1. The corresponding electromechanical coupling coefficient is

$$k_1^2=\frac{W_1}{W_1+W_2}. \tag{43}$$

The results of Sec. IV suggest another method for computing the coupling coefficient based on the variation of the natural frequency of vibration with the electrical boundary conditions. This definition can be seen as an extension of Ikeda's concept[26] developed in the field of piezoelectric delay lines

$$k_2^2=\frac{c^D-c^E}{c^D}=\frac{(f^D)^2-(f^E)^2}{(f^D)^2}, \tag{44}$$

where c^D (respectively, c^E) is the apparent elastic stiffness for the open-circuit (respectively, short-circuit) condition.

For the bar considered, k_1 is computed using Hom's method under a dc bias.[17] The following correspondence is used:

$$V_{\text{bias}}=V_{\text{init}}+\frac{V_{\text{step}}}{2}, \tag{45}$$

$$V_{\text{ac}}=\frac{V_{\text{step}}}{2}, \tag{46}$$

where V_{bias} and V_{ac} are the amplitudes of the bias voltage and the ac driving voltage, respectively. k_2 is evaluated from response to voltage and charge steps using Eq. (44). Results are displayed in Fig. 9 for various initial voltage and voltage steps. Both k_1 and k_2 increase with increasing initial voltage and step. They give similar results at high initial voltage but clearly differ at zero initial voltage. Saturation is observed around 45% at high electric field. The electrostrictive coupling coefficients obtained for PMN are always smaller than the usual piezoelectric coupling coefficient of PZT8 ceramics (k_{33} around 60%).

Computations are made at zero bias by choosing $V_{\text{step}}=-2\,V_{\text{init}}$ (or $Q_{\text{step}}=-2\,Q_{\text{init}}$). A flat response with displacement equal to its initial value is observed. When the voltage (respectively, charge) is very rapidly switched from V_{init} to $-V_{\text{init}}$ (respectively, Q_{init} to $-Q_{\text{init}}$) and no mechanical prestress is applied, no change is seen on the mechanical part because the electrostrictive effect is quadratic. Although Eq. (44) cannot be directly applied in this case, this result suggests a null coupling coefficient for a PMN-PT bar at zero bias consistent with Piquette's result[11] but different from Hom's result.[17]

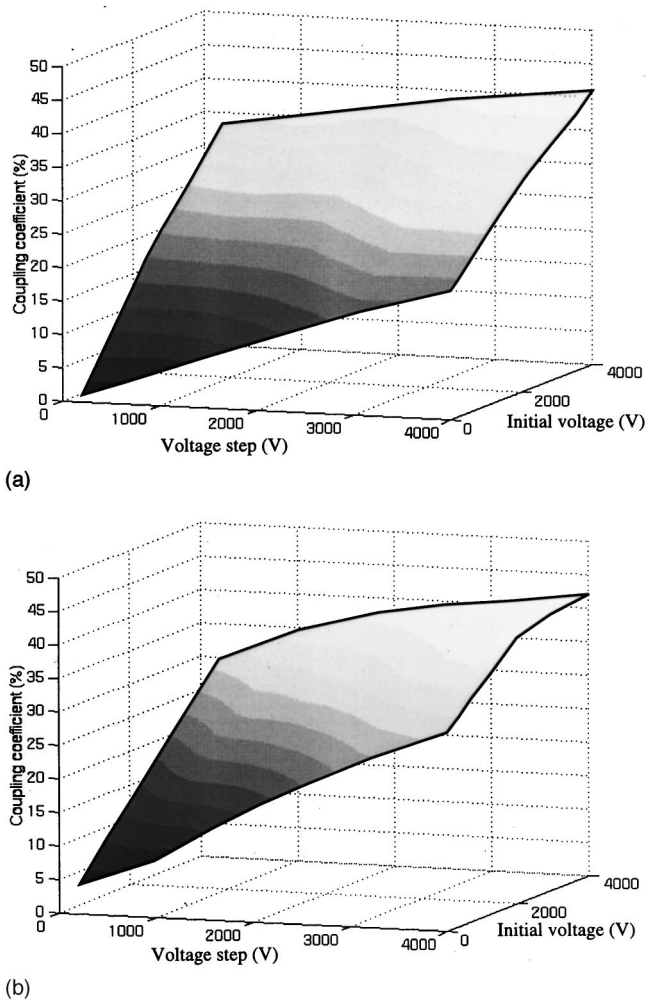


FIG. 9. Coupling coefficient of the PMN bar submitted to various initial voltages and voltage steps. (a) using Hom's method; (b) using numerical results and Ikeda's definition.

VI. CONCLUSION

The variational formulation established in this paper for the dynamic analysis of PMN electrostrictive ceramics is valid for any electromechanical transduction material as long as dissipative or hysteretic phenomena are not considered. It can be directly applied to other materials such as PZT ceramics in the nonlinear domain or electrostrictive polymers.

The finite-element implementation of this formulation leads to a nonlinear analysis in the time domain. Classical methods of analysis used to determine the characteristics of resonant transducers cannot be applied: eigenvalue analysis is impossible; time-domain response to an excitation at a prescribed frequency produces data which are difficult to use. The proposed method combines the response of the electrostrictive structure to voltage and charge steps and discrete Fourier transform. It provides an immediate insight on typical characteristics of resonant transducers such as resonance frequency and coupling coefficient interpreted according to Ikeda's definition.

ACKNOWLEDGMENTS

This work was supported by ONR under Contract No. N68171-95-M-6064 and by a DRET/DCN grant.

APPENDIX: LUMPED CONSTANT (LC) MODEL FOR DYNAMIC ANALYSIS OF THE PMN SPHERICAL SHELL

A lumped constants model is developed to describe the transient analysis of a PMN spherical shell with a radial electric field. To find a simple semianalytical solution, it is necessary to simplify Hom's model. The new set of equations assumes that the dielectric impermeability matrix $[\beta^T]$ used in Eq. (1) is constant, i.e., that the strain exhibits a quadratic dependence on the electric field. This assumption is verified in PMN ceramics when $E<E_0$ ($E_0\approx 0.6\,\mathrm{MV/m}$).[15] Assuming the usual hypotheses for thin-shell transducers[1] (thickness is small compared to diameter and wavelength; stress along the thickness is neglected), the constitutive equations are

$$\begin{aligned} S_1 &= s_{11}^D T_1 + s_{12}^D T_2 + Q_{12} D_3^2 \\ S_2 &= s_{12}^D T_1 + s_{11}^D T_2 + Q_{12} D_3^2 \\ E_3 &= -2 Q_{12} D_3 (T_1 + T_2) + \beta_{33}^T D_3, \end{aligned} \tag{A1}$$

where $S_1 = S_2 = u/a$ and $E_3 = \phi/e \cdot T_1$ and T_2 are, respectively, the meridional and hoop stresses.

The set of equations (A1) can be written

$$\begin{aligned} \frac{u}{a} &= s_c^D T + Q_{12} D_3^2 \\ \frac{\phi}{2} &= -2 Q_{12} D_3 T + \beta_{33}^T D_3, \end{aligned} \tag{A2}$$

where $s_c^D = (s_{11}^D + s_{12}^D)/2$ and $T = T_1 + T_2$.

Newton's law applied to an elementary volume of the shell described using spherical coordinates (Fig. 1) is written as

$$\rho\, a \frac{\partial^2 u}{\partial t^2} = -(T_1 + T_2) = -T, \tag{A3}$$

where a is the radius of the middle surface and ρ is the shell density. A finite-difference method is used for time discretization. Computation is performed at time $t_n = n\Delta t$ where Δt is the time step. A central difference method is used[20] to express velocity and acceleration at $t = t_n$

$${}^n\dot{u} = \frac{{}^{n+1}u - {}^{n-1}u}{\Delta t}, \tag{A4}$$

$${}^n\ddot{u} = \frac{{}^{n+1}u - 2\,{}^n u + {}^{n+1}u}{\Delta t^2}. \tag{A5}$$

Superscript n denotes the nth time step. After combining Eqs. (A2)–(A5), the discretized set of equations at time $t = t_n$ is

$${}^n D_3^3 + \frac{\left(s_c^D - 2 Q_{12} \dfrac{{}^n u}{a}\right)}{2 Q_{12}^2} {}^n D_3 + \frac{s_c^D \phi}{2 Q_{12}^2 e} = 0, \tag{A6}$$

$${}^{n+1}u = 2\,{}^n u - {}^{n-1}u + \frac{\Delta t^2}{\rho a s_c^D}\left(\frac{{}^n u}{a} - Q_{12}\,{}^n D_3^2\right). \tag{A7}$$

Initial conditions at $t = 0$ are obtained from static solution of Eqs. (A1) and (A3)

$${}^0 D_3 = \frac{{}^0\phi}{e \beta_{33}^T}, \tag{A8}$$

$${}^0 u = \frac{a Q_{12}}{e^2 \beta_{33}^{T^2}}\,{}^0\phi^2, \tag{A9}$$

and assuming that initial velocity is zero.

$${}^1 u = {}^{-1}u = {}^0 u + \frac{\Delta t^2}{2 \rho a s_c^D}\left(\frac{{}^0 u}{a} - Q_{12}\,{}^0 D_3^2\right). \tag{A10}$$

This semianalytical model only describes the first mode of the vibration, not the higher-order modes.

[1] O. B. Wilson, *Introduction to Theory and Design of Sonar Transducers* (Peninsula, Los Altos, CA, 1988).
[2] M. B. Moffett, J. M. Powers, and A. E. Clark, ''Comparison of Terfenol-D and PZT-4 power limitations,'' J. Acoust. Soc. Am. **90**, 1184–1185 (1991).
[3] K. M. Rittenmyer, ''Electrostrictive ceramics for underwater transducer application,'' J. Acoust. Soc. Am. **95**, 849–856 (1994).
[4] W. P. Mason, *Piezoelectric Crystals and Their Application to Ultrasonics* (Van Nostrand, Princeton, 1950), Chap. XII.
[5] A. F. Devonshire, ''Theory of ferroelectrics,'' Philos. Mag. **3**, 85–130 (1954).
[6] K. Uchino, ''Electrostrictive actuators: Materials and applications,'' Ceram. Bull. **65**, 647–652 (1986).
[7] D. Damajanovic and R. E. Newhnam, ''Electrostrictive and piezoelectric materials for actuator applications,'' J. Intell. Mater. Syst. Struct. **3**, 190–208 (1992).
[8] S. M. Pilgrim, M. Massuda, J. D. Prodey, J. M. Hock, and A. P. Ritter, ''Electrostrictive sonar drivers for flextensional transducers,'' in *Transducers for Sonics and Ultrasonics*, edited by M. D. McCollum, B. F. Hamonic, and O. B. Wilson (Technomic, Lancaster, 1993), pp. 95–102.
[9] C. L. Hom and N. Shankar, ''A fully coupled constitutive model for electrostrictive ceramic materials,'' J. Intell. Mater. Syst. Struct. **5**, 795–801 (1994).
[10] C. H. Sherman and J. L. Butler, ''Analysis of harmonic distorsion in electroacoustic transducers,'' J. Acoust. Soc. Am. **98**, 1596–1611 (1995).
[11] J. C. Piquette and S. E. Forsythe, ''Generalized material model for lead magnesium niobate (PMN),'' J. Acoust. Soc. Am. **104**, 2763–2772 (1998).
[12] C. L. Hom and N. Shankar, ''Modeling nonlinearity in electrostrictive sonar transducers,'' J. Acoust. Soc. Am. **104**, 1903–1913 (1998).
[13] J. C. Debus, B. Dubus, M. McCollum, and S. Black, ''Finite element modeling of PMN electrostrictive materials,'' Proceedings of the 3rd International Conference on Intell. Materials, edited by P. F. Gobin and J. Tatibouët, SPIE, pp. 913–918 (1996).
[14] J. C. Debus, B. Dubus, M. McCollum, and S. Black, ''2D modeling of PMN electrostrictive materials using the finite element code ATILA,'' *Proceedings of the 3rd European Conference on Underwater Acoustics*, edited by J. S. Papadakis (Crete University Press, Heraklion, 1996), Vol. II, pp. 1019–1024.
[15] J. C. Debus, B. Dubus, and J. Coutte, ''Finite element modeling of lead magnesium niobate electrostrictive materials: Static analysis,'' J. Acoust. Soc. Am. **103**, 3336–3343 (1998).
[16] K. Ghandi and N. W. Hagood, IV, ''Nonlinear modeling and characterization techniques for phase transitions in electromechanically coupled devices,'' PhD Thesis, Massachusetts Institute of Technology (1998).
[17] C. L. Hom, S. M. Pilgrim, N. Shankar, K. Bridger, M. Massuda, and S. R. Winzer, ''Calculation of quasi-static electromechanical coupling coefficients for electrostrictive ceramic materials,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **41**, 542–551 (1994).
[18] S. Sherrit, G. Catoiu, R. B. Stimpson, and B. K. Mukherjee, ''Modeling

and characterization of electrostrictive ceramics,'' S.P.I.E., Vol. 3324, 161–172 (1998).
[19] K. Washizu, *Variational Methods in Elasticity and Plasticity*, 2nd ed. (Pergamon, New York, 1975).
[20] O. C. Zienkiewicz and R. L. Taylor, *The Finite Element Method*, 4th ed. (McGraw-Hill, New York, 1989).
[21] ATILA Finite-Element code for Piezoelectric and Magnetostrictive Transducer Modeling Version 5.1.1, User's Manual, Institut Supérieur d'Electronique du Nord, Acoustics Laboratory (1997).
[22] L. S. Jacobsen and R. S. Ayre, *Engineering Vibrations* (McGraw-Hill, New York, 1958).
[23] E. A. McLaughlin, J. M. Powers, M. B. Moffett, and R. S. Janus, ''Characterization of PMN-PT-La for use in high-power electrostrictive projectors,'' J. Acoust. Soc. Am. **100**, 2729 (1996).
[24] J. C. Debus, S. Black, M. McCollum, and B. Dubus, ''2D modeling of electrostrictive materials using the finite element code ATILA,'' N68171-95-M-6064, NRL/USRD Report (1995).
[25] ''IEEE Standard on Piezoelectricity,'' IEEE Ultrason., Ferroelectr., and Freq. Contr. Soc., ANSI/IEEE Std 176–1987, 7 Sept. (1987).
[26] T. Ikeda, ''On the relations between electromechanical coupling coefficients and elastic constants in piezoelectric crystals,'' Jpn. J. Appl. Phys. **11**, 463 (1972).

New piezoelectric polymer for air-borne and water-borne sound transducers

Reiner Kressmann[a)]
Institute for Telecommunications, Darmstadt University of Technology, Merckstr. 25, D-64283 Darmstadt, Germany

(Received 13 September 2000; accepted for publication 18 January 2001)

Acoustic transducers made of a charged cellular polymer called EMFi have been designed and investigated with respect to air-borne and water-borne sound. The longitudinal transducer constant is around 90 pC/N, strongly exceeding the values of other piezoelectric polymers. This is mainly attributed to the very low Young's modulus of about 2 MPa. The acoustic impedance is only 2.6×10^4 kg/(m^2 s) and results in good matching to air but strong loading under water. Due to this strong loading, a pronounced reduction of resonance frequency from 300 kHz in air down to 17 kHz under water is observed. The experiments indicate that fluid loading is not only mass-like but also compliant, reducing the transducer's sensitivity below the resonance frequency of about -63 dB *re* 1 V/Pa (0.7 mV/Pa) in air to -71 dB *re* 1 V/Pa under water. This compliance is attributed to the medium's compressibility. Piezoelectricity of EMFi films is limited to temperatures below 70 °C; above, irreversible discharge of trapped charges takes place. Furthermore, a second type of EMFi, called ''OS'' was investigated, having a piezoelectric constant of 15 pC/N and a Young's modulus of 6 MPa. In quasi-static sensor measurements, the piezoelectric constant increases with the applied load. This nonlinearity explains the higher values reported in other publications on the same materials. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1354989]

PACS numbers: 43.38.Fx, 43.30.Yj, 77.84.Jd, 43.30.Jx [SLE]

I. INTRODUCTION

Polymers are well-known materials for acoustic transducers, both for air-borne and water-borne sound applications. The most important ones are Teflon due to its charge storing capabilities and piezoelectric polyvinylidene fluoride (PVDF). In the last few years, scientific interest on charge retention in cellular and porous polymers[1–8] and inorganic porous electrets[9] has grown rapidly. In this paper, the properties of transducers fabricated with two types of a novel polymer, the so-called electro-mechanical film (EMFi, provided by the Finnish company VTT), are presented and explained by a simple model. The two types are called ''HS'' (high sensitivity) and ''OS'' (ordinary sensitivity).

II. SAMPLES

The EMFi film is a foam based on polypropylene which is biaxially stretched during fabrication.[1] This results in the formation of lens-like air bubbles with diameters of about 10 to 70 μm and a thickness of about 5 μm resulting in a quite soft material. The manufacturer reports for the HS (high sensitivity) type of the material a thickness of 70 μm, a density of 330 kg/m^3, a Young's modulus of around 1 MPa, a relative permittivity of 1.2 and a sensitivity of 160 pC/N. The respective values for the OS type are thickness, 37 μm; density, 550 kg/m^3; Young's modulus, 9 MPa; relative permittivity, 1.6, and sensitivity, 30 pC/N.[1] The piezoelectricity is induced by means of corona charging resulting in charge storage inside the film.

a)Electronic mail: kre@nt.tu-darmstadt.de

EMFi films of several cm^2 size metallized on one side have been cut, glued on brass electrodes and mounted in aluminum housings to assure good shielding. Transducers with diameters of 20 and 35 mm active size were fabricated. For the underwater measurements, only carried out with the HS type, they have been covered with a thin waterproof film to protect them from contact with the water. A schematic of a transducer is given in Fig. 1.

III. EXPERIMENTAL WORK

Air-borne sound measurements have been carried out in an anechoic chamber under free-field conditions and a coupler. For the water-borne sound experiments an approximately cubic plastic tank of 1 m^3 size was used. The problems caused by its small opening of only 20 cm diameter are discussed together with the results in the following section.

Calibrated sensors were from Brüel and Kjær (quarter-inch microphone 4135, eighth-inch microphone 4138 and hydrophone 8103). Since the transducers' active capacitances are between 44 and 400 pF, a charge amplifier 2635 from Brüel and Kjær was used in the sensor experiments. Furthermore, an audio analyzer UPD from Rohde and Schwarz was used for signal generation and analysis.

A. Actuator response

The radiated sound pressure of a circular EMFi-HS actuator of 35 mm diameter driven sinusoidally with 100 V is shown in Fig. 2. The actuator was baffled slightly unsymmetrically with a wooden plate of 20×15 cm^2 size. The microphone was placed 60 cm apart from the sample on the main axis. The sound pressure increases with the square of frequency as expected below the resonance since in this

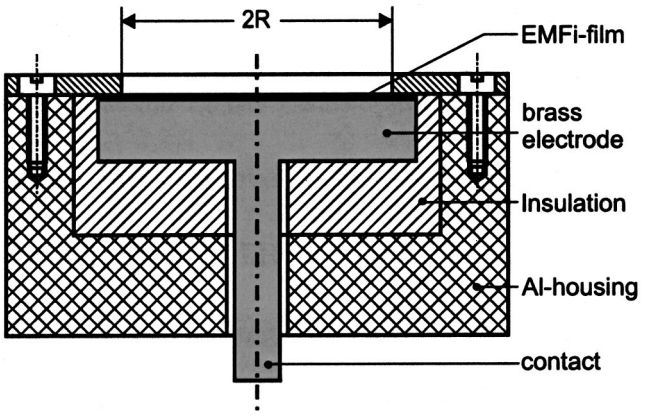


FIG. 1. Schematic of EMFi transducers with diameter $2R$.

range the deflection of the piezoelectric does not change with frequency. From interferometric measurements of dynamic deflection and dielectric measurements, a resonance frequency of about 300 kHz is known.[10] For the linear response, the transmitting sensitivity at 1 m distance is 4.7 mPa/V at 100 kHz.

The resonance frequency changes drastically, if the transducer is driven under water. The result shown in Fig. 3 is for an actuator of 35 mm diameter driven with 80 V at a distance of 20 cm. Though the boundary of the tank is strongly reverberant, it has been proved at a frequency of 10 kHz with gated actuation that the echoes from the walls are 10 dB weaker than the direct sound. The projector was mounted with its back close to the water surface to prevent contact at the electrical connectors with water. Therefore, the projector is baffled by the sound-soft water surface. Unfortunately, the adjustment of the reference hydrophone on the main axis of the sample was not very successful. This mainly results from the limited space for mounting transducers due to the small opening of the tank of just 20 cm diameter. Furthermore, the tank's upper wall does not allow the usage of heavy supports. Hence, the high frequency response of the sample can only be explained by assuming a deviation of 25 degrees off axis. Since this coincides with a -3 dB loss at 53 kHz it does not corrupt the data at lower frequencies.

The resonance shifts down to around 17 kHz which is attributed to the medium load. The transmitting sensitivity in water at 1 m distance is 8 mPa/V at 10 kHz.

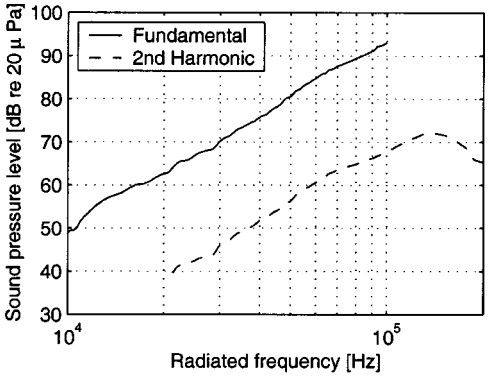


FIG. 2. Sound pressure level of an EMFi-HS actuator in air ($2R = 20$ mm) at 60 cm distance.

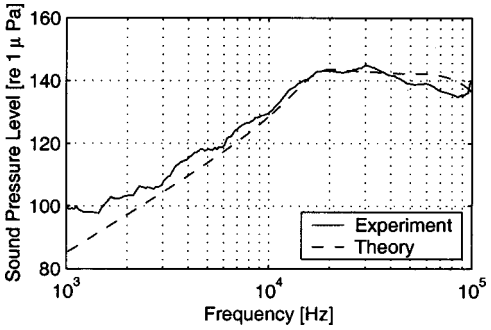


FIG. 3. Sound pressure level of an EMFi-HS projector ($2R = 35$ mm) at 20 cm distance 25° off axis. This unusual orientation results from the geometrical limitations of the small opening of the tank.

The transmitting sensitivity in air of an EMFi-OS actuator (not shown) of 20 mm diameter is 0.32 mPa/V at 100 kHz.

Both in water and in air, a quite strong second harmonic is present in the actuator experiments, whereas other harmonics do not occur. The second harmonic is due to an electrostatic effect as shown later in Eq. (7). Unpoled films only radiate the second harmonic and no fundamental. Poled films, too, show a similar behavior after thermal discharge. In this experiment, an EMFi actuator was placed inside an oven, and the radiated fundamental and second harmonic were recorded. The result is depicted in Fig. 4. The acoustic signals have not been very stable due to refraction caused by temperature gradients in the air between sample and microphone.

Discharge takes place between 70 °C to 100 °C and is irreversible.[1,3] This is not surprising, since also solid polypropylene is known to show weak charge storage capabilities.[11] The slight decrease of the sound pressure level during cooling is attributed to the lower Young's modulus at elevated temperature. After the experiment it could be seen at the first view that the sample could not further be used.

B. Sensor response

The sensitivity to air-borne sound of a transducer of 20 mm diameter is depicted in Fig. 5. At low frequencies, a three-port coupler with a diameter of 20 mm and a volume of 3 cm^3 was used. It guarantees a constant sound pressure up to 6 kHz. In the ultrasonics range, the substitution method in

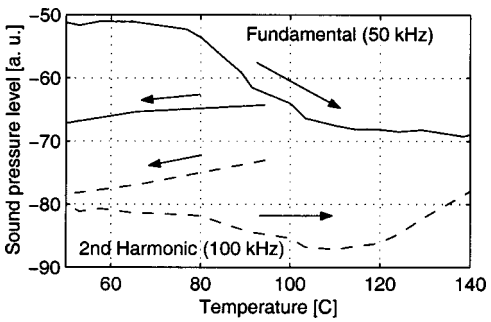


FIG. 4. Radiated sound pressure level in arbitrary units (a.u.) at 50 kHz [fundamental (–)] and 100 kHz [second harmonic (––)] versus temperature. Arrows indicate heating and cooling, respectively. The heating rate was 200 K/h.

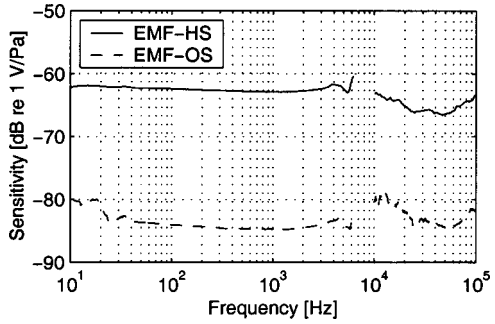


FIG. 5. Sensitivity of EMFi transducers to airborne sound ($2R=20$ mm). The low-frequency sensitivity was recorded with a three-port coupler with constant pressure up to 6 kHz. The sensitivity with respect to ultrasound was measured in free field using the substitution method. The difference of the values is attributed to the uncertainties of the measurements.

free field[12] using a baffled EMFi transducer as an actuator was used. The reference microphone and the EMFi sensor have been placed in the actuator's far field on the main axis and carefully adjusted with respect to inclination.

The frequency response of the EMFi-HS sensor is quite flat between 4 Hz and 6 kHz at about 0.7 mV/Pa. This corresponds to a longitudinal piezoelectric charge constant of about 100 pC/N, exceeding the data of PVDF, for instance, by about 10 dB. EMFi-OS sensors have a sensitivity of about 60 μV/Pa yielding a piezoelectric constant of about 15 pC/N. In the ultrasonics range, the sensors still work, but show strong directivity due to their size. The equivalent A-weighted noise level of the EMFi-HS microphone is 52 dB with the present setup. Even for sound pressure levels of up to 144 dB (at 300 Hz), the total harmonic distortion (THD) of both sensor types does not exceed that of the eighth-inch reference microphone (B & K 4138), which is just 3% at 168 dB. No change in the piezoelectric constant is observed for applied pressures between 20 mPa and 340 Pa.

The sensitivity of the EMFi-HS transducer used as hydrophone was measured with the comparison method and is shown in Fig. 6. Owing to the large reverberation time inside the tank, several frequency responses under various geometric configurations have been monitored. In this way, it is possible to average over the various spatial modes inside the tank. Below 6 kHz, the sound pressure level produced by the projector is too low to permit an accurate measurement of the sensitivity, and the number of modes inside the tank is too small to guarantee a successful spatial averaging.

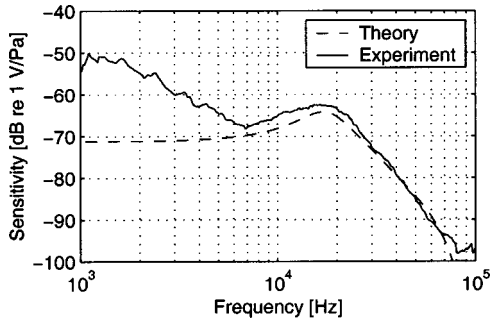


FIG. 6. Sensitivity of EMFi-HS under water (hydrophone, $2R=35$ mm). Below 6 kHz the data are corrupted by insufficient signal to noise ratio and unsuccessful spatial averaging.

From Fig. 6 it can be seen that the resonance is shifted down to 17 kHz as it appears in the actuator mode as well. Furthermore, the sensitivity below the resonance frequency is 0.3 mV/Pa, i.e., it is 8 dB lower in water than in air.

IV. THEORY

A. Mechanical–acoustical conversion

1. General remarks

In both air-borne and water-borne sound experiments the actuator can be modeled as a circular piston mounted in an infinite baffle[13] radiating the power $N=\mathrm{Re}\{Z_m\}v^2$ with the radiation impedance

$$Z_m=\rho c\,\pi R^2\left(1-\frac{J_1(2kR)}{kR}+j\,\frac{S_1(2kR)}{kR}\right) \tag{1}$$

and having a directivity ratio of

$$D=\frac{(kR)^2}{1-J_1(2kR)/kR}. \tag{2}$$

In these equations, v is the velocity of the vibrating piston, ρc is the characteristic impedance of the medium, R is the piston's radius, k is the wave number, and $J_1(x)$ and $S_1(x)$ are the Bessel and Struve function of first order, respectively.[14] Then, the sound pressure amplitude on the main axis at distance r can be computed:

$$p=\frac{1}{\sqrt{4\pi r^2}}\sqrt{\rho cND}=\frac{1}{2r}\,\rho R^2\omega^2 x, \tag{3}$$

where x is the deflection of the surface. The ω^2 increase of the pressure is confirmed experimentally up to about 100 kHz (see Fig. 2). The strong decrease of sound pressure of the second harmonic at frequencies above 150 kHz is mainly attributed to a decrease in sensitivity of the reference microphone.

The following calculations have been carried out for the HS type of EMFi films: The ω^2 increase with frequency holds from infra-sonic frequencies up to the mechanical resonance, above which x in Eq. (3) decreases. The mass per area is calculated from the geometry and the density of the film to be $M=7.7\times10^{-3}\,\mathrm{kg/m^2}$, in which it is considered that the dynamic mass is only one-third of the static one since the layer is clamped on one side. Any kind of impact of the backing on the resonance frequency can be excluded. The compliance is calculated from Young's modulus $Y=2$ MPa to be $F=3.5\times10^{-11}\,\mathrm{m^3/N}$. These data result in a resonance frequency of 307 kHz.

2. Fluid loading

For the water-borne sound measurement, carried out with transducers of 35 mm diameter, an additional mass load per area—computed from the imaginary part of Eq. (1) using a power expansion being valid for low frequencies—

$$M_s=\rho\,\frac{8R}{3\pi}=14.4\,\mathrm{kg/m^2} \tag{4}$$

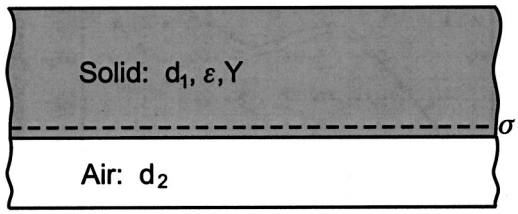


FIG. 7. Schematic of simplified model of EMFi film.

has to be considered. The shifted resonance frequency would be 7 kHz. Therefore, a further effect must be present, resulting in a slightly higher resonance and reduced sensor sensitivity at low frequencies. This effect seems to be an additional restoring force due to the medium: Considering that the same volume which is moved as a mass M_s is also compressed, the medium's compliant load times the area is

$$F_s = \kappa \frac{8R}{3\pi} = 8 \times 10^{-12}\,\mathrm{m^3/N}, \tag{5}$$

using the compressibility of water $\kappa = 5 \times 10^{-10}\,\mathrm{Pa^{-1}}$. As a consequence, the total compliance is reduced by a factor of about 5 which yields a resonance frequency of 18 kHz being in fair agreement with the experiment.

Stiffness terms in radiation impedances of fluid-loaded vibrating plates are well known in the literature for certain vibration modes.[15–17] In these experiments, however, we have no reason to assume any kind of mode distribution on the transducer's surface, and it must be premised that the surface vibrates piston-like. Furthermore, radiation impedances with negative imaginary part only appear in a small frequency range,[15] whereas the present experiments show, that both mass and compliant loading from the fluid take place over a broader frequency range. This effect might not have been observed in the past since its impact was small. From Eq. (4), it can be seen that mass loading increases with increasing sensor area.

B. Electro-mechanical conversion

The electro-mechanical coupling of a stack of two dielectrics of different elasticity with a charge layer in between which is similar to an electret microphone[18] has been already investigated[19] and extended to charged cellular polymers.[2] The latter model assumes a material composed of alternating air and solid layers with a finite charge density σ at the boundary. It can be further simplified as shown in Fig. 7. The following transducer constant, identical both for sensor and actuator mode, is deduced:

$$d_{33} = \frac{\varepsilon d}{Y} \frac{d_1 \sigma}{(d_1 + \varepsilon d_2)^2} = \frac{x}{U}. \tag{6}$$

In this equation, for HS-type transducers, $d_1 = 26\,\mu\mathrm{m}$ is the total thickness of the solid layers, $\varepsilon = 2.35$ is their permittivity, and $d_2 = 44\,\mu\mathrm{m}$ is the total thickness of the air layers. d_1 and d_2 are calculated from the total thickness d of 70 μm and the density of the film. From the measured sound pressure at the fundamental, a transducer constant of about 80 pm/V is computed from Eqs. (3) and (6). The respective data for the EMFi-OS actuator are $d_1 = 23\,\mu\mathrm{m}$, $d_2 = 14\,\mu\mathrm{m}$, $d_{33} = 15\,\mathrm{pm/V}$.

A Young's modulus of $Y = 2$ MPa (EMFi-HS) and $Y = 6$ MPa (EMFi-OS) was calculated from the electrostatic force per area A,

$$\frac{F_{el}}{A} = \frac{\varepsilon}{2d^2} U^2 = \frac{Y}{d} x \tag{7}$$

which results for the actuating mode in radiation of the second harmonic due to the nonlinearity with respect to the driving voltage (Fig. 2). Inserting these data into Eq. (6), a charge density of $\sigma = 6 \times 10^{-4}\,\mathrm{C/m^2}$ for EMFi-HS and of $\sigma = 1.6 \times 10^{-4}\,\mathrm{C/m^2}$ for EMFi-OS are computed. The actuator constant has been also deduced from interferometric measurement of deflection to be between 85 and 100 pm/V. From this experiment a resonance frequency of about 300 kHz was obtained[10] resulting in a Young's modulus of 2 MPa.

V. DISCUSSION

The obtained piezoelectric constants vary up to 20% from sample to sample, but are, on average, much smaller than the data reported in the literature.[2,20] These published higher values are observed by applying a static load of about 10^4 Pa to the sample, whereas in the present sensor experiments, the load was between 20 mPa and 340 Pa. The author's experiments with static loading have shown, that in this method a nonlinearity with respect to the loading appears. Furthermore, in the large-signal regime, the piezoelectrically induced charge increases, if just a small area of the film is loaded. In this case, the air can flow from the bubbles of the loaded area into the unloaded neighborhood resulting in an apparently softer film. This also explains the slow relaxation reported in Ref. 20 which was also observed in our experiments with quasi-static loading. The nonlinearity with respect to loading will be discussed in more detail in a future publication. A similar low value of 110 pm/V is reported in Ref. 21 for the actuating mode monitoring the sample's deflection under electrical load with an accelerometer.

Since the nonlinearity appears in the stress–strain relation, it affects both the measurement of piezoelectric constant and Young's modulus. But, the product of both, related to the surface charge at the interface [see Eq. (6)] is independent of this effect. Therefore, the lower Young's modulus of about 1 MPa reported in Refs. 2 and 20 leads to the same surface charge.

Charged porous polymers show longitudinal piezoelectric constants being much larger than those known for other polymers, e.g., polyvinylidene fluoride (PVDF). This follows mainly from the very low Young's modulus of EMFi-HS layers which also results in a very low acoustic impedance of about $Z = \sqrt{\rho Y} = 2.6 \times 10^4\,\mathrm{kg/(m^2\,s)}$. Therefore, cellular layers are matched better to air than to water. It should be clearly stated, that EMFi films are no piezoelectrics in the classical sense. Polypropylene does not contain any dipoles, and our measurement of the transversal coefficient yielded

the very low value of $d_{31}=2$ pC/N. The same value is reported in Ref. 10. Hence, EMFi films have been called pseudo-piezoelectrics.[2]

The observed low stability of EMFi films against heating is insufficient for many applications; it is inferior to that of other common piezoelectrics, e.g., PVDF and PZT. This problem might be overcome by using Teflon as a basic material known to have excellent charge storage properties.[11] Charge retention in porous polytetrafluorethylene has been already investigated,[4,5] and the piezoelectricity of various solid double layers with charge layer at the boundary was reported,[6–8] but no experiments on temperature dependence and transducers are reported up to now. In the latter publication, piezoelectric coefficients up to 600 pC/N are reported for double layers, obtained from quasi-static measurements. It should be remembered that our own measurements were all obtained on the simpler and mechanically more stable single layers.

VI. CONCLUSIONS

The presented acoustic experiments on transducers for water-borne and air-borne sound show that charged cellular polymers are promising candidates for several applications, mainly as hydrophones in shallow water or in air as ultrasonic sources or broadband sensors. The flexibility of the polymer allows the fabrication of large-scale transducers with almost arbitrary directivity pattern (e.g., a true omnidirectional transducer at high frequencies), if the film is fixed on a curved substrate which can also contain a more or less sophisticated electrode pattern. The underwater experiments seem to indicate that a low impedance material as the EMFi film is loaded by the medium simultaneously mass-like and compliant. The very weak transversal piezoelectricity of 2 pC/N (see above and Ref. 10) avoids bending of actuators glued on a substrate. For narrow-band applications, the resonance frequency can be tuned by applying an additional mass on top of the film which might be also desirable for protection against any kind of environmental influence.

The main future scope is to fabricate layers with higher temperature stability. Furthermore, the interaction between EMFi films and water should be investigated more deeply. Finally, the microscopic structure—both mechanically and electrically—of cellular polymers is of great interest to improve material properties and to understand the nonlinear behavior. Higher piezoelectric constants are possible by further softening films through enhanced gas content and higher charge density.

ACKNOWLEDGMENTS

The author thanks Professor G. M. Sessler, H. Berger, M. Fischer, J. Meyer, and F. Steffens for stimulating discussions and help with some of the experiments. Brüel and Kjær, Copenhagen, Denmark, is gratefully acknowledged for providing the hydrophone, and VTT, Tampere, Finland, for making the EMFi films available. Financial support of the Deutsche Forschungsgemeinschaft (DFG) is also acknowledged.

[1] J. Lekkala and M. Paajanen, ''EMFi—New electret material for sensors and actuators,'' *Proceedings of the 10th International Symposium on Electrets, Delphi* (IEEE, Piscataway, NJ, 1999), pp. 743–746.
[2] G. M. Sessler and J. Hillenbrand, ''Electromechanical response of cellular electret films,'' Appl. Phys. Lett. **75**, 3405–3407 (1999).
[3] J. v. Turnhout, R. E. Staal, M. Wübbenhorst, and P. H. de Haan, ''Distribution and stability of charges in porous polypropylene films,'' *Proceedings of the 10th International Symposium on Electrets, Delphi* (IEEE, Piscataway, NJ, 1999), pp. 785–788.
[4] Z. Xia, J. Jian, Y. Zhang, Y. Cao, and Z. Wang, ''Electret properties for porous polytetrafluorethylene (PTFE) Film,'' *CEIDP, 1997 Annual Report* (IEEE, Piscataway, NJ, 1997), pp. 471–474.
[5] R. Schwödiauer, G. Neugschwandtner, S. Bauer-Gogonea, S. Bauer, J. Heitz, and D. Bäuerle, ''Dielectric and electret properties of novel Teflon PTFE and PTFE-like polymers,'' *Proceedings of the 10th International Symposium on Electrets, Delphi* (IEEE, Piscataway, NJ, 1999), pp. 313–316.
[6] R. Gerhard-Multhaupt, Z. Xia, W. Künstler, and A. Pucher, ''Preliminary study of multi-layer space-charge electrets with piezoelectric properties from porous and non-porous Teflon films,'' *Proceedings of the 10th International Symposium on Electrets, Delphi* (IEEE, Piscataway, NJ, 1999), pp. 273–276.
[7] W. Künstler, Z. Xia, T. Weinhold, A. Pucher, and R. Gerhard-Multhaupt, ''Piezoelectricity of porous polytetrafluoroethylene single- and multiple-film electrets containing high charge densities of both polarities,'' Appl. Phys. A: Mater. Sci. Process. **70**, 5–8 (2000).
[8] G. Neugschwandtner, R. Schwödiauer, S. Bauer-Gogonea, and S. Bauer, ''Large piezoelectric effects in charged, heterogeneous fluoropolymer electrets,'' Appl. Phys. A: Mater. Sci. Process. **70**, 1–4 (2000).
[9] Y. Cao, Z. Xia, Q. Li, L. Chen, and B. Zhou, ''Study of porous dielectrics as electret materials,'' IEEE Trans. Dielectr. Electr. Insul. **5**, 58–62 (1998).
[10] G. S. Neugschwandtner, R. Schwödiauer, M. Vieytes, S. Bauer-Gogonea, S. Bauer, J. Hillenbrand, R. Kressmann, G. M. Sessler, M. Paajanen, and J. Lekkala, ''Large and broadband piezoelectricity in charged polypropylene foam electrets,'' Appl. Phys. Lett. **77**, 3827–3829 (2000).
[11] R. Kressmann, G. M. Sessler, and P. Günther, ''Space-charge electrets,'' in *Electrets*, 3rd ed., edited by R. Gerhard-Multhaupt (Laplacian, Morgan Hill, 1999), Vol. 2, Chap. 9, pp. 1–40.
[12] V. Nedzelnitsky, ''Calibration of pressure and gradient microphones,'' in *Encyclopedia of Acoustics*, edited by M. J. Crocker (Wiley, New York, 1997), Vol. 4, Chap. 157, pp. 1869–1879.
[13] H. F. Olson, *Acoustical Engineering* (van Nostrand, New York, 1957).
[14] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, 9th ed. (Dover, New York, 1970).
[15] M. Lax, ''The effect of radiation on the vibrations of a circular diaphragm,'' J. Acoust. Soc. Am. **16**, 5–13 (1944).
[16] P. W. Smith, Jr., ''The imaginary part of input admittance: a physical explanation of fluid-loading effects on plates,'' J. Sound Vib. **60**, 213–216 (1978).
[17] D. G. Crighton, ''The 1988 Rayleigh medal lecture: Fluid loading—The interaction between sound and vibration,'' J. Sound Vib. **133**, 1–27 (1989).
[18] G. M. Sessler and J. E. West, ''Applications,'' in *Electrets*, 3rd ed., edited by G. M. Sessler (Laplacian, Morgan Hill, 1999), Vol. 1, Chap. 7, pp. 347–381.
[19] R. Kacprzyk, E. Motyl, J. B. Gajewski, and A. Pasternak, ''Piezoelectric properties of nonuniform electrets,'' J. Electrost. **35**, 161–166 (1995).
[20] M. Paajanen, H. Välimäki, and J. Lekkala, ''Modelling the electromechanical film (EMFi),'' J. Electrost. **48**, 193–204 (2000).
[21] J. Hillenbrand and G. M. Sessler, ''Piezoelectricity in cellular electret films,'' IEEE Trans. Dielectr. Electr. Insul. **7**, 537–542 (2000).

Wave transmission through structural inserts

B. R. Mace[a)]
Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom

R. W. Jones
Department of Computer Science and Electrical Engineering, Luleå University of Technology, SE-97187 Luleå, Sweden

N. R. Harland
Department of Mechanical Engineering, The University of Auckland, Private Bag, Auckland, New Zealand

(Received 5 September 2000; revised 2 January 2001; accepted 3 January 2001)

The transmission of waves through two discontinuities in a one-dimensional waveguide system is considered. Attention is focused on transmission through a structural insert, which is defined here to be a waveguide segment which is inserted into an otherwise continuous structural member with different properties. A general expression for the net transmission through the insert is found. It has bandpass/stop characteristics and its frequency average is somewhat greater than that normally assumed due to the coherent interaction of the waves in the insert. The particular case is then considered where the insert comprises a three-layer composite beam inserted in a thin beam which vibrates in bending. The composite beam comprises two elastic faceplates and a core filled with a tunable electro- or magneto-rheological fluid. The net transmission and the stop bands depend on the properties of the insert. Since these properties are tunable by adjusting the field to which the tunable fluid is exposed, then so too are the transmission characteristics of the insert. [DOI: 10.1121/1.1352081]

PACS numbers: 43.40.Cw [CBB]

I. INTRODUCTION

Vibrations travel through structures as waves, carrying energy from one part to another. This paper considers one specific aspect of this, namely the net wave transmission through two discontinuities on a one-dimensional wave-bearing structure, and a technique for controlling this transmission is described.

The situation considered is shown in Fig. 1. Three one-dimensional waveguides *a*, *b*, and *c* are connected at junctions 1 and 2. The motion is time harmonic at frequency ω. It is desired to find the amplitudes $\mathbf{c}^+$ of the transmitted waves in terms of the amplitudes $\mathbf{a}^+$ of the waves incident on the first junction. The waveguides may be rods, beams, pipes, etc. In particular, the case is considered where waveguides *a* and *c* have identical properties which differ from those of waveguide *b*. This is referred to here as a structural insert: a segment *b* is inserted into (or attached in parallel with) an otherwise continuous structural member to control the transmission of vibration. The length of the insert is typically half a wavelength or more, so that it can support standing waves. It is seen below that the insert has bandpass/stop transmission characteristics. These frequency bands can be controlled by suitably designing the properties of the insert. Furthermore, in a tunable insert (e.g., the tunable fluid filled three-layer beam described in Sec. III C) the properties can be tuned on-line to accommodate design and modeling inaccuracies or to allow the stop bands to be continuously retuned in the face of incident waves of varying frequency content.

In the next section the reflection and transmission matrices of the insert are found. Then, some examples are considered. The first simple example concerns two point attachments to a single, uniform rod-like waveguide. Bandpass transmission characteristics are seen and the frequency average transmission is seen to differ from the value normally assumed, due to coherent interference of the reflected waves within section *b*. Then, bending waves in thin beams are considered. Finally, results are presented for a tunable fluid-filled insert, for which the stop bands are seen to be tunable. This insert is a composite beam element comprising two elastic outer layers with a core of tunable electro-rheological (ER) or magneto-rheological (MR) fluid. Wave propagation in such a waveguide was considered by Harland *et al.* (2000). The properties of the fluid can be varied by changing the applied field; hence, so too can the wave behavior of the composite beam and the net transmission through the insert.

II. TRANSMISSION AND REFLECTION MATRICES

Each waveguide in Fig. 1 carries n distinct wave modes. For example, for a bar in axial vibration only, $n=1$, for bending waves in a single plane in a thin beam $n=2$ (the propagating and near-field waves), while for a general one-dimensional structural member $n=6$ (axial and torsional waves, and bending waves in two perpendicular planes). For a symmetric tunable fluid-filled beam, $n=3$. The wave amplitudes at any point in the waveguide can be written as $n\times 1$ vectors of positive- and negative-going waves.

The junctions may mark a change in waveguide properties or some point at which there is some attachment to the

a)Electronic mail: brm@isvr.soton.ac.uk

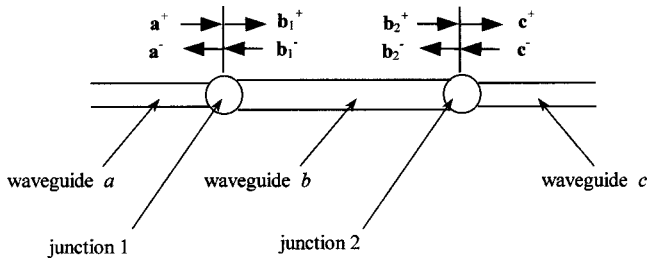


FIG. 1. Waveguides and junctions.

waveguide (e.g., a support, attached mass, etc.). A wave incident on a junction is partly transmitted and partly reflected, the amplitudes of these waves being given by the appropriate transmission and reflection matrices of the junction. In general, wave mode conversion occurs, in that an incident wave of one mode is scattered into reflected and transmitted waves of all the modes that are supported by the waveguides.

Consider now waves $\mathbf{a}^+$ incident on junction 1 between waveguides *a* and *b*. The wave amplitudes there are related by

$$\mathbf{a}^- = \mathbf{r}_1^{aa}\mathbf{a}^+ + \mathbf{t}_1^{ba}\mathbf{b}_1^-; \quad \mathbf{b}_1^+ = \mathbf{t}_1^{ab}\mathbf{a}^+ + \mathbf{r}_1^{bb}\mathbf{b}_1^-, \tag{1}$$

where **a**, **b**, and **c** are the vectors of wave amplitudes and the superscripts identify the direction of propagation. Here, $\mathbf{r}_1$ and $\mathbf{t}_1$ are matrices of reflection and transmission coefficients for junction 1, with the two superscripts, respectively, identifying the waveguides along which the incoming and outgoing waves propagate. These matrices can be determined from the continuity and equilibrium conditions applicable to the junction (e.g., Cremer *et al.*, 1988; Harland *et al.*, 2000). The reflection matrices $\mathbf{r}_1^{aa}$ and $\mathbf{r}_1^{bb}$ are square and of dimension n_a and n_b, respectively, while the transmission matrices are only square if $n_a = n_b$. The transmitted waves $\mathbf{b}_1^+$ then propagate along waveguide *b*. When incident upon the junction with waveguide *c* they are further reflected and transmitted, the reflected and transmitted wave amplitudes being given similarly by

$$\mathbf{b}_2^- = \mathbf{r}_2^{bb}\mathbf{b}_2^+ + \mathbf{t}_2^{cb}\mathbf{c}^-; \quad \mathbf{c}^+ = \mathbf{t}_2^{bc}\mathbf{b}_2^+ + \mathbf{r}_2^{cc}\mathbf{c}^-. \tag{2}$$

The reflected waves $\mathbf{b}_2^-$ propagate back along the waveguide. The wave amplitudes at the ends of waveguide *b* are related by the propagation relations

$$\mathbf{b}_2^+ = \mathbf{f}_b\mathbf{b}_1^+; \quad \mathbf{b}_1^- = \mathbf{f}_b\mathbf{b}_2^-, \tag{3}$$

where the propagation matrix $\mathbf{f}_b$ is diagonal, the jth diagonal element being $\exp(-ik_j l)$, where k_j is the wave number of the jth wave mode and l is the length of waveguide *b*. The wave numbers may be complex. If damping is present then the wave numbers have negative imaginary parts, so that the amplitude of a wave decays as it propagates. The wave field in waveguide *b* thus comprises the superposition of multiple reflections between the two junctions. It is the interference between successive reflections which produces the bandpass/stop behavior.

The net transmission and reflection matrices are such that, if $\mathbf{c}^- = 0$,

$$\mathbf{c}^+ = \mathbf{t}^{ac}\mathbf{a}^+; \quad \mathbf{a}^- = \mathbf{r}^{aa}\mathbf{a}^+, \tag{4}$$

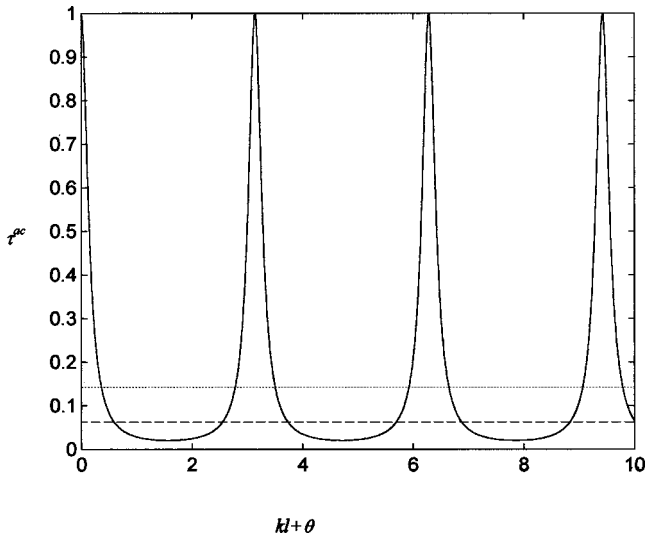


FIG. 2. Transmitted power τ^{ac} as a function of phase $kl+\theta$, single wave mode, $T=0.5$: --- T^4; ⋯ frequency average.

They can be found from the above equations and are given by

$$\mathbf{t}^{ac} = \mathbf{t}_2^{bc}(\mathbf{I} - \mathbf{f}_b\mathbf{r}_1^{bb}\mathbf{f}_b\mathbf{r}_2^{bb})^{-1}\mathbf{f}_b\mathbf{t}_1^{ab};$$
$$\mathbf{r}^{aa} = \mathbf{r}_1^{aa} + \mathbf{t}_1^{ba}(\mathbf{I} - \mathbf{f}_b\mathbf{r}_2^{bb}\mathbf{f}_b\mathbf{r}_1^{bb})^{-1}\mathbf{f}_b\mathbf{r}_2^{bb}\mathbf{f}_b\mathbf{t}_1^{ab}. \tag{5}$$

These matrices are the sum of a ''direct'' component (e.g., $\mathbf{r}_1^{aa}$) and components arising from the multiple reflections at the junctions. Similar expressions can be found for $\mathbf{t}^{ca}$ and $\mathbf{r}^{cc}$ by reversing superscripts and subscripts where relevant. The power carried by a specific wave component is proportional to $|t^{ac}|^2$ since the energy carried by a wave is proportional to the square of its amplitude.

It is normal when estimating the power transmitted through two or more discontinuities to ignore the effects of the coherent interference of the waves in region *b* (Cremer *et al.*, 1988). This is particularly the case when the discontinuities are well separated compared to the wavelength and the disturbances broadband. Here, however, the separation is of the order of the wavelength. The amplitude of the transmitted power is then approximated by

$$\hat{\mathbf{t}}^{ac} = \mathbf{t}_2^{bc}\mathbf{f}_b\mathbf{t}_1^{ab}. \tag{6}$$

Clearly, this may give discrete frequency errors when compared to the exact result in Eq. (5). However, the assumption is implicitly made that the approximation is accurate when averaged over frequency, since the relative phases of the waves in region *b* are uniformly random over a broad frequency band. As will be seen in the examples below, this is not true however. In particular, $\mathbf{t}^{ac}$ shows a pass/stop band structure and its frequency average is larger than that of $\hat{\mathbf{t}}^{ac}$.

III. EXAMPLES

A. Single wave mode

The net transmission is given in terms of the system properties by a rather complicated expression. However, it is possible to gain some physical insight by considering the simplest case, where the waveguides are undamped, where

each carries only one wave mode (for example axial or torsional waves in a rod) and where there are identical discontinuities at the two junctions.

The junctions are described by reflection and transmission coefficients $r=R\exp(-i\theta)$ and t, respectively, where $|t|=T$, R and T are real, and θ is some (typically slowly) frequency-dependent phase. From conservation of energy it follows that $R^2+T^2=1$. Thus, $\mathbf{r}_1^{bb}=\mathbf{r}_2^{bb}=r$, $\mathbf{t}_1^{ab}=\mathbf{t}_2^{bc}=t$ and $\mathbf{f}_b=\exp(-ikl)$, where $k=\omega/c$ is the wave number of the single wave mode, c being the wave speed. Equation (5) then reduces to

$$t^{ac}=\frac{e^{-ikl}}{1-e^{-i2(kl+\theta)}R^2}t^2. \tag{7}$$

The proportion of the incident power that is transmitted is given by

$$\tau^{ac}=|t^{ac}|^2=\frac{T^4}{1+R^4-2R^2\cos 2(kl+\theta)}. \tag{8}$$

Since k is proportional to frequency, the transmitted power is strongly dependent on frequency and shows pass/stop band behavior as illustrated in Fig. 2. Between the discontinuities, the net wave field consists of the superposition and interference of multiple reflections. At certain frequencies, when $kl+\theta=m\pi$, where m is any integer (the centers of the pass bands) these waves interfere constructively and the net transmission is large, with $|t^{ac}|=1$. Internal resonance occurs, with, from Eqs. (1)–(3),

$$|b^+|=\frac{1}{T}|a^+|; \quad |b^-|=\frac{R}{T}|a^+|. \tag{9}$$

At other frequencies destructive interference occurs and the net transmission may be small. The frequency average transmitted power is found by averaging over frequency or, equivalently here, wave number k. By averaging Eq. (8) over the range $0\leqslant 2(kl+\theta)<2\pi$, the frequency average is found to be

$$\langle\tau^{ac}\rangle=\frac{T^4}{1-R^4}, \tag{10}$$

and is thus somewhat larger (and perhaps substantially so) than $\hat{\tau}^{ac}=T^4$. The difference arises from the coherence of the wave field between the two discontinuities: those frequencies where the transmission is larger than average typically produce substantially larger transmission, thus biasing the average. The geometric mean (i.e., the average of $\log\tau^{ac}$) transmission equals T^4, however, and this value is more typical of the transmission at any discrete frequency.

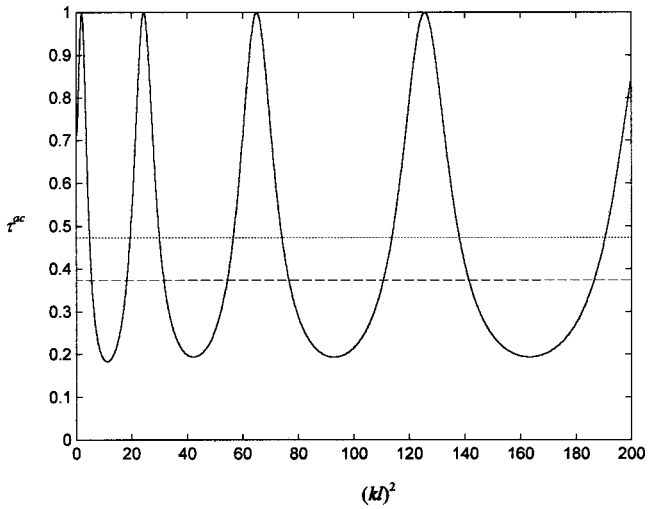


FIG. 3. Transmitted power τ^{ac} as a function of $(kl)^2\propto\omega$, beam insert with thickness equal to 25% of that of main waveguide; --- ''incoherent'' value, $\cdots$ frequency average.

B. Flexural waves—Thin beam inserts

The waveguides in the second example considered are each thin, Euler–Bernoulli beams which vibrate in bending. Such beams are commonly used as structural members. Now a second wave mode exists, namely the nonpropagating near field (Cremer *et al.*, 1988). This affects the overall behavior at low frequencies but not at higher frequencies, where near-field effects are localized around the junctions. Furthermore, the waves are dispersive, the wave number being $k=\sqrt[4]{m/EI}\sqrt{\omega}$, where m and EI are the beam mass per unit length and bending stiffness, respectively.

Suppose waveguides a and c are identical, while waveguide b is made of the same material but has a different thickness. Waveguide b can then be considered as a structural insert in an otherwise uniform member. The reflection and transmission matrices required in Eq. (5) are (Mace, 1984)

$$\mathbf{r}_1^{bb}=\mathbf{r}_2^{bb}=\frac{2}{\Delta}\begin{bmatrix}-2(\beta^2-1)\gamma-i\beta(1-\gamma)^2 & (1+i)\beta(1-\gamma^2)\\ (1-i)\beta(1-\gamma^2) & -2(\beta^2-1)\gamma+i\beta(1-\gamma)^2\end{bmatrix},$$
$$\mathbf{t}_1^{ba}=\mathbf{t}_2^{bc}=\frac{4}{\Delta}\begin{bmatrix}(1+\beta)(1+\gamma) & -(1-i\beta)(1-\gamma)\\ -(1+i\beta)(1-\gamma) & (1+\beta)(1+\gamma)\end{bmatrix}, \tag{11}$$

where the parameters

$$\beta=\frac{k_a}{k_b}; \quad \gamma=\frac{(EIk^2)_a}{(EIk^2)_b}; \quad \Delta=(1+\beta)^2(1+\gamma)^2-(1+\beta^2)(1-\gamma)^2. \tag{12}$$

Furthermore

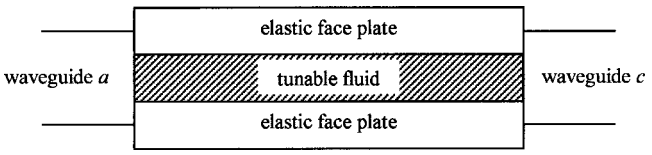


FIG. 4. Tunable fluid-filled beam insert.

$$\mathbf{f}_b = \begin{bmatrix} e^{-ikl} & 0 \\ 0 & e^{-kl} \end{bmatrix}. \tag{13}$$

The matrix $\mathbf{t}_l^{ab}$ follows from Eq. (11) with a suitable change of subscripts.

The transmitted energy per unit incident energy, τ^{ac}, shows similar pass/stop band characteristics (Fig. 3), with successive bands having different bandwidths due to dispersion. Figure 3 also shows the frequency average transmission and the conventional value predicted under the assumption of incoherent reflections within the insert $|t_1^{ab} t_2^{bc}|^2$. The near-field term in $\mathbf{f}_b$ is only important at low frequencies. At higher frequencies the behavior becomes analogous to the case discussed in Sec. III A, except that the wave motion is dispersive.

Clearly, this effect can be exploited in regions of the structure with different static load-bearing requirements, both to minimize structural mass and to reduce higher frequency vibration transmission. However, a large change in section is required to produce substantial attenuation, which would limit applicability. More sophisticated inserts, such as the split-beam isolator considered by Shoavi (1991), can show substantially better stop-band behavior without compromising static strength.

C. Tunable fluid-filled insert

As a final example consider the case of the system shown in Fig. 4. This consists of a uniform Euler–Bernoulli (EB) beam with an insert (waveguide *b*) which consists of a tunable fluid-filled composite beam. The insert comprises two elastic faceplates of thickness h_1 with a central core of thickness h_2 filled with a tunable fluid such as an electro-rheological (ER) or magneto-rheological (MR) fluid. When the fluid is in the preyield state the composite beam behaves like a sandwich beam with a visco-elastic core, as modeled by Mead and Markus (1969). However, the shear modulus G_2 of the middle layer is tunable. It depends on the electric or magnetic field to which the tunable fluid is subjected. For typical fluids the shear modulus can be approximated by (Don, 1993)

$$G_2 = G' + iG''; \quad G' = A_0 + A_1 F^B; \quad G'' = C_0 + C_1 F^D, \tag{14}$$

where *A*, *B*, *C*, and *D* are fluid-dependent constants and *F* is the applied field (be it electric or magnetic). Thus, in the preyield state the shear properties of such fluids are tunable and can be changed by varying the electric or magnetic field the fluid experiences.

TABLE I. Properties of Euler–Bernoulli beam.

Thickness	5×10^{-3}
E	7.27×10^{10}
ρ	2.7×10^{3}

TABLE II. Properties of insert.

Thickness of elastic layers	1×10^{-3}
Thickness of fluid layer	3×10^{-3}
l	0.12

The wave propagation, reflection, and transmission characteristics of the tunable fluid filled beam were considered by Harland *et al.* (2000). Three wave modes exist, the wave numbers being the solutions to

$$-k^6 - g[1+Y]k^4 + \frac{\overline{m}}{\overline{EI}}k^2\omega^2 + g\frac{\overline{m}}{\overline{EI}}\omega^2 = 0, \tag{15}$$

where $\overline{EI} = 2E_l h_1^3 L_y/12$ is the total bending stiffness of the elastic layers, $\overline{m}$ is the total mass per unit length of beam, $g = 2G_2^*/E_l h_1 h_2$ is the complex shear parameter and $Y = (h_1+h_2)^2 L_y E_1 h_1/2\overline{EI}$ is known as Kerwin's geometric parameter. Here, E_1 is the elastic modulus of the outer layers and L_y is the width of the beam. Two wave modes are in essence bending-like modes while the third is a ''push–pull''-type motion in the outer faceplates. At higher field strengths, though, such a simplified interpretation is not possible.

The wave numbers of these waves depend on the shear modulus of the central core and hence on the applied field. The wave numbers are thus tunable, hence, so too are the pass/stop band characteristics of the insert. [The reflection and transmission matrices also depend on the applied field, but to a much lesser degree (Harland *et al.*, 2000).] In particular the stop bands can be tuned to minimize the transmitted energy in a desired frequency range. This gives the potential to tune the transmission characteristics of the structure on-line.

As a numerical example, consider a tunable fluid-filled beam inserted into a thin, Euler–Bernoulli beam. The beam and insert properties are given in Tables I and II, respectively, with the properties of the elastic outer layers of the insert being the same as those of the main beam. The properties of the tunable fluid are given in Table III and are taken from manufacturers' data or by fitting expressions of the form of Eq. (3) to the available data.

Figure 5 shows the transmitted power per unit incident power for various electric fields for the ER fluid-filled case. In Fig. 5(a) there is a distinct minimum at a particular frequency, and this frequency is tunable by changing the electric field. The bandwidth is approximately equal to the center frequency. Such an insert is therefore especially suited to the isolation of structure-borne vibrations in those cases where

TABLE III. Properties of tunable fluids.

	ER fluid	MR fluid
G'	$5\times10^4 F^2$	$4.00\times10^3 + 7.96\times10^5 F^{1.65}$
G''	$4340 F^{0.35}$	$4.00\times10^3 + 7.96\times10^5 F^{1.65}$
Units of F	kV/mm	kOe/mm
ρ	1.06×10^3	3.28×10^3

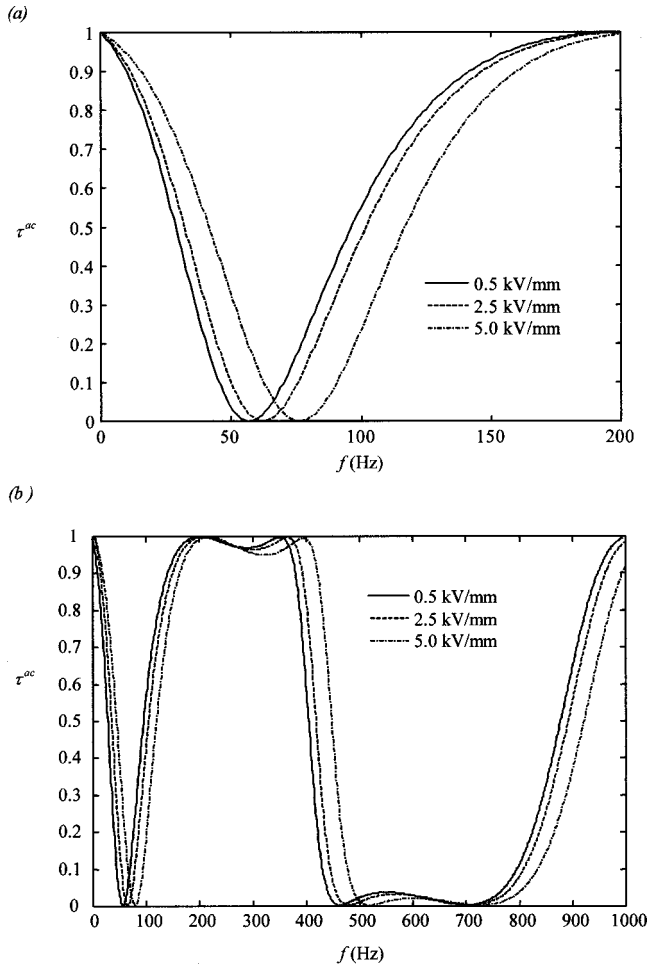


FIG. 5. Transmitted power per unit incident power for ER fluid-filled insert.

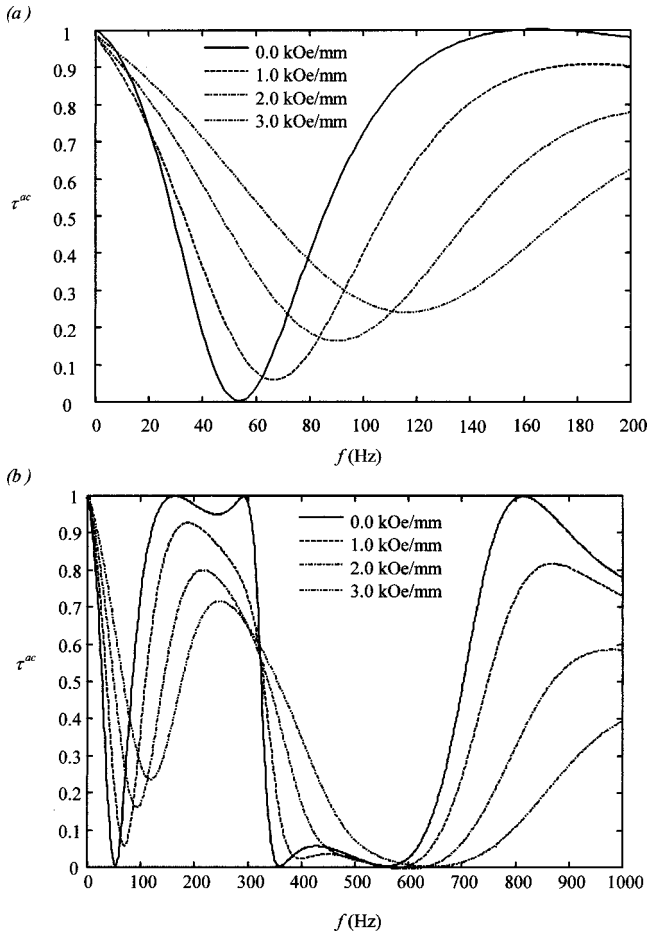


FIG. 6. Transmitted power per unit incident power for MR fluid-filled insert.

the incident power is somewhat narrow band. At higher frequencies [Fig. 5(b)] a clear pass/stop band structure can be seen. This is also tunable, but to a lesser extent.

Figure 6 shows the MR fluid-filled case. The degree of tunability is substantially greater than in the ER case, since the MR fluid is much stronger. Moreover, the MR fluid introduces substantially larger damping levels. This tends to lead to somewhat higher transmitted energy levels in the stop bands, while at the same time giving substantial levels of energy absorption. The MR case is thus better suited to adding tunable damping rather than tunable isolation.

IV. CONCLUDING REMARKS

This paper considered the net vibration transmission through two discontinuities in a one-dimensional waveguide system, and in particular through a structural insert. The transmission shows pass/stop band characteristics, the bands depending on the insert properties. The frequency average transmission is somewhat greater than that normally assumed due to the coherent interaction of the waves in the insert. This behavior offers scope for vibration control either at the design stage or as a retrofit. If a tunable insert is used, such as the tunable fluid-filled composite beam considered as an example, the stop bands may be tuned and retuned on-line.

Of course, there may be further reflections from some third discontinuity at some point downstream in waveguide *c*. If these are strong enough then further reflected waves will be generated in the downstream region, and that will affect the net transmission: the situation becomes more complicated. However, the effects of these downstream reflections will be small if the reflections are not strong [cf. R^2 in Eq. (7) is small] or if the effects of damping are large enough [cf. $\exp(-i(2kl+\theta))$ in Eq. (7) becomes small].

Cremer, L., Heckl, M., and Ungar, E. E. (**1988**). *Structure-Borne Sound* (Springer, Berlin).

Don, D. L. (**1993**). ''An Investigation of Electrorheological Material Adaptive Structures,'' Master's thesis, Lehigh University.

Harland, N. R., Mace, B. R., and Jones, R. W. (**2001**). ''Wave propagation, transmission and reflection in electrorheological fluid-filled beams,'' J. Sound Vib. (in press).

Mace, B. R. (**1984**). ''Wave reflection and transmission in beams,'' J. Sound Vib. **97**, 237–246.

Mead, D. J., and Markus, S. (**1969**). ''The forced vibration of a three-layer, damped sandwich beam with arbitrary boundary conditions,'' J. Sound Vib. **10**, 163–175.

Shoavi, E. (**1991**). ''Reduction of Vibration Transmission by Destructive Interference,'' Ph.D. thesis, The University of Auckland.

Time-domain simulation of damped impacted plates. I. Theory and experiments

Antoine Chaigne[a] and Christophe Lambourg[b]
Ecole Nationale Supérieure des Télécommunications, Département TSI, CNRS URA 820, 46 Rue Barrault, 75634 Paris Cedex 13, France

(Received 3 December 1999; revised 11 January 2001; accepted 13 January 2001)

A time-domain formulation for the flexural vibrations in damped rectangular isotropic and orthotropic plates is developed, in order to investigate transient excitation of plates by means of sound synthesis. The model includes three basic mechanisms of damping (thermoelasticity, viscoelasticity and radiation) using a general differential operator. The four rigidity factors of the plate are modified by perturbation terms, each term corresponding to one specific damping mechanism. The first damping term is derived from the coupling between the thermoelastic stress–strain relations and the heat diffusion equation. The second term is obtained from the general differential formulation of viscoelasticity. The third term is obtained through a Padé approximation of the damping factor which governs the coupling of the plate with the surrounding air. The decay factors predicted by the model reproduce adequately the dependence on both dimensions and frequency of the decay factors measured on rectangular plates of various sizes and thicknesses made of four different materials (aluminum, glass, carbon fiber, and wood). The numerical resolution of the complete problem, including initial and boundary conditions, and the comparison between real and simulated sounds are presented in a companion paper [J. Acoust. Soc. Am. **109**, 1433–1447 (2000)]. [DOI: 10.1121/1.1354200]

PACS numbers: 43.40.Dx [CBB]

I. INTRODUCTION

A noticeable part of our daily acoustic environment is made of transient sounds radiated by vibrating structures. These sounds may have pleasant qualities, like those produced by stringed and percussive musical instruments, for example. On the contrary, most of the sounds produced by impacts will likely be considered as undesirable noise.[1,2]

A frequency approach is generally not sufficient for understanding the complexity of transients, since most of the auditory information is often included in the short initial part of the sound. Therefore, a time-domain approach seems more appealing. Another meaningful and attractive aspect of this approach is that the sound resulting from time-domain modeling can be used for musical purposes, in the context of sound synthesis, and for the constitution of systematic series of stimuli for psychoacoustical tests where the goal is to investigate the relationships between physical parameters of structures and subjective judgments.

The geometry of real vibrating structures are often very complex, and it is not always conceivable to model them with great accuracy. However, one efficient strategy consists of approximating structures of complex shapes by structures of simpler geometry, such as plates and shells, in order to grasp the most essential features of the vibratory and acoustical problems. To a first approximation, some properties of panels and of some stringed musical instruments such as the guitar, for example, can be reasonably understood by studying the vibroacoustical behavior of thin plates. The interest in plates is reflected in the considerable amount of reliable results and models published on these structures.

From an auditory point of view, the quality of impacts, and of musical percussive sounds, is highly correlated to the frequency-dependent decay times of the excited eigenfrequencies. A consequence of the dependence on frequency of the decay times is that the power spectrum of the sounds varies with time. The decay times are due to the various mechanisms of loss in the materials. These mechanisms are not identical for all materials: metallic plates will be subjected to thermoelastic damping, whereas viscoelasticity will be one major cause for the attenuation of sound in materials such as glass and carbon fibers, for example. As a consequence, there is a strong need for developing accurate models of these kinds of damping in order to reproduce convincingly the sounds generated by a large variety of materials.

In order to address some of the problems presented above with the help of simulations, a time-domain numerical model of damped isotropic and orthotropic plates excited by an impact has been developed. Previous work by the authors on the comparison between numerical simulations and experiments for an undamped impacted isotropic plate has shown the efficiency of the method.[3]

The paper is organized as follows. In Sec. II the conditions under which three fundamental mechanisms of damping in plates (thermoelasticity, viscoelasticity and radiation) can be expressed by means of a general differential operator are shown. The influence of each of these three mechanisms on the decay times of the eigenfrequencies of the plate is presented. In Sec. III, the procedure for measuring accurately the eigenfrequencies and decay times of the free vibrations, for freely suspended plates, is presented. From these mea-

[a]Present address: ENSTA-UME, Chemin de la Hunière 91761 Palaiseau cedex, France; electronic mail: chaigne@ensta.fr
[b]Present address: 64 rue des Moines, 75017 Paris, France.

surements, the values of the rigidity constants and of the damping constants are derived for four different isotropic and orthotropic materials (aluminum, glass, carbon fibers and wood) which illustrate various combinations of damping mechanisms. For each plate material, the dependence with frequency of the decay factors predicted by the model is compared to measurements. The presentation of the complete model, including initial impact by a mallet and boundary conditions, the numerical resolution of the problem, and the comparison between real and simulated sounds is made in a companion paper.[4]

II. THEORY

This section starts with a short review of the undamped plate equations. This review is followed by the successive presentation of the three mechanisms of damping under study: thermoelasticity, viscoelasticity, and radiation. It is shown, in particular, how the rigidity coefficients of the plate can be modified in order to account for each of these mechanisms. The section ends with the summary of the damped plate equations used in the simulations. The equations are generally written in the orthotropic case. In some situations, however, the simplified equations, which are valid in the isotropic case only, are highlighted.

A. The undamped orthotropic plate equations

In Cartesian coordinates, the free flexural vibrations of an orthotropic Kirchhoff–Love plate of thickness h, with symmetry axes of the material parallel to the sides of the plate in the (x,y) plane, are governed by the following equations:[5,6]

$$\begin{pmatrix} M_x(x,y) \\ M_y(x,y) \\ M_{xy}(x,y) \end{pmatrix} = -h^3 \begin{pmatrix} D_1 & D_2/2 & 0 \\ D_2/2 & D_3 & 0 \\ 0 & 0 & D_4/2 \end{pmatrix} \begin{pmatrix} W_{,xx}(x,y) \\ W_{,yy}(x,y) \\ W_{,xy}(x,y) \end{pmatrix},$$

$$\rho h s^2 W(x,y) = M_{x,xx}(x,y) + M_{y,yy}(x,y) + 2M_{xy,xy}(x,y), \quad (1)$$

where $W(x,y,t)$ is the vertical transverse displacement. The subscripts $,xx$, $,yy$ and $,xy$ denote the partial derivatives. $M_x(x,y)$, $M_y(x,y)$ and $M_{xy}(x,y)$ are the bending and twisting moments, ρ is the density and D_i are the four rigidity constants of the plate. For convenience, the equations are written in the Laplace domain. As a consequence, a partial derivative versus time $\partial/\partial t$ is replaced by a multiplication by s, where s is the Laplace variable. This rule is used throughout this paper.

B. Thermoelasticity

Thermoelastic damping can affect the vibrations in plates with significant thermal conductivity, such as metallic plates. A model for this damping can be derived from the equations that govern the coupling between plate vibrations and the diffusion of heat.[7,8] In the present paper, the method previously used by Cremer for isotropic bars is extended to orthotropic plates, in order to derive the expressions for the complex rigidity constants.[9]

The temperature change θ is assumed to be sufficiently small so that the stress–strain relationships can be linearized. Therefore, we get the stress components:

$$\begin{aligned} \sigma_{xx} &= -12z\left(D_1 W_{,xx} + \frac{D_2}{2} W_{,yy}\right) - \phi_x \theta, \\ \sigma_{yy} &= -12z\left(D_3 W_{,yy} + \frac{D_2}{2} W_{,xx}\right) - \phi_y \theta, \\ \sigma_{xy} &= -6zD_4 W_{,xy}, \end{aligned} \quad (2)$$

where ϕ_x and ϕ_y are the thermal coefficients of the material. In the particular case of an isotropic material, we have $\phi_x = \phi_y = \phi$. Equation (2) has to be complemented by the heat diffusion equation, assuming that θ only depends on z:

$$\kappa \theta_{,zz} - \rho C s \theta = -zT_0 s(\phi_x W_{,xx} + \phi_y W_{,yy}), \quad (3)$$

where T_0 is the absolute temperature, C is the specific heat at constant strain, and κ is the thermal conductivity. Following the method used by Cremer,[9] it is assumed that $\theta(z)$ is given by an equation of the form:

$$\theta(z) = \theta_0 \sin\frac{\pi z}{h} \quad \text{for } z \in \left[-\frac{h}{2}, \frac{h}{2}\right], \quad (4)$$

which takes into account the fact that there is no heat transmission between plate and air. Through integration of $z\sigma_{ij}$ over the thickness h of the plate, one obtains, in the Laplace domain, the relationship between the bending and twisting moments M_{ij} and the partial derivatives of the transverse displacement W of the plate. It is found, finally, that the rigidity factors are now complex functions of the form:

$$\begin{aligned} \tilde{D}_1(s) &= D_1 + \phi_x^2 \frac{s\zeta}{1+\tau s}, \\ \tilde{D}_2(s) &= D_2 + 2\phi_x \phi_y \frac{s\zeta}{1+\tau s}, \\ \tilde{D}_3(s) &= D_3 + \phi_y^2 \frac{s\zeta}{1+\tau s}, \\ \tilde{D}_4(s) &= D_4, \end{aligned} \quad (5)$$

where the thermoelastic relaxation time τ and the thermal constant ζ are given by:

$$\tau = \frac{\rho C h^2}{\kappa \pi^2} \quad \text{and} \quad \zeta = \frac{8T_0 h^2}{\kappa \pi^6}. \quad (6)$$

As a consequence of Eq. (6), the thermoelastic losses decrease as the thickness h of the plate increases. Notice also that D_4 is real. As a consequence, the particular flexural modes of the plate, such as the torsional modes, for example, which involve this rigidity factor will be relatively less affected by thermoelastic damping than the other modes. In other words, the thermoelastic damping factors depend on the modal shapes. These important features of thermoelastic damping will be confirmed by measurements of the decay times in aluminum plates of various sizes (see Sec. III). In what follows, the complex rigidities are written in the form:

$$\tilde{D}_i(s)=D_i[1+\tilde{d}_{it}(s)]=D_i\left[1+\frac{sR_i}{s+c_1/h^2}\right], \quad i=[1,2,3],$$
$$\tilde{D}_4(s)=D_4. \tag{7}$$

In practice, due to the order of magnitude for the thermal constants of usual materials, the $\tilde{d}_{it}(s)$ can be considered as perturbation terms of the complex rigidities, due to thermoelasticity. For convenience, these terms are written in Eq. (7) using the following nondimensional numbers:

$$R_1=\frac{8T_0\phi_x^2}{\pi^4 D_1\rho C};\quad R_2=\frac{16T_0\phi_x\phi_y}{\pi^4 D_3\rho C};\quad R_3=\frac{8T_0\phi_y^2}{\pi^4 D_3\rho C}. \tag{8}$$

In addition, the decay factor $1/\tau$ is written here in the form c_1/h^2 in order to highlight the fact that this factor is proportional to the inverse squared thickness of the plate.

C. Viscoelasticity

A large class of materials is subjected to viscoelastic damping.[10] This is the case for three of the four materials investigated in this study: glass, carbon fiber and wood. A convenient method for representing the viscoelastic phenomena in a material is to use a differential formulation between stress (σ) and strain (ε) tensors of the form:[11]

$$\sigma+\sum_{w=1}^{N} q_w\frac{\partial^w\sigma}{\partial t^w}=E\left[\varepsilon+\sum_{v=1}^{N} q_v\frac{\partial^v\varepsilon}{\partial t^v}\right]. \tag{9}$$

As a consequence, the differential equations involving the flexural displacements and moments in thin plates contain time derivatives up to order N. By taking the Laplace transform of both sides in Eq. (1), one can define the complex rigidities due to viscoelasticity:

$$\tilde{D}_i(s)=D_i(1+\tilde{d}_{iv}(s))=D_i\frac{1+\sum_{v=1}^{N}s^v p_{iv}}{1+\sum_{w=1}^{N}s^w q_w}. \tag{10}$$

Equation (10) is a particular class of representation for the complex rigidities. This operator is bounded under the condition:[12]

$$q_N\neq 0. \tag{11}$$

Another restrictive condition on the coefficients in Eq. (10) follows from the fact that the energy for deforming the material from its initial undisturbed state must be positive. In other words, this means that the model must be a so-called ''dissipative'' model.[13] For a generalized viscoelastic strain–stress relationship the Laplace transform of which is given by:

$$\tilde{\sigma}_{ij}=\sum_{k,l}\tilde{a}_{ijkl}(s)\tilde{\epsilon}_{kl}. \tag{12}$$

Gurtin and Sternberg[12] have proved that one necessary condition for the model to be dissipative is:

$$\Im m(X_{ij}\tilde{a}_{ijkl}(j\omega)X_{kl})\geqslant 0 \quad \text{for} \quad \omega\geqslant 0 \tag{13}$$

for any real symmetric tensor X_{ij}. For our viscoelastic plate model, Eq. (13) together with Eq. (10) yield the following conditions:

$$p_{1N}>0,\quad p_{3N}>0,\quad p_{4N}>0,$$
$$p_{1N}p_{3N}-\frac{D_2p_{2N}^2}{4D_1D_3}>0. \tag{14}$$

In our method, a so-called Wiechert model (also called generalized Maxwell model) is used. It is a convenient formulation of Eq. (10) which fulfills the conditions (14) and ensures that the model is dissipative.[14] This model is written:

$$\tilde{d}_{iv}(s)=\sum_{n=1}^{N}\frac{sR_{in}}{s+s_n}. \tag{15}$$

As for thermoelastic damping, the viscoelastic losses in the materials considered here are relatively small and thus, the corresponding terms can be viewed as first-order correction terms of the rigidity terms.

One property of Eq. (15) is that the loss factor tends to zero as ω tends to zero. Therefore, in order to account for the fact that experimental decay factors usually tend to a constant (positive) value as ω vanishes, it is necessary to introduce, in addition, an arbitrary viscous damping term proportional to velocity $-R_f(\partial w/\partial t)$ in the bending wave equation [Eq. (1)]. The values obtained for R_f are the result of least-square fit procedures performed on the measured decay factors (see Sec. III). It has been found that introducing this term reduces significantly the number N of factors in the viscoelastic model, for a given degree of accuracy between the model and the experimental data. In practice, it turns out that, for the presently investigated materials, and for the accuracy required, the number N of factors in Eq. (15) remains less than or equal to 2.

D. Damping due to radiation (isotropic plates)

The interaction of vibrating plates with the surrounding air is a fundamental problem in vibroacoustics.[15–17] In this section, the goal is to develop an approximate model in the time-domain which takes into account the effect of radiation on the temporal decay of the free bending vibrations for isotropic plates.

The leading idea is to derive a differential operator for the radiation losses (i.e., a polynomial formulation in the Laplace domain), similar to the one presented in the previous sections for thermoelastic and viscoelastic damping. This should allow us to group all three mechanisms of damping for the free vibrations of plates into a single unified formulation.

The experiments described here, where the plates are freely suspended by light threads, show that in the ''low-frequency'' domain (i.e., below the critical frequency), viscoelastic and/or thermoelastic losses are the main causes of damping (see Sec. III for the four materials under investigation). In this particular situation, there is no damping due to coupling at the boundaries. Above the critical frequency, the damping is mainly governed by radiation for low dissipative materials, such as aluminum and glass. In what follows, it is assumed that the radiation losses become relevant when the wavelength is significantly lower than the dimensions of the plate. Therefore, an asymptotic infinite plate model is used, which simplifies substantially the formulation of the radia-

tion damping. In what follows, no attempt is made to model accurately the mode-by-mode radiation of the finite plate below the critical frequency.

The equations governing the flexural motion of an infinite isotropic plate surrounded by air are the following:[18]

$$\rho h \frac{\partial^2 W(x,y,t)}{\partial t^2} + h^3 D\left(\frac{\partial^4}{\partial x^4} + 2\frac{\partial^4}{\partial x^2 \partial y^2} + \frac{\partial^4}{\partial y^4} \right) W(x,y,t) = f_z(x,y,t) + p(x,y,0-,t) - p(x,y,0+,t), \tag{16}$$

$$\left. \frac{\partial p}{\partial z} \right|_{z=0+} = -\rho_a \frac{\partial^2 W}{\partial t^2}, \tag{17}$$

$$\Delta p - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = 0, \quad \text{for } z>0 \text{ and } z<0, \tag{18}$$

where $D=D_1=D_3=(D_2+D_4)/2$. For a traveling wave of the form $W(x,y)\exp j(\tilde{\omega}t-\underline{k}.\underline{x})$ where k is the wave number and $\tilde{\omega}=\omega+j\alpha_r$ the complex frequency, one can derive from Eqs. (16)–(18) the dispersion equation for the fluid-loaded plate:[18]

$$-\tilde{\omega}^2\left(1+\frac{2\rho_a}{\rho h}\frac{1}{\sqrt{k^2-\frac{\tilde{\omega}^2}{c^2}}} \right) + \frac{Dh^2}{\rho}k^4 = 0. \tag{19}$$

Under the assumption of a light fluid, one can equivalently derive the radiation decay factor α_r from Eq. (19) by reformulating this equation through introduction of a rigidity modulus $\tilde{D}$ of the form:

$$\tilde{D} \simeq D\left(1-\frac{2\rho_a c}{\omega_c \rho h}\frac{1}{\sqrt{\Omega-\Omega^2}} \right), \tag{20}$$

with

$$\omega_c = c^2\sqrt{\frac{\rho}{h^2 D}} \quad \text{and} \quad \Omega = \frac{\omega}{\omega_c}. \tag{21}$$

Now, in order to find an appropriate form for Eq. (20), the complex rigidity $\tilde{D}$ is rewritten in the form of a third order Padé development. The optimization procedure is performed for $s=j\omega$ in the audio range. In the time domain, this yields:

$$\tilde{D}(s) \hat{=} D\left(1+\frac{2\rho_a c}{\omega_c \rho h}\frac{\Sigma_{m=1}^{3} b_m (s/\omega_c)^m}{\Sigma_{n=0}^{3} a_n (s/\omega_c)^n} \right) = D(1+\tilde{d}_r(s)), \tag{22}$$

with $a_0=1.1669$, $a_1=1.6574$, $a_2=1.5528$, $a_3=1$, $b_1=0.0620$, $b_2=0.5950$, $b_3=1.0272$. For orthotropic plates, it is known that Eq. (19) depends on the direction of propagation of the flexural wave in the plate. As a consequence, it was not possible to formulate the radiation damping in the time-domain using an approximation similar to Eq. (22). Therefore, the two examples below which show a relevant contribution of radiation losses in the decay factors, will be limited to isotropic materials: glass and aluminum. For the two examples of orthotropic materials (wood and carbon fiber) presented in Sec. III, the radiation losses are generally masked by the relatively high internal (viscoelastic) losses. In these cases, the discussion on the radiation damping will be restricted to limiting situations, for waves propagating along the axes of symmetry of the material.

E. Summary of the damping model

It is assumed that the three previously examined physical mechanisms of damping are uncorrelated. It is also assumed that the perturbation term $\tilde{d}_i(s)$, for each complex rigidity, is given by the sum of three perturbation terms, each of them corresponding to one specific mechanism. Therefore, in the most general case, the four complex rigidities in our model are given by:

$$\tilde{D}_i(s) = D_i(1+\tilde{d}_i(s)) = D_i(1+\tilde{d}_{it}(s)+\tilde{d}_{iv}(s)+\tilde{d}_r(s)), \quad i=1,..,4, \tag{23}$$

where $\tilde{d}_{it}$ refers to thermoelastic damping, $\tilde{d}_{iv}$ refers to viscoelastic damping and $\tilde{d}_r$ refers to radiation damping. It is clear that the relative contribution of each perturbation term in the measured global damping varies with the material. In some specific cases, one or two terms will become negligible, as will be shown in Sec. III.

In summary, the system of equations used in our model for simulating damped plates is similar in form to Eq. (1), except that each rigidity constant D_i is now replaced by a complex rigidity $\tilde{D}_i(s)$, which represents a differential operator. In addition, a viscous damping term with constant R_f is inserted in the bending wave equation:

$$\begin{pmatrix} \tilde{M}_x(x,y,s) \\ \tilde{M}_y(x,y,s) \\ \tilde{M}_{xy}(x,y,s) \end{pmatrix} = -h^3 \begin{pmatrix} \tilde{D}_1(s) & \tilde{D}_2(s)/2 & 0 \\ \tilde{D}_2(s)/2 & \tilde{D}_3(s) & 0 \\ 0 & 0 & \tilde{D}_4(s)/2 \end{pmatrix} \times \begin{pmatrix} \tilde{W}_{,xx}(x,y,s) \\ \tilde{W}_{,yy}(x,y,s) \\ \tilde{W}_{,xy}(x,y,s) \end{pmatrix},$$

$$\rho h(s^2+R_f s)\tilde{W}(x,y,s) = \tilde{M}_{x,xx}(x,y,s) + \tilde{M}_{y,yy}(x,y,s) + 2\tilde{M}_{xy,xy}(x,y,s) + \tilde{f}_z(x,y,s). \tag{24}$$

In Eq. (24), $\tilde{f}_z(x,y,s)$ represents the source term due to impact which is presented with more details in the companion paper.[4]

III. MEASUREMENTS

A. Experimental set-up

Measurements of eigenfrequencies and decay times are performed on rectangular plates in an anechoic room. The plates are suspended vertically by means of a nylon thread in order to avoid transmission losses. In a first series of experiments, the plate is set into vibration by means of an impact hammer. An FFT (fast Fourier transform) frequency analysis is then performed, with a B&K 2032 frequency analyzer, on the pressure signal recorded at a distance of approximately 1 m from the plate. This analysis yields a first estimate for the eigenfrequencies of the freely vibrating plate.

In a second series of experiments, the plate is acoustically excited by means of a loudspeaker driven by a Philips PM 5193 programmable frequency synthesizer. The frequency of the synthesizer is initially set equal to one of the previously estimated eigenfrequencies of the plate. The transverse velocity of one arbitrary point on the plate is measured by means of a Polytec OFV 2600-352 laser vibrometer. This procedure ensures noncontact excitation and measurements. The frequency is adjusted until the magnitude of the plate velocity reaches its maximum. This yields an accurate estimation of the eigenfrequency. The acoustic source is then suddenly turned off, and the velocity signal delivered by the vibrometer is recorded on a DAT tape. This signal is analyzed using the matrix pencil method.[19] In this parametric method, the signal is assumed to be composed of a finite sum of exponentially decaying sinusoids, whose frequencies, decay times, magnitude, and phases are estimated. In our case, this signal processing technique yields the decay times of the successive eigenfrequencies of the plate with great accuracy.

In case of high damping, the acoustic power of the sound source is not sufficient for exciting flexural vibrations in the plate with enough magnitude. This is the case, for example, in this study for both the carbon fiber and wooden plates, for frequencies above 500 Hz. In this frequency range, an impact hammer excitation had to be used.

B. Link between measured decay times and the damping model

Within the framework of the model, and with the assumption of ''small damping,'' the four loss factors η_i predicted by the model are given by:[20]

$$\eta_i(\omega) \simeq \Im m(\tilde{D}_i(j\omega))/D_i = \Im m(\tilde{d}_i(j\omega)), \quad i=1,..,4. \tag{25}$$

Notice that in Eq. (25) the loss factors are frequency-dependent. As mentioned above, the available measured quantities, for each plate, are the eigenfrequencies ω_p and the decay factors α_p. These decay factors are the inverse of the decay times τ_p, which are measured as the plate vibrates freely with frequency ω_p. The link between the measured α_p and the loss factors defined in Eq. (25) are now examined. The simple case of proportional damping is discussed first.

In a limited number of cases, it is justified to make the assumption of proportional damping.[21] With the additional assumption of small damping, it can be shown that the loss factors and the decay factors, for a given mode p are linked by the expression:[22]

$$\alpha_p(\omega_p) = \frac{\omega_p}{2}\,\eta(\omega_p). \tag{26}$$

This is the case, for example, for glass plates (see below).

In the general case, however, this assumption of proportional damping is too restrictive and does not account for the observed complicated frequency dependence of the decay factors. In this case, it can be shown that the decay factors, for an orthotropic plate, are given by:[20]

$$\alpha_p(\omega_p) = \sum_{i=1}^{4} \frac{\omega_p}{2}\,\eta_i(\omega_p) J_{i,p} \tag{27}$$

with

$$J_{i,p} = \frac{a_i(u_p,u_p)}{\omega_p^2 \|u_p\|^2}, \tag{28}$$

where u_p is the eigenvector of mode p in the undamped case, and where the scalar products $a_i(u_p,u_p)$ are defined by:[20]

$$\begin{aligned}
a_1(u_p,u_p) &= \int_S h^3 D_1 \left(\frac{\partial^2 u_p}{\partial x^2}\right)^2 dS,\\
a_2(u_p,u_p) &= \int_S h^3 D_2 \frac{\partial^2 u_p}{\partial x^2}\frac{\partial^2 u_p}{\partial y^2} dS,\\
a_3(u_p,u_p) &= \int_S h^3 D_3 \left(\frac{\partial^2 u_p}{\partial y^2}\right)^2 dS,\\
a_4(u_p,u_p) &= \int_S h^3 D_4 \left(\frac{\partial^2 u_p}{\partial x \partial y}\right)^2 dS.
\end{aligned} \tag{29}$$

In Eqs. (29), S is the surface of the plate of thickness h and density ρ and $\|u_p\|^2$ is the norm defined as:

$$\|u_p\|^2 = \omega^2 \int_S \rho h u_p^2\, dS. \tag{30}$$

In the isotropic case, Eq. (27) reduces to

$$\alpha_p(\omega_p) = \frac{\omega_p}{2}[\,\eta_1(\omega_p) I_{1,p} + \eta_4(\omega_p) I_{4,p}], \tag{31}$$

where

$$I_{1,p} = J_{1,p} + J_{3,p} + \frac{2D_1}{D_2} J_{2,p} \quad \text{and} \quad I_{4,p} = J_{4,p} - \frac{D_4}{D_1} J_{2,p}. \tag{32}$$

C. Determination of the rigidity constants

1. First estimates of the rigidities

The method used here is largely inspired by the work of McIntyre and Woodhouse.[20] Using Rayleigh's method with polynomial functions as trial functions, a first estimate for the rigidity factor D_4 is obtained from measurement of the first torsional (or twisting) mode f_{11}:

$$D_4 = f_{11}^2 \frac{4\pi^2}{144} \rho \frac{(l_x l_y)^2}{h^2}. \tag{33}$$

The rigidity constants D_1 and D_3 are estimated from measurements of the first flexural modes f_2 and f_3 in the x and y directions, respectively. Assuming that these modes can be approximately predicted from the free–free bar equations, we get:[20]

$$D_1 \simeq 0.0789 \frac{\rho f_2^2 l_x^4}{h^2} \quad \text{and} \quad D_3 \simeq 0.0789 \frac{\rho f_3^2 l_y^4}{h^2}. \tag{34}$$

Finally, a first estimate for D_2 is obtained by considering that, for orthotropic plates, we have the relation:

$$D_2 = 2\sqrt{\nu_{xy}\nu_{yx} D_1 D_3}. \tag{35}$$

Setting the product $\nu_{xy}\nu_{yx}$ to an arbitrary, though realistic, value of 0.01, we get $D_2 \simeq 0.2\sqrt{D_1 D_3}$ as a starting value. In the isotropic case, we simply have:

$$D_1 = D_3 \quad \text{and} \quad D_2 = 2D_1 - D_4. \tag{36}$$

2. Accurate calculation of the rigidity constants

The next step consists of calculating the eigenfrequencies ω_p of the plate from the first estimates of the rigidity constants D_i. This calculation is made by means of the classical Rayleigh–Ritz method, using polynomial trial functions.[21] The procedure is typically applied to the lowest 20 modes of the plate. Correcting terms ΔD_i are then obtained using a first-order approximation:

$$\Delta\omega_p^2 = \sum_{i=1}^{4} \frac{\partial(\omega_p^2)}{\partial D_i}\Delta D_i \quad \text{with} \quad \frac{\partial(\omega_p^2)}{\partial D_i} = J_{i,p}\frac{\omega_p^2}{D_i}, \tag{37}$$

where the $\Delta\omega_p^2$ are the differences between the measured and calculated squared frequencies. The procedure is iterated until the $\Delta\omega_p^2$ are less than a fixed accuracy limit. In practice, a maximum of 3 iterations is necessary for the required accuracy in audio applications (the order of magnitude fixed here is about 1%).

D. Examples of results on four different materials

In order to illustrate the relevance of the above-presented model of damping, the eigenfrequencies and decay times of the free vibrations for rectangular plates made of four different materials are measured and compared to the predictions of the model. For each material, the contributions are specific. For the aluminum plate, it turns out that the damping is mainly governed by thermoelastic and radiation mechanisms. For the three other plates, the damping is due to viscoelasticity and radiation. In the thermoelastic and radiation cases, the damping parameters in Eqs. (7) and (22), respectively, can be compared to predictions based on the physical properties of the material. In the viscoelastic case, there is no similar theoretical background: the model and the damping constants are the results of a least-square fit performed on the experimental data. From the observation of the results obtained for the frequency-dependent decay factors, which generally highlight the most relevant damping mechanisms for each material, damping models are built, with the help of standard optimization procedures, and are tested against the experimental data. In the figures, the experimental points are indicated by ''○,'' whereas the theoretical points are indicated by ''×.'' In some cases (radiation, viscoelasticity for the glass plate), the theoretical model is represented by a solid line.

1. Aluminum

Measurements were conducted on two different plates (see Table I). The first 30 modes were measured by means of acoustic excitation and optical measurement, with the help of a laser vibrometer, in an anechoic room, following the previously described procedure.

Figures 1(a) and (b) show the decay factors (in s^{-1}) as a function of frequency between 0 and 2 kHz, for plates a_1 and a_3, respectively. These figures show that the frequency dependence of the decay factors is far from being monotonic, but rather show an apparently random distribution of values, however with a general tendency to increase with frequency. This apparent randomness is more pronounced for the thinner plate a_1 than for the thicker one a_3. In addition, it can be seen that the mean value of the decay factor is about four times greater for plate a_1 than for plate a_3. This is an indication that the damping in this frequency range might be due to thermoelasticity, see Sec. II, since the decay factors vary as the inverse of squared thickness.

Figures 1(c) and (d) show the variation of the decay factors with frequency, up to 12 kHz, for the same two plates. These figures show clearly that the damping increases strongly above 6 kHz (for the thin plate a_1), and above 3 kHz (for the thick plate a_3), which correspond to the critical frequencies for these two plates. The solid lines superimposed on the experimental points are obtained from the asymptotic model of radiation for isotropic plates developed in Sec. II. This comparison shows that, for these metallic plates, the radiation clearly becomes the dominant damping mechanism above the critical frequency. In Figs. 1(a) and (b) it can be seen that, in the frequency range where thermoelasticity damping is the main cause of damping, the radiation model is significantly lower than the measured decay factors and, consequently, will not affect the prediction of damping in this range.

In order to build a model for these plates, the rigidity constants D_1 and D_4 are derived from the measurements of the 20 first modes, using the method described above. The iteration process is repeated until the average relative frequency error between measurements and simulations is less than 0.5%. The values obtained for D_1 and D_4 are listed in Table I.

For the damping model, in view of the previously described measurements, it is assumed that the perturbation terms in the complex rigidities are only due to thermoelasticity and to radiation. As mentioned in Sec. II, the thermoelastic term only affects $\tilde{D}_1$. Therefore, we write:

$$\tilde{d}_1(s) = \tilde{d}_{1t}(s) + \tilde{d}_r(s) \quad \text{and} \quad \tilde{d}_4(s) = \tilde{d}_r(s). \tag{38}$$

As mentioned in Sec. II, a viscous damping term with constant R_f is added to the bending wave equation, in order to account for the finite damping near the zero frequency. Finally, from Eq. (38), the following decay factors, for each mode p, are derived:

$$\alpha_p = \frac{\omega_p}{2}\Im m(\tilde{d}_{1t}(j\omega_p)I_{1,p} + \tilde{d}_r(j\omega_p)) + \frac{R_f}{2}. \tag{39}$$

Figures 1(a)–(d) show the agreement between the measured decay factors and the simulations obtained with the damping parameters shown in Table I. It is found, in particular, that the thermoelastic constants derived from least-square optimization procedures are of the same order of magnitude as the general constants found in the literature for aluminum. For frequencies below 2 kHz (for plate a_1), or below 1.2 kHz (for plate a_3), it can be seen that the model accounts for the apparent erratic law of dependence of damping on frequency. For frequencies above the critical frequencies (6 kHz for a_1, 3 kHz for a_3), the asymptotic radiation model

TABLE I. Measured elastic and geometrical parameters of the plates.

Aluminum plates				
$\rho = 2660\ \mathrm{kg\,m^{-3}}$	$D_1 = 6160$ MPa	$D_4 = 8600$ MPa		
plate a_1:	$l_x = 304$ mm	$l_y = 192$ mm	$h = 2$ mm	$f_c = 6$ kHz
plate a_3:	$l_x = 419.5$ mm	$l_y = 400$ mm	$h = 4$ mm	$f_c = 3$ kHz
Glass plate				
$\rho = 2550\ \mathrm{kg\,m^{-3}}$	$D_1 = 6700$ MPa	$D_4 = 10\,270$ MPa		
plate v_1:	$l_x = 229.5$ mm	$l_y = 220.5$ mm	$h = 2$ mm	$f_c = 5.8$ kHz
Carbon plate				
$\rho = 1540\ \mathrm{kg\,m^{-3}}$	$D_1 = 8437$ MPa	$D_2 = 463$ MPa	$D_3 = 852$ MPa	$D_4 = 2267$ MPa
plate c_1:	$l_x = 399$ mm	$l_y = 200$ mm	$h = 2.2$ mm	critical zone=[3.6,11.3] kHz
Wooden plate				
$\rho = 388\ \mathrm{kg\,m^{-3}}$	$D_1 = 1013$ MPa	$D_2 = 27.5$ MPa	$D_3 = 53.7$ MPa	$D_4 = 221$ MPa
plate b_1:	$l_x = 515$ mm	$l_y = 412.5$ mm	$h = 4.8$ mm	critical zone=[2.4,10.3] kHz

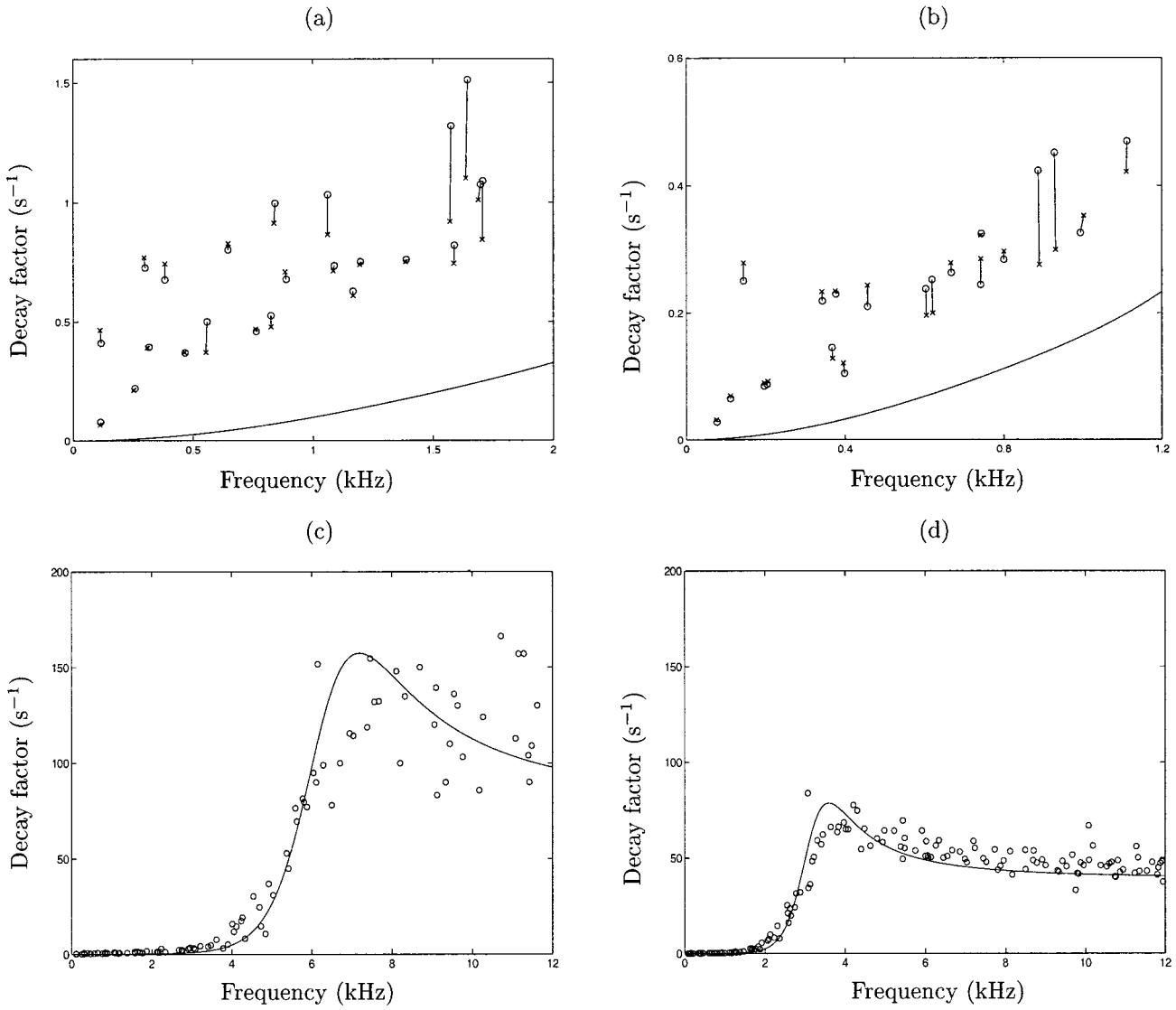


FIG. 1. Aluminum plates. Comparison between predicted and measured decay factors (in $\mathrm{s^{-1}}$) vs frequency. ''○'' measured points; ''×'' predicted points with thermoelastic damping model. Solid line: asymptotic radiation model. (a) and (c) plate a_1 ($h = 2$ mm). (b) and (d) plate a_3 ($h = 4$ mm). (a) and (b) 0–2 kHz. (c) and (d) 0–12 kHz.

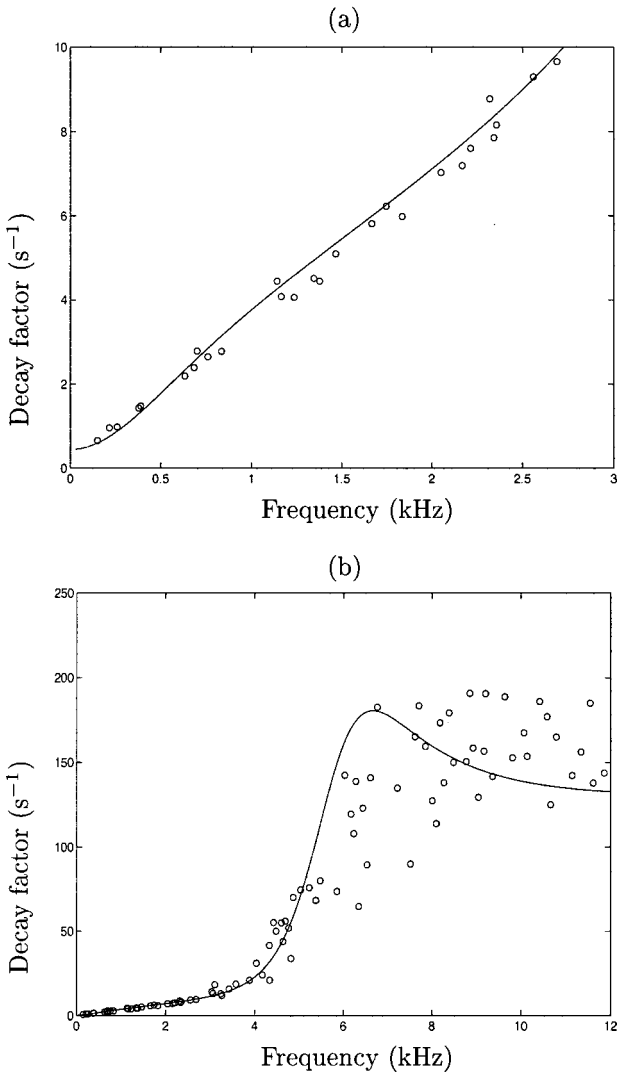


FIG. 2. Glass plate. Comparison between predicted and measured decay factors (in s^{-1}) vs frequency. (a) Solid line: viscoelastic model; ''○'' measured points. (b) Solid line: complete model (viscoelasticity+radiation); ''○'' measured points.

based on the dispersion equation for the infinite plate accounts globally for the frequency dependence of decay factors measured on finite plates. It can be seen that the standard deviation of the experimental points increases as the thickness of the plate decreases.

2. Glass

Figure 2(a) shows the decay factors measured on the rectangular glass plate for $f<3$ kHz, i.e., below its critical frequency (5.8 kHz). The results show an almost linear relationship between the decay factors and the eigenfrequencies, which completely differs from the results obtained in the case of aluminum. In addition, the decay factors are nearly ten times greater than those obtained for an aluminum plate of comparable size and thickness (plate a_1). Comparing these experimental results to Eq. (26) shows that the assumption of proportional damping is pertinent here. As a consequence, the decay factors are independent of the modal shapes of the plate. Therefore, we can write in this range:

$$\alpha_p=\frac{\omega_p}{2}\,\eta(\omega_p) \quad \text{with} \quad \eta(\omega_p)=\eta_1(\omega_p)=\eta_4(\omega_p). \tag{40}$$

Above the critical frequency [see Fig. 2(b)], the measured decay factors for the glass plate are comparable to those of the aluminum plate. Here again, the asymptotic radiation model fits well with the average value of the experimental points: a noticeable increase around the critical frequency followed by an almost constant value ($150\,s^{-1}$). The standard deviation in this range is of the same order of magnitude as that observed for aluminum plate a_1.

Below the critical frequency, the damping model for the glass plate is built with a two-cell viscoelastic model which ensures that the loss factors η_1 and η_4 are almost constant below the critical frequency:

$$\tilde{d}_{iv}(s)=\tilde{d}_v(s)=\frac{sR_1}{s+s_1}+\frac{sR_2}{s+s_2}, \quad i=1,4. \tag{41}$$

The efficiency of this viscoelastic model is shown in Fig. 2(a). The decay factors are reproduced within less than 5%. Above the critical frequency, the radiation is modeled with the perturbation term $\tilde{d}_r(s)$ given by Eq. (22). Finally, a viscous damping R_f is added, as with aluminum plates, in order to account for the damping near the zero frequency. As for the aluminum plates, it is assumed that the two mechanisms of damping are uncorrelated, and that the two perturbation terms, viscoelasticity and radiation, are added in the complex rigidities. In summary, the decay factors predicted by the model used for the glass plate are given by:

$$\alpha_p=\frac{\omega_p}{2}\,\Im m(\tilde{d}_v(j\omega_p)+\tilde{d}_r(j\omega_p))+\frac{R_f}{2}. \tag{42}$$

3. Carbon fibers

The rectangular plate made of carbon fibers used for the measurements is orthotropic, with axes of symmetry parallel to the edges. For this plate, measurements of eigenfrequencies and decay rates, used for the estimation of the damping constants and reported here, are limited to the range [0,2] kHz. The reason is that this material is highly dissipative, compared to glass and aluminum, for example. As a consequence, the bending vibrations are damped very rapidly and it is difficult to measure the decay times above 2 kHz with great accuracy. In order to allow comparisons with previous studies on such materials, the results are presented here in terms of quality factors. Figure 3 shows the dependence of quality factors on frequency for the carbon plate. The majority of the modes exhibit constant Q values around 250, although a substantial number of modes show an apparent erratic distribution of Q values, as it was the case for the aluminum plates below the critical frequency. Due to low thermal conductivity, the thermoelastic losses are negligible in such a material, and it is reasonable to assume here that the internal losses are almost entirely due to viscoelasticity.

With the assumption of a proportional damping, it was not possible here to predict the decay factors with sufficient accuracy. Therefore, in order to account for the measured decay factors, it is assumed here that the complex rigidities $\tilde{D}_i$ take the form:

$$\tilde{d}_{iv}(s)=\frac{sR_{1i}}{s+s_{1i}}+\frac{sR_{2i}}{s+s_{2i}},\quad i=(1,3,4)\ \text{ and }\ d_{2v}(s)=0, \tag{43}$$

where three of the four rigidities are perturbed by a specific two-cell viscoelastic function. As for the other materials, a viscous damping is added to the bending wave equation. Thus, the predicted decay factors are given by:

$$\alpha_p(\omega_p)=\frac{\omega_p}{2}\Im m(\tilde{d}_{1v}(j\omega_p)J_{1,p}+\tilde{d}_{3v}(j\omega_p)J_{3,p}$$

$$+\tilde{d}_{4v}(j\omega_p)J_{4,p})+\frac{R_f}{2}. \tag{44}$$

Since D_2, and thus $J_{2,p}$, are relatively small compared to the other terms in Eq. (44), there is no need for adding a perturbation term to D_2, since it has no influence on the decay factors. In terms of quality factors, Eq. (44) becomes:

$$Q_p=\pi\frac{f_p}{\alpha_p}. \tag{45}$$

Unlike isotropic materials, it was not possible to derive a satisfactory approximation for the radiation damping of orthotropic plates, because of the angular dependence of the dispersion equation (see Sec. II). In some fortunate cases, the material and the geometry of the plate are such that the radiation damping is masked by other mechanisms and can be neglected. This is the case, in our study, for the carbon plate and, to a lesser extent, for the wooden plate. Otherwise, we must content ourselves with a simplified approach to the radiation modeling for orthotropic plates by considering two limiting cases: a wave traveling either in the x or in the y direction. In these cases, an upper bound f_{c3} and a lower bound f_{c1} for the critical frequency can be found by replacing D in Eq. (21) by D_3 and D_1, respectively. This yields the limits of a ''critical zone.'' Using Eq. (22), one can then derive two asymptotic curves for predicting the contribution of radiation damping in the decay factors following a procedure similar to the isotropic case. For the carbon plate under study c_1 (see Table I), for example, this zone is given by the interval [3.6,11.3] kHz. This means qualitatively that the radiation losses are higher for the eigenmodes that are spatially oriented along the x-axis.

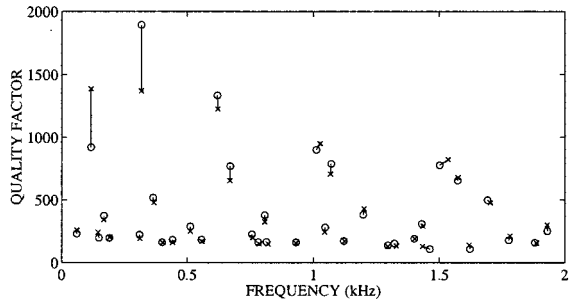


FIG. 3. Carbon plate. Comparison between predicted and measured quality factors vs frequency. ''○'' measured points; ''×'' predicted points with viscoelastic damping model. The solid lines indicate the distance between measured and predicted values.

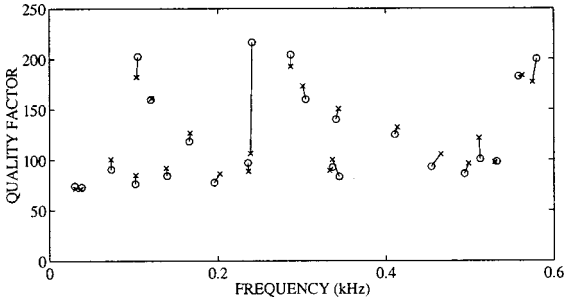


FIG. 4. Wooden plate (Spruce). Comparison between predicted and measured quality factors vs frequency. ''○'' measured points; ''×'' predicted points with viscoelastic damping model. The solid lines indicate the distance between measured and predicted values.

Figure 3 shows the comparison between measured and calculated quality factors for plate c_1 using Eq. (44), which amounts to assuming that the damping is only due to viscoelasticity. It can be seen that the model is able to predict most of the quality factors with great accuracy in the interval [0,2] kHz, except for a limited number of low-frequency modes with high Q.

4. Wood (Spruce)

The fourth measured plate b_1 is made of Spruce. This rectangular plate is cut along the Longitudinal-Radial plane (LR plane or quarter-cut).[23] This plate was selected and prepared by an instrument maker for designing a guitar soundboard. Compared to the other measured orthotropic plate in carbon c_1, the wooden plate exhibits a more pronounced degree of orthotropy $D_3/D_1\simeq 20$ (see Table I). A modal analysis performed on this plate showed some distinct asymmetry of some modal shapes which might be the consequence of heterogeneities in the material. Measurements of the thickness at different points of the plate showed relative differences of nearly 5% which also contributes to the inhomogeneous properties of the plate.

The flexural vibrations in the material are damped very rapidly, which is immediately perceptible by tapping the plate with the finger, as an instrument maker usually does for testing them. Due to the small decay times, measurements were only possible below 3 kHz. Accurate measurements were obtained for the first 23 eigenfrequencies (see Fig. 4). These measurements were used for estimating the elastic and damping constants (see Tables I and II).

For this plate, the structure of the viscoelastic model is the same as the one given in Eq. (43) for the carbon plate c_1. Consequently, the prediction of the decay factors is governed by an equation similar to Eq. (44). As for the carbon plate, the results shown in Fig. 4 are presented in terms of Q factors, using an equation similar to Eq. (45). This figure shows that the quality factor for this material is almost constant and nearly equal to 80 for a large number of modes in the observed range, with again an apparent erratic distribution for the other modes. There is excellent agreement between measured and predicted Q values, except for one mode around 230 Hz.

Here also, the radiation damping is neglected, due to the high rate of energy dissipation in the material. However, this

TABLE II. Measured damping parameters of the plates.

Thermoelastic damping (aluminum)			
$R_1 = 8.45 \times 10^{-3}$	$c_1 = 8.0 \times 10^{-4}$ rad m^2 s^{-1}		
Viscoelastic damping (glass)			
$R_1 = 1.63 \times 10^{-3}$	$s_1 = 5180$ rad s^{-1}	$R_2 = 1.962 \times 10^{-3}$	$s_2 = 55\,100$ rad s^{-1}
Viscoelastic damping (carbon)			
$R_{11} = 1.32 \times 10^{-3}$	$s_{11} = 10.1 \times 10^3$ rad s^{-1}	$R_{12} = 5.0 \times 10^{-3}$	$s_{12} = 94.0 \times 10^3$ rad s^{-1}
$R_{31} = 8.8 \times 10^{-3}$	$s_{31} = 2.5 \times 10^3$ rad s^{-1}	$R_{32} = 44.0 \times 10^{-3}$	$s_{32} = 70.0$ rad s^{-1}
$R_{41} = 10.4 \times 10^{-3}$	$s_{41} = 2.27 \times 10^3$ rad s^{-1}	$R_{42} = 14.4 \times 10^{-3}$	$s_{42} = 40.0 \times 10^3$ rad s^{-1}
Viscoelastic damping (wood)			
$R_{11} = 8.18 \times 10^{-3}$	$s_{11} = 3.2 \times 10^3$ rad s^{-1}	$R_{12} = 10.0 \times 10^{-3}$	$s_{12} = 50.2 \times 10^3$ rad s^{-1}
$R_{31} = 16.7 \times 10^{-3}$	$s_{31} = 1.1 \times 10^3$ rad s^{-1}	$R_{32} = 70.0 \times 10^{-3}$	$s_{32} = 50.2$ rad s^{-1}
$R_{41} = 15.2 \times 10^{-3}$	$s_{41} = 1.75 \times 10^3$ rad s^{-1}	$R_{42} = 35.0 \times 10^{-3}$	$s_{42} = 50.2 \times 10^3$ rad s^{-1}
Viscous damping			
aluminum $R_f = 0.032$ s^{-1}	glass $R_f = 0.88$ s^{-1}	carbon $R_f = 0.8$ s^{-1}	wood $R_f = 2.4$ s^{-1}

assumption is less justified here than for the carbon plate c_1 because of both the relatively larger size and lower density of the wooden plate. Schematically, it is found that the critical frequency for flexural waves propagating along the x-axis is here equal to 2.4 kHz, which means that the radiation damping will become significant for frequencies near and above this value. In comparison, the radiation damping for plate c_1 was found to be significant for frequencies greater than or equal to 3.6 kHz. The consequence of this is that the global damping in the model will be underestimated for frequencies above 2.4 kHz, as it will be seen in the simulations.[4] On the contrary, adding a simplified radiation contribution using an ''equivalent isotropic'' model by means of a perturbation term similar to the one shown in Eq. (22) yields an overestimation of the global damping. In the case of the present wooden plate, it has been shown that these approximations in the radiation damping do not affect substantially the estimation of the quality factors, since the radiation effects are masked by the intrinsic damping of the material.

IV. CONCLUSIONS

The aim of the work reported in this paper was to derive an appropriate model for the time-domain simulation of damped flexural vibrations in plates. The main interest of this work is to investigate the link between tone qualities and material properties by means of sound synthesis. This goal is not only of prime importance for the design of musical instruments, but also for a better understanding and control of all sources of noise, especially those resulting from impacts.

In the literature, one can find many papers dealing with the frequency domain modeling of damping in structural dynamics.[24] In comparison, relatively few papers address this problem in the time-domain. Therefore, it was decided to put emphasis here on the damping model. In musical acoustics, previous work on strings, bars, and membranes has shown the importance of accurate damping parameters in the realism of the simulated sounds.[25,26]

The strategy selected here was to take three basic mechanisms of damping in plates into account by adding perturbation terms in the rigidities. The advantage of this method is to yield a simplified system that governs the vibrations of the plate itself, without any coupling to acoustical and thermal fields. From these equations, the eigenfrequencies and decay rates of the plates can be estimated. The elastic plate model is based on the Kirchhoff–Love approximations, and is treated for a simple rectangular geometry.

The method has been validated on two isotropic (glass and aluminum) and two orthotropic (carbon fiber and wood) materials. The damping phenomena in each of these materials are, to some extent, representative of the three investigated mechanisms: thermoelasticity and radiation in aluminum, viscoelasticity and radiation in glass, viscoelasticity in carbon and wood. For each of these materials, the comparison between measured and predicted eigenfrequencies and decay factors show that the model is able to reproduce fairly well the complicated frequency dependence of the decay factors, and the influence of the plate geometry (size, thickness) on the decay times. However, the simplified asymptotic radiation model is only valid for isotropic materials and is not relevant enough for orthotropic materials. For these latter materials, it is likely that it will be necessary to model the complete fluid-structure interaction in order to predict the decay rates with sufficient accuracy, as has been done for other vibrating structures.[26]

The next step of the work will now consist of the numerical formulation of the problem in order to simulate waveforms in the time domain. In order to facilitate the future finite-difference formulation of the damped plate equations, a convenient differential operator was selected for the damping terms. Attention was paid, in particular, to the mathematical properties of this formulation, in order to ensure the well-posedness of the problem in terms of dissipativity. The numerical model and the results are presented in a companion paper.[4]

[1] T. Igarashi, M. Goto, and A. Kawasaki, ''Studies on impact sound (First report. The sound generated by a ball colliding a plate),'' Bull. JSME **28**, 148–154 (1985).
[2] T. Igarashi and T. Aimoto, ''Studies on impact sound (Second report. Mechanism of sound generation),'' Bull. JSME **28**, 1247–1254 (1985).
[3] S. Schedin, C. Lambourg, and A. Chaigne, ''Transient sound fields from impacted plates: Comparison between numerical simulations and experiments,'' J. Sound Vib. **221**, 471–490 (1999).

[4] C. Lambourg, A. Chaigne, and D. Matignon, ''Time-domain simulation of damped impacted plates. Part II. Numerical model and results,'' J. Acoust. Soc. Am. **109**, 1433–1447 (2001).
[5] A. W. Leissa, *Vibrations of Plates*, NASA SP-160 (NASA, Washington, D.C., 1969).
[6] R. F. S. Hearmon, *An Introduction to Applied Anisotropic Elasticity* (Oxford University Press, London, 1961).
[7] M. A. Biot, ''Thermoelasticity and irreversible thermodynamics,'' J. Appl. Phys. **27**, 240–253 (1956).
[8] C. Zener, ''Internal friction in solids—I. Theory of internal friction in reeds,'' Phys. Rev. **32**, 230–235 (1937).
[9] L. Cremer, M. Heckl, and E. E. Ungar, *Structure-Borne Sound*, 2nd ed. (Springer-Verlag, Berlin, 1972).
[10] R. M. Christensen, *Theory of Viscoelasticity. An Introduction* (Academic, New York, 1982).
[11] E. Skudrzyk, *Simple and Complex Vibratory Systems* (The Pennsylvania State University Press, Univeristy Park, PA, 1968).
[12] M. E. Gurtin and E. Sternberg, ''On the linear theory of viscoelasticity,'' Arch. Ration. Mech. Anal. **11**, 291–356 (1962).
[13] W. F. Ames, *Numerical Methods for Partial Differential Equations*, 3rd ed. (Academic, San Diego, 1992).
[14] D. Bonsignour, ''Caractéristiques acoustiques d'un matériau viscoélastique,'' Acustica **23**, 214–222 (1970).
[15] P. J. T. Filipi, O. Lagarrigue, and P.-O. Mattei, ''Perturbation method for sound radiation by a vibrating plate in a light fluid: Comparison with the exact solution,'' J. Sound Vib. **177**, 259–275 (1994).
[16] A. Frendi, L. Maestrello, and A. Bayliss, ''Coupling between plate vibration and acoustic radiation,'' J. Sound Vib. **177**, 207–226 (1994).
[17] G. Maidanik, ''The influence of fluid loading on the radiation from orthotropic plates,'' J. Sound Vib. **3**, 288–299 (1966).
[18] M. C. Junger and D. Feit, *Sound, Structures, and Their Interaction* (MIT Press, Cambridge, 1986).
[19] J. Laroche, ''The use of the matrix pencil method for the spectrum analysis of musical signals,'' J. Acoust. Soc. Am. **94**, 1958–1965 (1993).
[20] M. E. McIntyre and J. Woodhouse, ''On measuring the elastic and damping constants of orthotropic sheet materials,'' Acta Mech. **36**, 1397–1416 (1988).
[21] L. Meirovitch, *Analytical Methods in Vibrations* (MacMillan, New York, 1967).
[22] J. Sanchez Hubert and E. Sanchez Palencia, *Vibration and Coupling of Continuous Systems—Asymptotic Methods* (Springer-Verlag, Berlin, 1989).
[23] V. Bucur, *Acoustics of Wood* (CRC Press, Boca Raton, 1995).
[24] J. Argyris and H. P. Mlejnek, *Dynamics of Structures. Text on Computational Mechanics—Volume V* (Elsevier, Amsterdam, 1991).
[25] A. Chaigne and V. Doutaut, ''Numerical simulations of xylophones. I. Time-domain modeling of the vibrating bars,'' J. Acoust. Soc. Am. **101**, 539–557 (1996).
[26] L. Rhaouti, A. Chaigne, and P. Joly, ''Time-domain modeling and numerical simulation of a kettledrum,'' J. Acoust. Soc. Am. **105**, 3545–3562 (1999).

Time-domain simulation of damped impacted plates. II. Numerical model and results

Christophe Lambourg,[a] Antoine Chaigne,[b] and Denis Matignon
Ecole Nationale Supérieure des Télécommunications, Département TSI, CNRS URA 820, 46 Rue Barrault, 75634 Paris Cedex 13, France

(Received 13 December 1999; revised 11 January 2001; accepted 13 January 2001)

A time-domain model for the flexural vibrations of damped plates was presented in a companion paper [Part I, J. Acoust. Soc. Am. **109**, 1422-1432 (2001)]. In this paper (Part II), the damped-plate model is extended to impact excitation, using Hertz's law of contact, and is solved numerically in order to synthesize sounds. The numerical method is based on the use of a finite-difference scheme of second order in time and fourth order in space. As a consequence of the damping terms, the stability and dispersion properties of this scheme are modified, compared to the undamped case. The numerical model is used for the time-domain simulation of vibrations and sounds produced by impact on isotropic and orthotropic plates made of various materials (aluminum, glass, carbon fiber and wood). The efficiency of the method is validated by comparisons with analytical and experimental data. The sounds produced show a high degree of similarity with real sounds and allow a clear recognition of each constitutive material of the plate without ambiguity. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1354201]

PACS numbers: 43.40.Dx [CBB]

I. INTRODUCTION

A time-domain model of damped isotropic and orthotropic plates was derived in a companion paper (Part I) in order to investigate the relevance of material properties in the quality of sounds produced by vibrating structures. This plate model includes damping terms expressed by means of a general differential operator similar to the one commonly used in viscoelasticity. In the previous paper it was shown under which conditions this operator is able to account for viscoelastic, thermoelastic and radiation losses in materials.[1]

In this paper (Part II, Sec. II) the equations of motion for the damped plate are briefly reviewed in the general case. For convenience, these equations are written in the Laplace domain with complex rigidity factors. The system is complemented by initial and boundary conditions. The initial impact is modeled by Hertz's law of contact. Most of the presented results are obtained in the case of freely suspended plates. However, the model is able to account for other simple boundary conditions, such as clamped or simply supported plates.

The main purpose of this paper is to validate a vibrational model for the impacted plate. No attempt is made to compute the complete acoustic field around the plate with great accuracy. However, in order to facilitate the comparison with real sounds, rather than with vibrational quantities (such as velocity or acceleration at a given point of the plate), a simple radiation model is to calculate radiation from the plate displacement, in order to obtain simulated waveforms having the same dimension as recorded sounds.

In Sec. III, the equations of the model are put into a numerical form using an explicit finite-difference scheme of second order in time and fourth order in space. Similar schemes were successfully applied in the past to the coupling between plate vibration and acoustic radiation by Frendi *et al.*[2] and by Schedin *et al.* for the time-domain simulation of undamped isotropic plates subjected to impacts.[3] The numerical formulation is followed by a necessary analysis of stability and dispersion properties, since no previous references were found in the literature for similar problems which included damping.

The results presented in Sec. IV show the efficiency of the simulation. Following the mode of presentation previously used for xylophones,[4] the method is first validated by comparisons between analytical and numerical results in simple situations and, second, by comparisons between measurements and simulations for damped rectangular plates made either of isotropic or orthotropic materials.

II. THE GOVERNING EQUATIONS

A. The damped-plate equations

The flexural vibrations of a rectangular Kirchhoff–Love plate are considered here (see Fig. 1 for the geometry of the problem). The equations are written in the orthotropic case. The losses are expressed by means of a differential operator similar to the one used in viscoelasticity. For a plate of small thickness h, the transverse displacement $W(x,y,t)$, as a function of the coordinates x,y and time t, is governed by the following equations (expressed in the Laplace domain with Laplace variable s):[1,5]

[a]Present address: 64 rue des Moines, 75017 Paris, France.
[b]Present address: ENSTA-UME, Chemin de la Hunière 91761 Palaiseau cedex, France, electronic mail: chaigne@ensta.fr

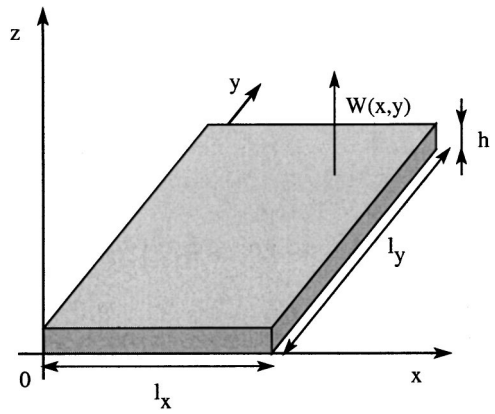


FIG. 1. Geometry of the plate of dimensions l_x and l_y and of thickness h. The flexural displacement $W(x,y)$ is oriented along the z axis.

$$\begin{pmatrix} \tilde{M}_x(x,y,s) \\ \tilde{M}_y(x,y,s) \\ \tilde{M}_{xy}(x,y,s) \end{pmatrix} = -h^3 \begin{pmatrix} \tilde{D}_1(s) & \tilde{D}_2(s)/2 & 0 \\ \tilde{D}_2(s)/2 & \tilde{D}_3(s) & 0 \\ 0 & 0 & \tilde{D}_4(s)/2 \end{pmatrix} \times \begin{pmatrix} \tilde{W}_{,xx}(x,y,s) \\ \tilde{W}_{,yy}(x,y,s) \\ \tilde{W}_{,xy}(x,y,s) \end{pmatrix},$$

$$\rho h(s^2+R_f s)\tilde{W}(x,y,s) = \tilde{M}_{x,xx}(x,y,s) + \tilde{M}_{y,yy}(x,y,s) + 2\tilde{M}_{xy,xy}(x,y,s) + \tilde{f}_z(x,y,s), \tag{1}$$

where the subscripts $,xx$, $,yy$ and $,xy$ denote the partial derivatives of the variables. The first three equations correspond to the viscoelasticlike strain-stress relationships for thin orthotropic plates. $\tilde{M}_x(x,y,s)$, $\tilde{M}_y(x,y,s)$ and $\tilde{M}_{xy}(x,y,s)$ are the Laplace transforms of the bending and twisting moments and $\tilde{D}_i(s)$ are the four complex rigidities. $\tilde{W}(x,y,s)$ is the Laplace transform of the displacement. The fourth equation in Eq. (1) derives from Newton's second law, where ρ is the density of the material, R_f a viscous damping coefficient, and $f_z(x,y,t)$ an excitation source term.

The finite-difference formulation of the problem is facilitated by expressing the strain–stress relationships in terms of differential equations. In the present model, the complex rigidities are written in the form:

$$\tilde{D}_i(s) = D_i(1+\tilde{d}_i(s)) = D_i \frac{1+\sum_{v=1}^{N} s^v p_{iv}}{1+\sum_{w=1}^{N} s^w q_w}, \tag{2}$$

where the coefficients p_{iv} and q_w depend on the damping mechanisms (see Table II). The D_i correspond to the static rigidities of the plate. The $\tilde{d}_i$ are perturbation terms due to damping. In order to ensure that the model is dissipative the coefficients p_{iv} and q_w must fulfill a number of constraints.[1]

B. The boundary conditions

For a plate of finite size, Eq. (1) must be completed by the boundary conditions. Only three ideal conditions for edges parallel to the axes of symmetry of the material are considered here. For the edges located at $x=0$ and $x=l_x$ (see Fig. 1), these conditions are[6]

- for a free edge (F):

$$M_x=0 \text{ and } \frac{\partial M_x}{\partial x}+2\frac{\partial M_{xy}}{\partial y}=0, \tag{3}$$

- for a simply supported edge (SS):

$$W=0 \text{ and } M_x=0, \tag{4}$$

- for a clamped edge (C):

$$W=0 \text{ and } \frac{\partial W}{\partial x}=0. \tag{5}$$

Similar conditions are given for the edges located at $y=0$ and $y=l_y$ by interchanging x and y in Eqs. (3)–(5).

C. The interaction between the plate and the impactor

In Eq. (1), it is assumed that the load $f(x,y,t)$ due to the interaction between the plate and the impactor is governed by Hertz's law of contact.[7] This method has been used with success in the past for the time-domain modeling of impacted xylophone bars[8] and for a mixed time-frequency investigation on impacts on plates.[9] The motion W_i of the impactor is a solution of the following system:

$$f(x,y,t) = g(x-x_0,y-y_0)F_H(t),$$

$$F_H(t) = \begin{cases} [(W(x_0,y_0,t)-W_i(t))/k_H]^{3/2} \\ \quad \text{if } W_i(t) < W(x_0,y_0,t), \\ 0 \quad \text{otherwise} \end{cases} \tag{6}$$

$$\frac{d^2W_i(t)}{dt^2} = \frac{F_H(t)}{m_i},$$

where $g(x-x_0,y-y_0)$ is a normalized spatial window centered at the impact point, F_H is the interaction force, $W-W_i$ is the relative displacement between the impactor and the plate at the impact point, m_i is the mass of the impactor and k_H is Hertz's constant which depends on the surface geometries and elastic properties of the two bodies in contact. In Eq. (6), the vibrations in the impactor are neglected and it is assumed that the contact surface remains constant during the impact. These assumptions are reasonable when the dimensions of the impactor are small compared to those of the plate.[9]

D. The radiation equation

For the purpose of simplicity, only the case of baffled plates will be considered here. Thus, for computing the radiated pressure from the plate displacement W, the following time-domain formulation of the Rayleigh integral was used:

$$p(\mathbf{r},t) = -\frac{\rho_a}{2\pi}\int\int_{S_0}\frac{1}{|\mathbf{r}-\mathbf{r}_0|}\frac{\partial^2 W}{\partial t^2}\left(\mathbf{r}_0, t-\frac{|\mathbf{r}-\mathbf{r}_0|}{c}\right)dS_0, \tag{7}$$

where p is the sound pressure at the position $\mathbf{r}$ in the sound field, ρ_a is the air density, c is the speed of sound in air, S_0 is the surface of the plate and $\mathbf{r_0}$ refers to the position of the source points on the plate surface.

III. THE NUMERICAL MODEL

The present work is limited to rectangular plates for which the boundaries coincide with the nodes of a regular rectangular grid. For this class of geometries, it is known that the Finite-Difference Methods (FDM) are convenient for solving vibration problems in the time domain.[10,11]

In this section, the use of a 2–4 finite-difference scheme for solving the impacted plate equations is studied in terms of stability and numerical dispersion. This scheme is of second order in time and fourth order in space. The numerical parameters to be determined are the time step Δt and the two spatial steps Δx, Δy along the x and y axis, respectively. The appropriate selection of these parameters is imposed both by the numerical stability condition and by the order of accuracy required for the dispersion.

In what follows, the guideline for the dispersion criteria was given by the accuracy required in the context of audio applications. According to Moore,[12] the minimum value of the relative difference limen for pitch of a pure tone is about 0.5% at 2 kHz, and becomes greater than 5% for frequencies above 5 kHz. Since there are no available data for the difference limen relative to tones made of simultaneous inharmonic frequencies, the pure tone difference limen was used as reference for the required accuracy of the numerical model.

A. Explicit finite-difference schemes

The transverse displacement of the plate is computed at the nodes ($x=l\Delta x$, $y=m\Delta y$, $t=n\Delta t$) of a rectangular grid. The value of a field variable $v(x,y,t)$ expressed on this mesh is denoted $v_{l,m}^{n}=v(l\Delta x,m\Delta y,n\Delta t)$.

The elastic and inertial terms, involving second order partial derivatives versus time or space, are discretized by means of centered finite difference operators (see Appendix A). The damping terms are approximated by decentered operators which has the advantage to keep the explicit character of the numerical formulation. This strategy is justified by the fact that the damping terms $\tilde{d}_i$ defined in Eq. (2) are assumed to be first order correction terms. Centered operators for the damping terms would have led to an implicit scheme. The time derivatives in the damping terms $\tilde{d}_i$ are approximated by backward difference operators which, in terms of the z-transform, amounts to replacing the continuous Laplace variable s by the discrete operator $(1-z^{-1})/\Delta t$. Thus, the discrete approximation of $\tilde{d}_i$, written in descending power of z, is written:

$$\frac{\Sigma_{r=0}^{N}\chi_{ir}z^{-r}}{1+\Sigma_{r=1}^{N}\psi_r z^{-r}}=\frac{\Delta t^N+\Sigma_{r=1}^{N}\Delta t^{N-r}(1-z^{-1})^r p_{ir}}{\Delta t^N+\Sigma_{r=1}^{N}\Delta t^{N-r}(1-z^{-1})^r q_r}. \quad (8)$$

In Eq. (8), the constants χ_{ir} and ψ_r are obtained by identifying term by term the two sides of Eq. (8).

The finite-difference approximation of Eq. (1) is then given by:

$$\begin{aligned}
(M_x)_{l,m}^{n} &= -h^3 D_1 \sum_{r=0}^{N} \chi_{1r}(D_{xx}W)_{l,m}^{n-r} \\
&\quad - h^3 \frac{D_2}{2}\sum_{r=0}^{N}\chi_{2r}(D_{yy}W)_{l,m}^{n-r} - \sum_{r=1}^{N}\psi_r(M_x)_{l,m}^{n-r} \\
(M_y)_{l,m}^{n} &= -h^3 D_3 \sum_{r=0}^{N} \chi_{3r}(D_{yy}W)_{l,m}^{n-r} \\
&\quad - h^3 \frac{D_2}{2}\sum_{r=0}^{N}\chi_{2r}(D_{xx}W)_{l,m}^{n-r} - \sum_{r=1}^{N}\psi_r(M_y)_{l,m}^{n-r} \\
(M_{xy})_{l+1/2,m+1/2}^{n} &= -h^3 \frac{D_4}{2}\sum_{r=0}^{N}\chi_{4r}(D_{xy}W)_{l+1/2,m+1/2}^{n-r} \\
&\quad - \sum_{r=1}^{N}\psi_r(M_{xy})_{l+1/2,m+1/2}^{n-r} \\
W_{l,m}^{n+1} &= (2-R_f\Delta t)W_{l,m}^{n} + (R_f\Delta t-1)W_{l,m}^{n-1} \\
&\quad + \frac{\Delta t^2}{\rho h}((D_{xx}M_x)_{l,m}^{n} + (D_{yy}M_y)_{l,m}^{n} \\
&\quad + 2(\tilde{D}_{xy}M_{xy})_{l,m}^{n} + (f_z)_{l,m}^{n}),
\end{aligned} \quad (9)$$

where the discrete spatial operators D_{xx}, D_{yy}, D_{xy} and $\tilde{D}_{xy}$ are fourth order approximations (see Appendix A). It can be seen from Eq. (9) that the transverse displacement W is recursively determined at time $n+1$ from its values at previous time steps.

B. Boundary conditions

The numerical approximation of the boundary conditions is obtained by means of the image method. It consists of replacing the boundaries by virtual sources located at points external to the plate. In what follows, only the boundary conditions for a free straight edge located at $x=0$, for a semi-infinite plate, is discussed [see Eq. (3)]. The results can be easily extended to simply supported and clamped edges.[13] The numerical formulation of the boundary conditions described below is conducted so as to maintain the overall accuracy of the scheme and to limit the computational domain.

It is assumed that the semi-infinite plate is located in $x>0$ (see Fig. 1) and that the transverse displacement W, at time $n\Delta t$ and at every previous time step, is known at each point $l\Delta x,m\Delta y$ such that $l\geqslant -1$. The values of W at time $n+1$ are then obtained by using the following procedure:

- First, the bending and twisting moments are computed at time $n\Delta t$. This computation is performed by applying the first three equations in Eq. (9) to each point of the grid for which $x>0$. For $x=0$, M_x is equal to zero and M_y is derived from the displacement by using the second equation in Eq. (9), after replacing $D_{xx}^{(4)}$ by the second order operator $D_{xx}^{(2)}$ in order to limit the spatial extension of the boundary domain. $(M_{xy})_{-1/2,m}^{n}$ is given by the third equation in Eq. (9) where $D_{xy}^{(4)}$ is replaced by the operator $D_x^{(2)}oD_y^{(4)}$ defined as:

$$(D_x^{(2)} o D_y^{(4)} W)^n_{-1/2,m+1/2}$$
$$= \frac{1}{24\Delta x \Delta y}[(W^n_{0,m-2} - 27W^n_{0,m-1} + 27W^n_{0,m} - W^n_{1,m+2}) - (W^n_{-1,m-2} - 27W^n_{-1,m-1} + 27W^n_{-1,m} - W^n_{-1,m+2})]. \quad (10)$$

- The application of the image method also involves the computation of the values of M_x at points $x=-\Delta x$. This is performed by considering the second condition in Eq. (3), for which a third order decentered approximation in Δx and fourth order in Δy leads to:

$$(M_x)^n_{-1,m} = 3(M_x)^n_{1,m} - \tfrac{1}{2}(M_x)^n_2 - 3\Delta x((\tilde{D}_y^{(4)} M_{xy})^n_{-1/2,m} + (\tilde{D}_y^{(4)} M_{xy})^n_{1/2,m}), \quad (11)$$

where

$$(\tilde{D}_y^{(4)} M_{xy})^n_{l+1/2,m} = \frac{9}{8\Delta y}((M_{xy})^n_{l+1/2,m+1/2} - (M_{xy})^n_{l+1/2,m-1/2}) - \frac{1}{24\Delta y}((M_{xy})^n_{l+1/2,m+3/2} - (M_{xy})^n_{l+1/2,m-3/2}). \quad (12)$$

- The final task consists in computing W at time $(n+1)\Delta t$. This is performed by using the fourth equation in Eq. (9) for the points $l>0$, and by replacing the fourth order operators in Δx by second order ones in this equation at points $l=0$. The displacement at the image points $l=-1$ is derived from the free edge condition $M_x=0$ at $x=0$, for which a second order in Δx, and fourth order in Δy approximation leads to the following expression of $W^{n+1}_{-1,m}$:

$$W^{n+1}_{-1,m} = -\frac{\Delta x^2 D_2}{2D_1}(D_{yy}^{(4)} W)^{n+1}_{0,m} + 2W^{n+1}_{0,m} - W^{n+1}_{1,m}. \quad (13)$$

The influence of the damping terms is neglected in the discrete approximation of the boundary conditions. The schemes for the three other edges of the plate obey the same principles as the one presented above for $x=0$.

C. Stability

The numerical parameters of explicit difference schemes must fulfill a stability condition which guarantees the convergence of the numerical solution.[14] In this paragraph, the stability condition for the 2–4 scheme applied to the undamped plate model is presented. The results are then extended to the damped model. A distinction is made in what follows between two definitions of stability:[14]

- The numerical scheme is called *weakly stable* if the solution at a fixed time $T=n\Delta t$ is bounded for all Δt with respect to the Euclidean norm $\|W^n\| = (\Sigma_{l,m}|W^n_{l,m}|^2)^{1/2}$.
- The scheme is called *strongly stable* if the solution remains bounded for all T.

A strong stability condition can be obtained for the undamped model. In the damped case, paradoxically, only a weak stability condition can be given.

1. Undamped model

In this paragraph, the fluid damping constant R_f and the damping coefficients p_{ir} and q_r are set equal to zero. The Fourier method is used for determining the stability criterion.[14] It consists of studying the modulus of one particular solution $W^n_{l,m} = w_n \exp(jk_x l\Delta x + jk_y m\Delta y)$ of the homogeneous equation. Injecting $W^n_{l,m}$ in Eq. (9) for $f^n_{l,m}=0$ leads to the following equation for w_n:

$$w_{n+1} + (b(k_x,k_y)^2\Delta t^2 - 2)w_n + w_{n-1} = 0, \quad (14)$$

with:

$$b(k_x,k_y)^2 = 16\left(\frac{\xi_1^2}{\Delta x^4}\left(X^2+\frac{1}{3}X^4\right)^2 + \frac{\xi_3^2}{\Delta y^4}\left(Y^2+\frac{1}{3}Y^4\right)^2 + \frac{\xi_2^2}{\Delta x^2\Delta y^2}\left(X^2+\frac{1}{3}X^4\right)\left(Y^2+\frac{1}{3}Y^4\right) + \frac{\xi_4^2}{\Delta x^2\Delta y^2}\left(X+\frac{1}{6}X^3\right)^2\left(Y+\frac{1}{6}Y^3\right)^2\right), \quad (15)$$

and where

$$\xi_i = \sqrt{\frac{h^2 D_i}{\rho}}, \quad X = \sin\frac{k_x\Delta x}{2}, \quad Y = \sin\frac{k_y\Delta y}{2}. \quad (16)$$

The stability criterion is satisfied if the discriminant of the characteristic polynomial associated with Eq. (14) is negative for all k_x and k_y. This leads to the following condition:

$$\Delta t \leqslant \frac{1}{2\sqrt{\left(\frac{4}{3}\right)^2\left(\xi_1^2 + \frac{\xi_2^2}{r_1^2} + \frac{\xi_3^2}{r_1^4}\right) + \left(\frac{7}{6}\right)^4\frac{\xi_4^2}{r_1^2}}}, \quad (17)$$

where $r_1 = \Delta y/\Delta x$.

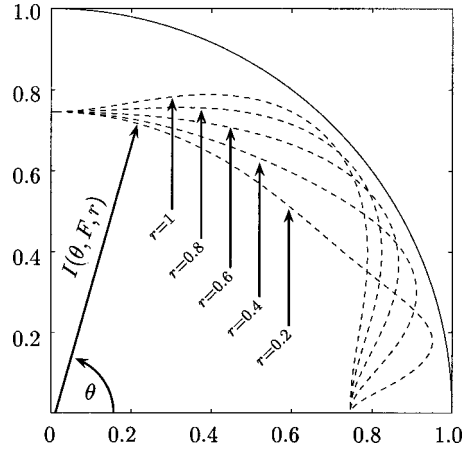


FIG. 2. Numerical isotropy $I(\theta,F,r)$ for a reduced frequency F equal to 0.2 and for different values of the degree r of anisotropy: $r=\{0.2,0.4,0.6,0.8,1\}$.

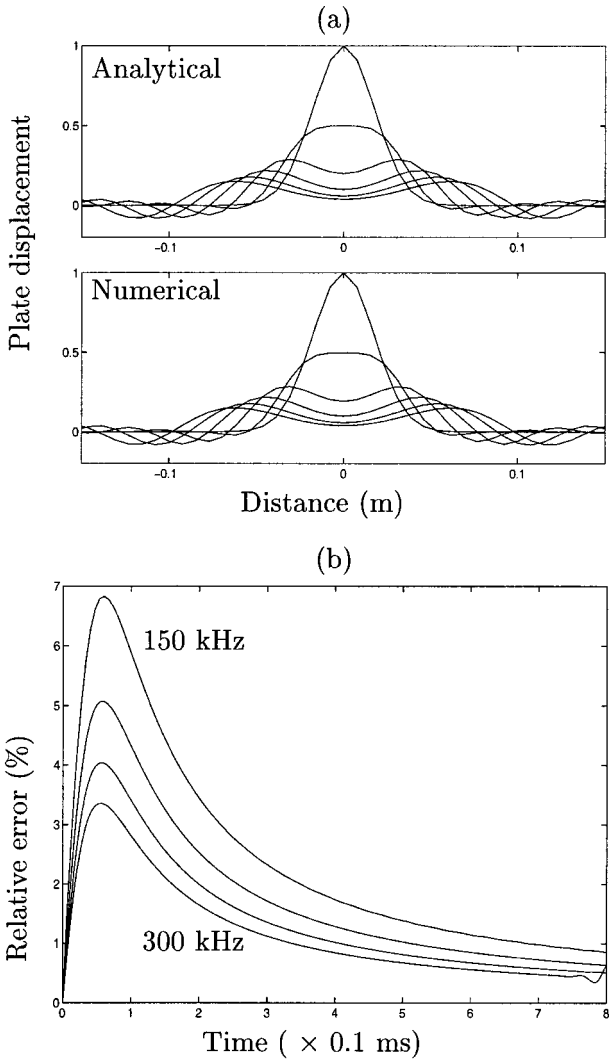


FIG. 3. (a) Infinite isotropic plate profile at successive instants of time $t=\{0,0.05,0.10,0.15,0.20,0.25\}$ ms. Top: analytical solution. Bottom: numerical solution. (b) Relative error (in %) between analytical and numerical solutions versus time for four different sampling frequencies $f_e=\{150,200,250,300\}$ kHz.

2. Damped model

A model including the damping terms is now considered. To facilitate the presentation, the case of a damped bending bar (1D problem) is first presented and the results are then extended to plates (2D).

Similarly to Eq. (1), the equations governing the transverse displacement $W(x,t)$ of a thin bar are written:

$$\tilde{M}_x=-EI\frac{1+\Sigma_{r=1}^{N}p_r s^r}{1+\Sigma_{r=1}^{N}q_r s^r}\tilde{W}_{,xx}$$
$$\rho S(s^2+R_f s)\tilde{W}=\tilde{M}_{x,xx}+\tilde{f}_z. \tag{18}$$

The corresponding FD scheme is written:

$$(M_x)_l^n=-IE\sum_{r=0}^{N}\chi_r(D_{xx}W)_l^{n-r}-\sum_{r=1}^{N}\psi_r(M_x)_l^{n-r},$$
$$W_l^{n+1}=(2-R_f\Delta t)W_l^n+(R_f\Delta t-1)W_l^{n-1}-\frac{\Delta t^2}{\rho S}((D_{xx}M_x)_l^n+(f_z)_l^n), \tag{19}$$

where E is Young's modulus, I is the geometrical moment and S is the cross-section of the bar. The term f_z denotes the external force density by unit length. D_{xx} is a fourth order discrete operator.

A necessary condition for weak stability is obtained by studying the amplification matrix[14] (see Appendix B). This condition is written:

$$\Delta x^2\geqslant 2\Delta t\sqrt{\frac{IEp_N}{\rho S q_N}}. \tag{20}$$

It is found that the condition expressed in Eq. (20) is similar to the one obtained with the Fourier method in the undamped case, except that E is now replaced by Ep_N/q_N. This term corresponds to the high frequencies' asymptotic value of the complex Young's modulus.

The stability condition for the 2D system Eq. (9) is determined in the same way as for the 1D case (see Appendix B). It is found that the weak stability condition of the damped 2–4 scheme is similar to Eq. (17), after replacing each D_i by $D_i p_{iN}/q_N$. Since p_{iN}/q_N is always greater than unity for a dissipative model, this result shows that the stability condition is more restrictive when damping terms are taken into account than for the undamped plate. However, for the investigated materials (aluminum, steel, glass and wood), it has been found that the consecutive increase in size of the minimum space step remains below a few percent, compared to the undamped case.

D. Accuracy

One method for estimating the deviation between exact and approximate solutions consists of calculating the difference between continuous and discrete frequencies.[10,15] For simplicity, only the undamped model will be discussed here.

Inserting the solution $W(x,y,t)=(Ae^{j\omega t}+Be^{-j\omega t})\exp(jk_x x+jk_y y)$ in Eq. (1), without damping terms and for $f_z=0$, yields the continuous dispersion equation:

$$k=\sqrt{\frac{\omega}{\xi_\theta}}\quad\text{with}\quad k=\sqrt{k_x^2+k_y^2},\quad \theta=\arctan\frac{k_y}{k_x},\quad \xi_\theta=\frac{h^2}{\rho}(D_1\cos^4\theta+(D_2+D_4)\cos^2\theta\sin^2\theta+D_3\sin^4\theta). \tag{21}$$

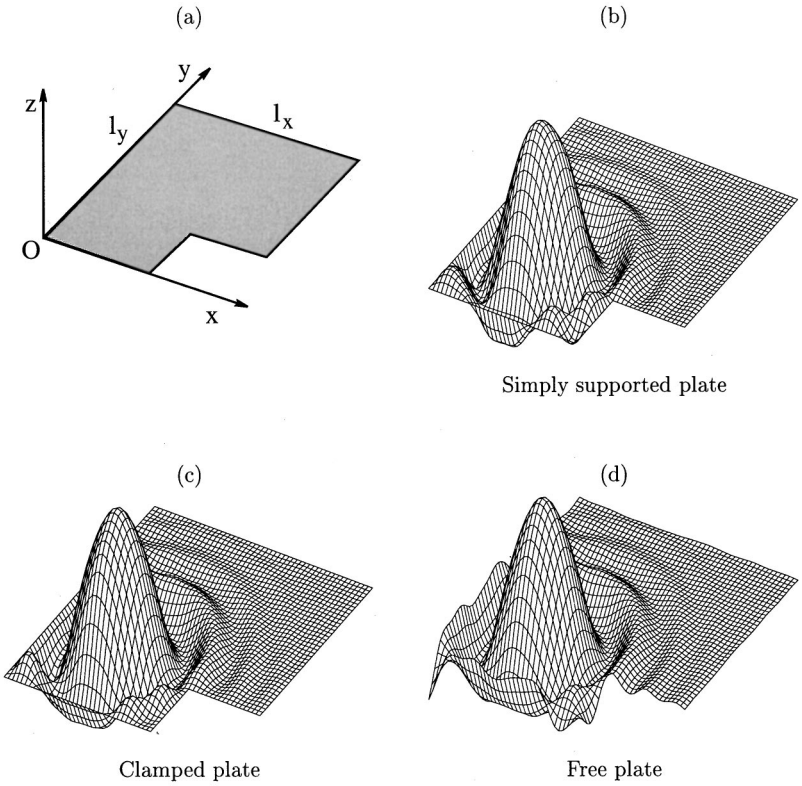


FIG. 4. Bending vibrations of a L-shaped isotropic plate, a few ms after the impact, with different boundary conditions. (a) Geometry of the plate at rest. (b) Simply supported plate. (c) Clamped plate. (d) Free plate.

Similarly, provided that the stability condition is verified, the introduction of the component $W_{l,m}^{n}=(Ae^{j\omega_{\text{num}}n\Delta t}+Be^{-j\omega_{\text{num}}n\Delta t})\exp(jk_x l\Delta x+jk_y m\Delta y)$ in Eq. (9) leads to the numerical dispersion equation:

$$\omega_{\text{num}}=\frac{2}{\Delta t}\arcsin\frac{b(k_x,k_y)\Delta t}{2}, \tag{22}$$

where $b(k_x,k_y)$ is defined in Eq. (15).

In order to simplify the mathematical derivations, it is assumed here that $\xi_2=0$ and that $\xi_4=\sqrt{2\xi_1\xi_3}$ which constitutes a reasonable approximation for a large class of orthotropic materials.[16] This allows us to characterize the anisotropy of the material with only one parameter $r=\xi_3/\xi_1$. Under these conditions Eqs. (21) and (22), combined with Eq. (17), yield:

$$\begin{aligned}\Omega_{\text{num}}=2\arcsin\Bigg[\frac{36}{\sqrt{9410}}\Bigg(\left(X^2+\frac{X^4}{3}\right)^2+\left(Y^2+\frac{Y^4}{3}\right)^2\\ +2\left(X+\frac{X^3}{6}\right)^2\left(Y+\frac{Y^3}{6}\right)^2\Bigg)^{1/2}\Bigg]\\ X=\sin\left(\left(\frac{4705}{2592}\right)^{1/4}\frac{\sqrt{\Omega}\cos\theta}{\sqrt{\cos^2\theta+r^2\sin^2\theta}}\right)\\ Y=\sin\left(\left(\frac{4705}{2592}\right)^{1/4}\frac{r\sqrt{\Omega}\sin\theta}{\sqrt{\cos^2\theta+r^2\sin^2\theta}}\right),\end{aligned} \tag{23}$$

where $\Omega=\omega\Delta t=2\pi F$ and $\Omega_{\text{num}}=\omega_{\text{num}}\Delta t$ are the reduced angular frequencies.

Finally, the relative error in frequency is given by:

$$e(\Omega,\theta)=\frac{\Omega-\Omega_{\text{num}}}{\Omega}. \tag{24}$$

It has been observed that about 15 time steps per period are required with the 2–4 scheme in order to keep the error in frequency smaller than 5% at 20 kHz, whereas more than 40 time-steps by period would have been needed with the 2–2 scheme. In the audio range 0–20 kHz, this degree of accuracy requires a sampling rate approximately equal to 300 kHz for the 2–4 scheme, whereas 800 kHz would have been needed with the 2–2 scheme.

For a material with a given anisotropy degree r, Eq. (23) shows that the numerical error not only depends on frequency, but also on the propagation angle θ in the plate. A convenient method for representing this error is to define an isotropy index:[15]

$$I(\theta,F,r)=\frac{c_{\text{num}}}{c}, \tag{25}$$

where c_{num} is the numerical phase velocity and c the exact phase velocity. Figure 2 represents the variations of I with θ, for a given reduced frequency $F=0.2$, and for different values of the degree r of anisotropy. Notice that the angle θ_m corresponding to the maximum index I_M moves from $\pi/4$ to zero as the anisotropy degree r of the material decreases. However, the ratio I_M/I_m between the maximum and the minimum values of the isotropy index remains unchanged as r varies.

In our applications, the plate equation has been solved with a 2–4 scheme and with a sampling rate of 192 kHz. This leads to an error of 8% in the estimation of the phase velocity at 20 kHz and of 2.5% at 10 kHz.

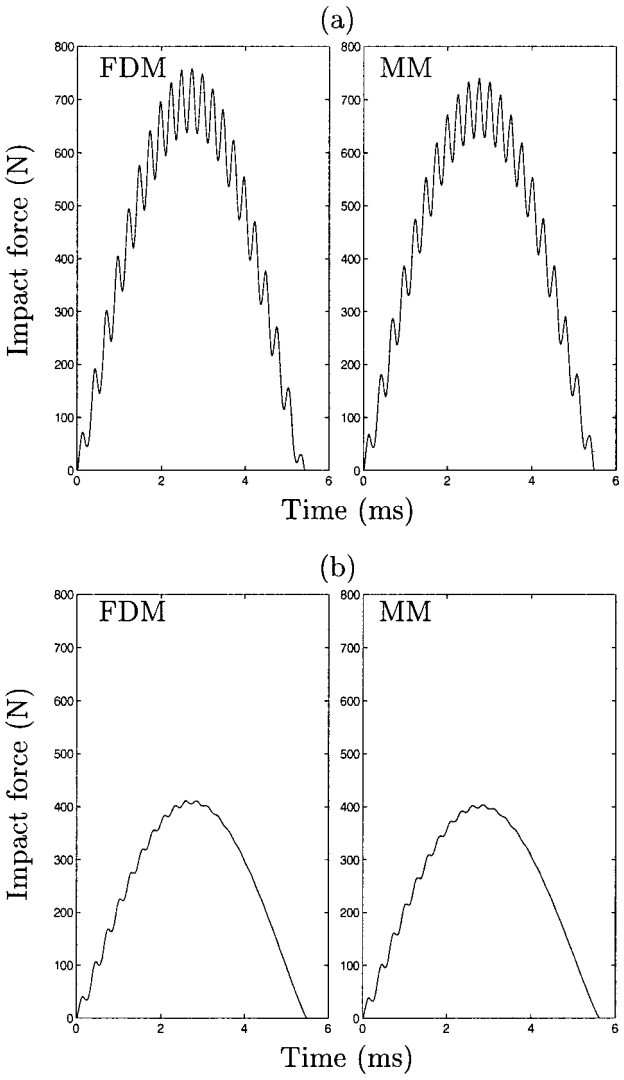


FIG. 5. Interaction force between an orthotropic plate and an impactor during the contact time, for two different initial velocities of the impactor: (a) 1.45 m/s, (b) 0.6 m/s. Left: Simulation with the present Finite-Differences Method (FDM). Right: Simulation made by McMillan (Ref. 9).

E. Computation of the plate-impactor interaction

The discrete formulation of Eq. (6) is given by

$$f_{l,m}^n = -g_{l,m} F_h^n$$

$$W_i^{n+1} = 2W_i^n - W_i^{n-1} + \frac{\Delta t^2}{m_s} F_h^n \tag{26}$$

$$F_h^{n+1} = \begin{cases} [(W_p^{n+1} - W_i^{n+1})/k_H]^{3/2} & \text{if } W_i^{n+1} < W_p^{n+1} \\ 0 & \text{otherwise.} \end{cases}$$

W_p is the displacement of the plate at impact point (x_0, y_0). The spatial window $g_{l,m}$ is a triangular function of width $2\Delta x$ and $2\Delta y$, centered at point (x_0, y_0).

F. Computation of the sound pressure

The semi-discrete formulation of the Rayleigh integral presented in Eq. (7) is obtained by using a standard trapezoidal rule:

$$p(\mathbf{r},t) = -\frac{\rho_a}{2\pi} \sum_{l=0}^{N_x-1} \sum_{m=0}^{N_y-1} \frac{\Delta x \Delta y}{4} \frac{d^2}{dt^2} \left(\frac{W_{l,m}(t - R_{l,m}/c)}{R_{l,m}} + \frac{W_{l+1,m}(t - R_{l+1,m}/c)}{R_{l+1,m}} + \frac{W_{l,m+1}(t - R_{l,m+1}/c)}{R_{l,m+1}} + \frac{W_{l+1,m+1}(t - R_{l+1,m+1}/c)}{R_{l+1,m+1}} \right). \tag{27}$$

In this equation, N_x and N_y are the nearest integers smaller than or equal to $l_x/\Delta x$ and $l_y/\Delta y$, respectively, where $0 < x < l_x$ and $0 < y < l_y$ is the plate domain. $R_{l,m}$ is the distance between the listening point and each mesh point of the plate and $R_{l,m}/c$ is the corresponding time delay. After time-domain discretization, this quantity is approximated by the nearest multiple of Δt such that:

$$W_{l,m}(n\Delta t - R_{l,m}/c) \simeq W_{l,m}^{n - i_{l,m}} \quad \text{with} \quad i_{l,m} = \text{int}\left(\frac{R_{l,m}}{c\Delta t} \right). \tag{28}$$

Finally, the discrete formulation of Eq. (27) is written:

$$p(\mathbf{r}, n\Delta t) = p^n = D_{tt} \sum_{i=i_{\min}}^{i_{\max}} \sum_{l=0}^{N_x} \sum_{m=0}^{N_y} \delta_{i i_{l,m}} A_{l,m} W_{l,m}^{n-i}, \tag{29}$$

where

$$\begin{aligned} &i_{\min} = \min(i_{l,m}), \quad i_{\max} = \max(i_{l,m}) \\ &D_{tt} \quad \text{discrete backward time-derivative operator} \\ &A_{l,m} = -\frac{\Delta x \Delta y \rho_a}{2\pi R_{l,m}} B_{l,m} \end{aligned} \tag{30}$$

$$B_{l,m} = \begin{cases} 1 & \text{if } 0 < l < N_x \text{ and } 0 < m < N_y \\ 1/2 & \text{if } ((l = 0 \text{ or } N_x) \text{ and } (0 < m < N_y)) \\ & \text{or } ((m = 0 \text{ or } N_y) \text{ and } (0 < l < N_x)) \\ 1/4 & \text{otherwise.} \end{cases}$$

IV. RESULTS OF SIMULATIONS

In order to assess the validity of the method, this section starts with a comparison between the numerical result and the theoretical solution obtained in a simple case where the plate displacement is given analytically. This presentation is followed by the simulation of the plate for three different boundary conditions. The force pulse obtained with our numerical scheme is then compared to the force pulse calculated for the same plate impacted with the same impactor using the method developed by McMillan.[9] Finally, simulated waveforms, and their corresponding spectra, are compared to measurements for four different materials (aluminum, glass, carbon fiber and wood).

A. Comparison with analytical displacement

An analytical solution to the plate equation can be obtained only for a limited number of cases. One example of such an analytical solution can be found in textbooks for an infinite isotropic plate subjected to the initial Gaussian shape:[6]

TABLE I. Elastic and geometrical parameters of plates and impactors.

Aluminum plates				
$\rho = 2660\ \mathrm{kg\ m^{-3}}$	$D_1 = 6160$ MPa	$D_4 = 8600$ MPa		
plate a_1:	$l_x = 304$ mm	$l_y = 192$ mm	$h = 2$ mm	$f_c = 6$ kHz
plate a_3:	$l_x = 419.5$ mm	$l_y = 400$ mm	$h = 4$ mm	$f_c = 3$ kHz
Glass plate				
$\rho = 2550\ \mathrm{kg\ m^{-3}}$	$D_1 = 6700$ MPa	$D_4 = 10270$ MPa		
plate v_1:	$l_x = 229.5$ mm	$l_y = 220.5$ mm	$h = 2$ mm	$f_c = 5.8$ kHz
Carbon plate				
$\rho = 1540\ \mathrm{kg\ m^{-3}}$	$D_1 = 8437$ MPa	$D_2 = 463$ MPa	$D_3 = 852$ MPa	$D_4 = 2267$ MPa
plate c_1:	$l_x = 399$ mm	$l_y = 200$ mm	$h = 2.2$ mm	critical zone=[3.6,11.3] kHz
Carbon–Epoxy plate				
$\rho = 1581\ \mathrm{kg\ m^{-3}}$	$D_1 = 6912$ MPa	$D_2 = 51.5$ MPa	$D_3 = 4750$ MPa	$D_4 = 1015$ MPa
plate m_1:	$l_x = 70.7$ mm	$l_y = 70.7$ mm	$h = 2$ mm	
Wooden plate				
$\rho = 388\ \mathrm{kg\ m^{-3}}$	$D_1 = 1013$ MPa	$D_2 = 27.5$ MPa	$D_3 = 53.7$ MPa	$D_4 = 221$ MPa
plate b_1:	$l_x = 515$ mm	$l_y = 412.5$ mm	$h = 4.8$ mm	critical zone=[2.4,10.3] kHz
Titanium impactor				
radius $R = 10$ mm	$m_i = 1.12$ kg			
$k_H = 4.0\times10^{-6}\ \mathrm{m\ N^{-2/3}}$				
Rubber impactor				
radius $R = 10$ mm	$m_i = 23.6$ g			
$k_H = 9.0\times10^{-6}\ \mathrm{m\ N^{-2/3}}$				

TABLE II. Damping parameters.

Thermoelastic damping (aluminum)			
$\tilde{D}_1(s) = D_1\left(1+\frac{sR_1}{s+(c_1/h_2)}\right)$,	$\tilde{D}_4(s) = D_4$		
$R_1 = 8.45\times10^{-3}$	$c_1 = 8.0\times10^{-4}\ \mathrm{rad\ m^2\ s^{-1}}$		
Viscoelastic damping (glass)			
$\tilde{D}_i(s) = D_i\left(1+\frac{sR_1}{s+s_1}+\frac{sR_2}{s+s_2}\right)$,	$i = [1, 4]$		
$R_1 = 1.63\times10^{-3}$	$s_1 = 5180\ \mathrm{rad\ s^{-1}}$	$R_2 = 1.962\times10^{-3}$	$s_2 = 55\,100\ \mathrm{rad\ s^{-1}}$
Viscoelastic damping (carbon)			
$\tilde{D}_i(s) = D_i\left(1+\frac{sR_{i1}}{s+s_{i1}}+\frac{sR_{i2}}{s+s_{i2}}\right)$,	$i = [1, 3, 4]$	$\tilde{D}_2(s) = D_2$	
$R_{11} = 1.32\times10^{-3}$	$s_{11} = 10.1\times10^3\ \mathrm{rad\ s^{-1}}$	$R_{12} = 5.0\times10^{-3}$	$s_{12} = 94.0\times10^3\ \mathrm{rad\ s^{-1}}$
$R_{31} = 8.8\times10^{-3}$	$s_{31} = 2.5\times10^3\ \mathrm{rad\ s^{-1}}$	$R_{32} = 44.0\times10^{-3}$	$s_{32} = 70.0\ \mathrm{rad\ s^{-1}}$
$R_{41} = 10.4\times10^{-3}$	$s_{41} = 2.27\times10^3\ \mathrm{rad\ s^{-1}}$	$R_{42} = 14.4\times10^{-3}$	$s_{42} = 40.0\times10^3\ \mathrm{rad\ s^{-1}}$
Viscoelastic damping (wood)			
$\tilde{D}_i(s) = D_i\left(1+\frac{sR_{i1}}{s+s_{i1}}+\frac{sR_{i2}}{s+s_{i2}}\right)$,	$i = [1, 3, 4]$	$\tilde{D}_2(s) = D_2$	
$R_{11} = 8.18\times10^{-3}$	$s_{11} = 3.2\times10^3\ \mathrm{rad\ s^{-1}}$	$R_{12} = 10.0\ 10^{-3}$	$s_{12} = 50.2\times10^3\ \mathrm{rad\ s^{-1}}$
$R_{31} = 16.7\times10^{-3}$	$s_{31} = 1.1\times10^3\ \mathrm{rad\ s^{-1}}$	$R_{32} = 70.0\times10^{-3}$	$s_{32} = 50.2\ \mathrm{rad\ s^{-1}}$
$R_{41} = 15.2\times10^{-3}$	$s_{41} = 1.75\times10^3\ \mathrm{rad\ s^{-1}}$	$R_{42} = 35.0\times10^{-3}$	$s_{42} = 50.2\times10^3\ \mathrm{rad\ s^{-1}}$
Radiation damping (isotropic plates)			
$\tilde{D}_i(s) = D_i\left(1+\frac{2\rho_a c}{\rho h\omega_c}\frac{\Sigma_1^3 b_m(s/\omega_c)^m}{\Sigma_0^3 a_n(s/\omega_c)^n}\right)$	$i = [1, 4]$	$\omega_c = c^2\sqrt{\frac{\rho}{h^2 D_1}}$	
$\rho_a = 1.2\ \mathrm{kg\ m^{-3}}$	$c = 344\ \mathrm{m\ s^{-1}}$		
$a_0 = 1.1669$	$a_1 = 1.6574$	$a_2 = 1.5528$	$a_3 = 1$
$b_1 = 0.0620$	$b_2 = 0.5950$	$b_3 = 1.0272$	
Fluid damping R_f in $\mathrm{s^{-1}}$			
aluminum $R_f = 0.032$	glass $R_f = 0.88$	carbon $R_f = 0.8$	wood $R_f = 2.4$

TABLE III. General simulation parameters.

Typical time and spatial steps		
$f_e = 1/\Delta t = 192$ kHz	$\Delta x = \Delta y = 9.5$ mm	
Impact position		
$x_o = l_x/2$	$y_o = l_y/2$	
Listening point		
$x = l_x/2$	$y = l_y/2$	$z = 24$ cm

$$W(0,r) = W_0 e^{-r^2/a^2}. \tag{31}$$

In this case, the displacement at time t is given by:

$$W(t,r) = \frac{W_0}{1+\tau^2} e^{-r^2/a^2(1+\tau^2)} \times \left[\cos\frac{r^2\tau}{a^2(1+\tau^2)} + \tau \sin\frac{r^2\tau}{a^2(1+\tau^2)} \right], \tag{32}$$

where

$$\tau = \frac{4t}{a^2}\sqrt{h^2 D_1 \rho}. \tag{33}$$

Figure 3(a) shows the shape of the plate at successive instants of time (between 0 and 0.25 ms) for an aluminum plate with thickness $h = 2$ mm, and for an initial Gaussian shape of width $a = 2.5$ cm. The other dimensions of the plate are sufficiently large (1 m$\times$1 m) so that no reflexions from the boundaries are observed during the selected time scale. No differences can be easily seen between the exact solution [top of Fig. 3(a)] and the simulated solution obtained with

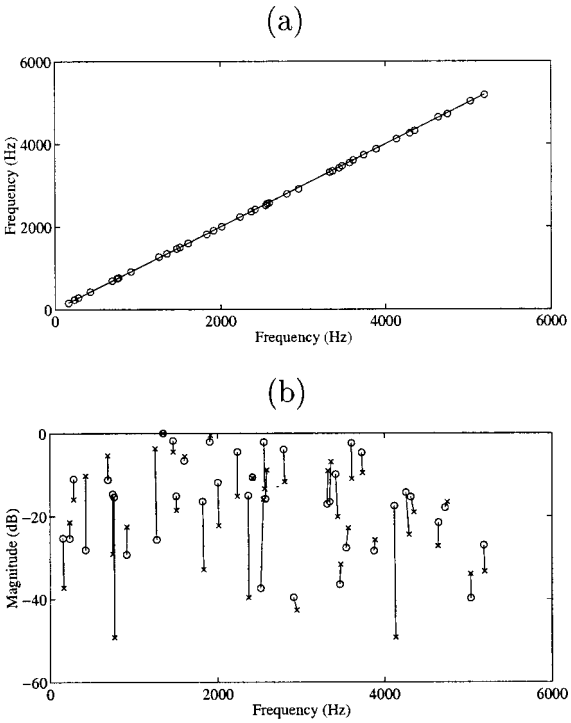


FIG. 7. Glass plate. FFT analysis of the sound pressure during the first 20 ms. (a) Simulated vs measured frequencies in the range 0–5 kHz (''○''). The solid line indicates perfect matching between the two frequency sets (slope equal to 1). (b) Comparison between simulated (''×'') and measured (''○'') magnitudes of the main FFT peaks.

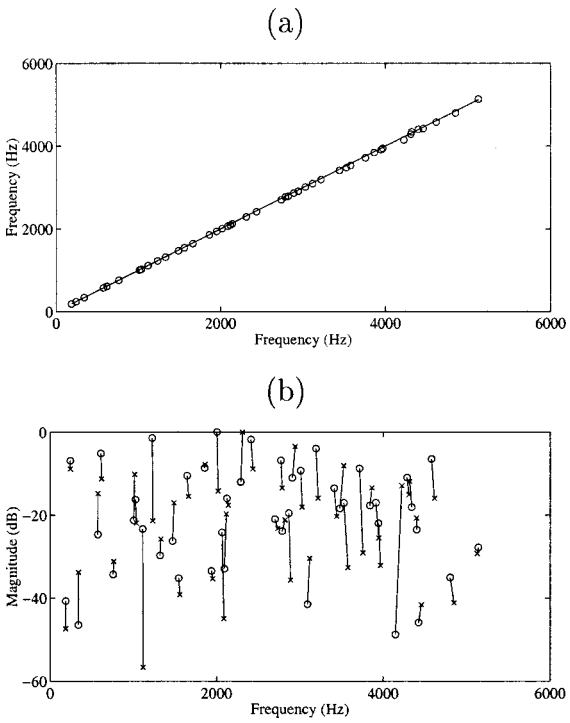


FIG. 6. Aluminum plate. FFT analysis of the sound pressure during the first 20 ms. (a) Simulated vs measured frequencies in the range 0–5 kHz (''○''). The solid line indicates perfect matching between the two frequency sets (slope equal to 1). (b) Comparison between simulated (''×'') and measured (''○'') magnitudes of the main FFT peaks.

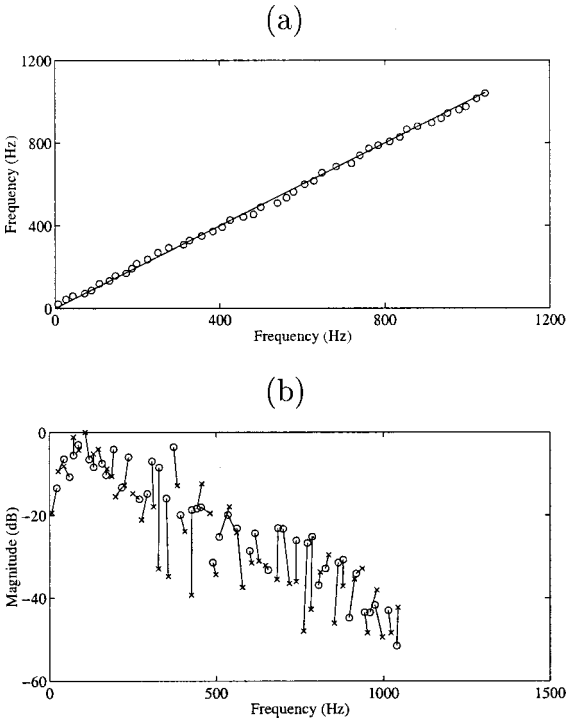


FIG. 8. Carbon plate. FFT analysis of the sound pressure during the first 20 ms. (a) Simulated vs measured frequencies in the range 0–1.5 kHz (''○''). The solid line indicates perfect matching between the two frequency sets (slope equal to 1). (b) Comparison between simulated (''×'') and measured (''○'') magnitudes of the main FFT peaks.

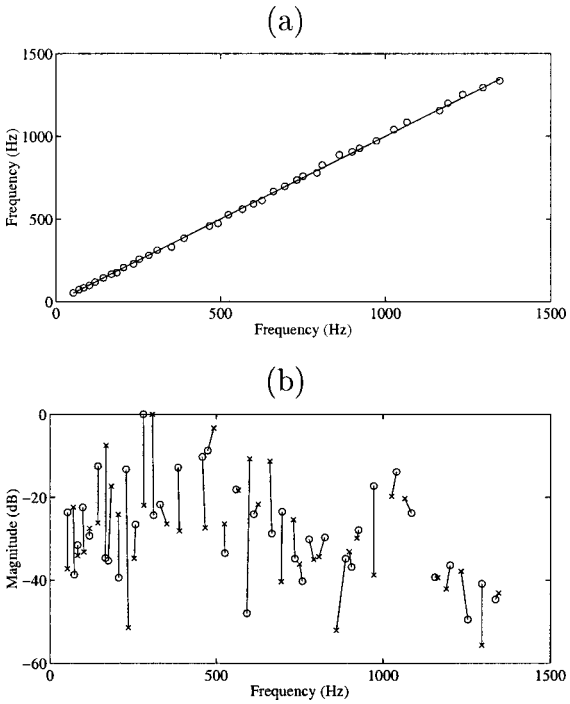


FIG. 9. Wooden plate. FFT analysis of the sound pressure during the first 20 ms. (a) Simulated vs measured frequencies in the range 0–1.5 kHz (''○''). The solid line indicates perfect matching between the two frequency sets (slope equal to 1). (b) Comparison between simulated (''×'') and measured (''○'') magnitudes of the main FFT peaks.

our finite-difference scheme [bottom of Fig. 3(a)]. Therefore, in order to better show the discrepancies between these two solutions, Fig. 3(b) shows the relative error between the two waveforms as a function of time for four different sampling frequencies. For a sampling frequency equal to 200 kHz, for example, it can be seen that the relative error remains always smaller than 5% and tends rapidly to less than 1%.

B. Boundary conditions

Figure 4 illustrates the ability of the numerical method to calculate the flexural vibrations of the plate under various boundary conditions: simply supported [Fig. 4(a)], clamped [Fig. 4(b)] and free plate [Fig. 4(c)]. This figure shows, in addition, that the scheme is able to simulate L-shaped plates, provided that the sides are parallel to the axes. Thus, the model is not restricted to simple rectangular plates.

C. Impact simulation: Comparison with another method

In her thesis, McMillan tackles similar problems of impacted plates, using a method based on Green functions of the vibrating structure at the impact point.[9] Figure 5 shows the simulated forces obtained with the two methods, using the same geometric and elastic parameters, in the case of a Carbon–Epoxy plate struck by a Titanium impactor (see Table I). The force histories obtained with the present finite-difference method are on the left-hand side of the figure, whereas the force histories obtained with the McMillan method are on the right-hand side. Figure 5(a) corresponds to an impact with initial velocity 1 m/s, whereas Fig. 5(b) corresponds to an impact velocity of nearly 0.6 m/s. It can be seen that the two methods yield identical waveforms. The ripples in the force pulses are due to the fact that the inverse of the lowest eigenfrequency of the plate here is about 20 times lower than the interaction time, a consequence of the particular selected set of parameters.

D. Comparison between measured and simulated pressure

The simulated sounds are now compared with measured sounds. Due to the large number of excited modes in each case, no quantitative information can be easily derived from the visual comparison between measured and simulated waveforms. In this respect, comparisons in the spectral domain are more informative. All spectra are calculated by FFT (fast Fourier transform) on the first 20 ms on the sound and normalized (in dB) with respect to the magnitude of the highest peak. In order to allow quantitative comparison between measured and simulated spectra, only the most significant peaks are displayed in the figures. These data are extracted from the FFT analysis by means of a simple peak detection algorithm. The values of the parameters used for the simulations can be found in Tables I–III.

Figure 6 shows the results obtained for an aluminum plate excited with a xylophone rubber mallet with initial velocity 1.45 m/s. Figure 6(a) shows that the number and frequencies of the spectral peaks, between 0 and 5 kHz, are very well reproduced. The spectral envelopes of both real and simulated sounds are comparable. The displayed spectral domain is limited here to 5 kHz for reasons of clarity, and also because the magnitude of the peaks above 5 kHz are relatively small (less than -40 dB below the maximum), a consequence of both excitation spectrum and damping. The damping here is mainly due to thermoelastic and radiation losses.[1]

Figure 7 shows the results obtained for a glass plate. Here again, the measured and simulated frequencies and magnitudes of the spectral peaks look very similar, which is confirmed by listening to the corresponding sounds. The losses in this material are mainly due to viscoelasticity and radiation.[1]

A first example of impact against an orthotropic plate is given in Fig. 8 which compares measurements with simulations in the case of a plate made of carbon fibers. It can be seen on the spectra that the high frequencies are rapidly damped: the magnitude of the peaks are less than -40 dB below the maximum for frequencies above 1.2 kHz. Here, the impact produces a duller sound than with aluminum and glass. The damping is almost entirely due to viscoelastic losses in the material, and the radiation losses can be neglected.[1] Similar conclusions can be drawn from Fig. 9 which shows the results obtained in the case of wood (Spruce). Here, the damping of the vibrations in the plate is mainly due to viscoelastic losses. However, due to the dimensions of the plate, the radiation losses cannot be totally neglected. The discrepancies between measurements and simulations for some of the peaks are probably due to the approximations made in the radiation model.

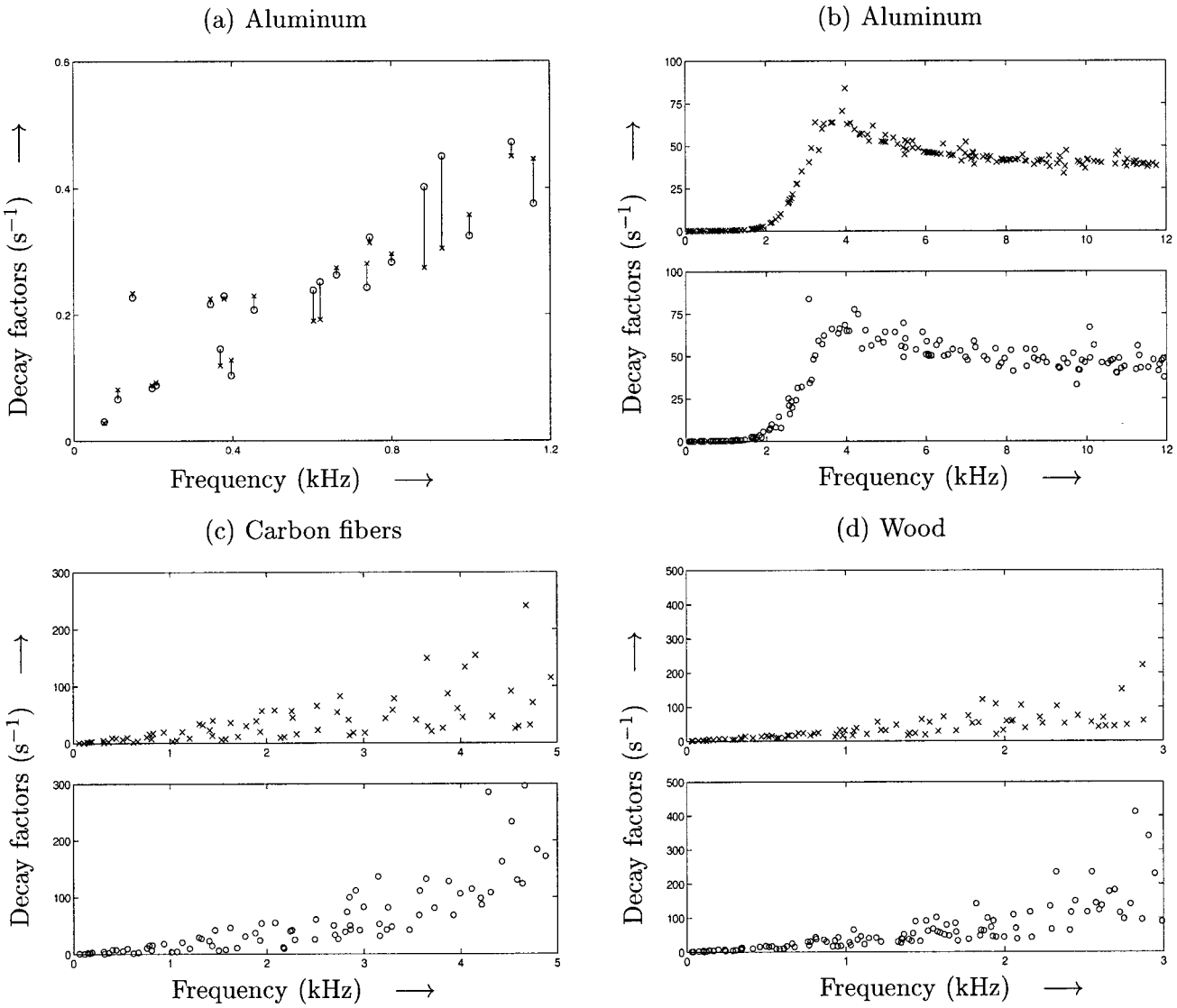


FIG. 10. Decay factors. (a) Aluminum plate. Range 0–1.2 kHz. (''○'') Experiments; (''×'') simulations. (b) Aluminum plate. Range 0–12 kHz. Top: simulations; bottom: measurements. (c) Carbon plate. Range 0–5 kHz. Top: simulations; bottom: measurements. (d) Wooden plate. Range 0–3 kHz. Top: simulations; bottom: measurements.

In a previous paper, measured decay factors were compared to the values predicted by the continuous plate model solved in the frequency domain.[1] To completely assess the validity of the method, Fig. 10 shows, in addition, the comparison between measured and simulated decay factors, where the simulated damping factors are now obtained from a frequency analysis of the simulated waveforms. The analysis method used here is the Matrix Pencil method.[17] Figure 10(a) shows that the thermoelastic model of losses accounts for the apparent erratic distribution of the decay factors in the low-frequency range for metallic plates. This distribution essentially follows from the fact that the decay time of each eigenfrequency depends on its modal shape. In this frequency range (below 2 kHz, for the plate shown on this figure) the radiation damping is negligible. The low values of the decay factors below the critical frequency play a major role in duration and tone quality of the sound generated by the free vibrations of the plate. Figure 10(b) shows the relevance of the asymptotic radiation model which accounts for the losses above the critical frequency of the plate. For this plate (aluminum a_3, see parameters in Tables I and II), the average decay factor above 4 kHz is about 100 times the average decay factor below 1 kHz.

Figures 10(c) and (d) show the relevance of the viscoelastic model for orthotropic plates. It can be seen that the viscoelastic losses fully account for the measured decay factors, except for the wooden plate above 2 kHz where the radiation losses are not negligible. Here again, the model is able to reproduce the apparent erratic distribution of decay factors with frequency. This distribution is not due to error in the measurements (a commonly erroneous conclusion) but rather to the fact that there is a specific damping model for each complex rigidity (see Table II). Notice further that 13 damping parameters only are needed here for reproducing adequately the temporal evolution of more than 40 modes. This is an interesting feature of the method in terms of data reduction.

V. CONCLUSION

A time-domain model for the simulation of flexural vibrations and sounds radiated by damped isotropic and orthotropic plates has been presented. In order to demonstrate the particular interest of the method for the synthesis of transients, the study was made for plates excited by an impact.

The key point of the model is that the losses in the plates are modeled in the time domain with the help of a general differential operator applied to the rigidities of the material. This formulation gives great flexibility for modeling the damping of plates. As shown in a companion paper, this formulation allows the modeling of the three main mechanisms of damping in plates: viscoelasticity, thermoelasticity, and radiation.[1] Therefore, it makes it possible to reproduce adequately the complicated frequency dependence of the decay times observed in a large variety of materials with only a small number of damping parameters.

The sounds synthesized with this model allow a clear recognition of the materials without ambiguity. This accurate reproduction would not be possible with a simple description of the damping, as with a single loss factor, for example. All measured and simulated sounds presented in this paper can be heard at the following Web address: `http://wwwy.ensta.fr/~chaigne/plaque/compar_exp_sim.html`.

The equations are solved numerically by means of a finite-difference scheme of second order in time and fourth order in space (2–4 scheme). The presence of complex rigidities makes it necessary to investigate thoroughly the stability of the scheme, which contributes to filling a gap in the literature. It has been shown that a weak stability condition can be obtained from the study of the amplification matrix, following the method by Richtmyer and Morton.[14]

The accuracy of the method is studied in terms of numerical dispersion. A comparison with another scheme of lower order in space (2–2 scheme) shows that, for a given accuracy, the required computing time with the selected 2–4 scheme is about 4 times lower than with the 2–2 scheme.

In order to show the validity of the method, a comparison is made first between an exact analytical solution and the numerical result obtained for an infinite undamped isotropic plate. The impact model is further validated by comparing the calculated force pulses imparted to the plate with the force pulses obtained with another method.[9] Finally, simulated sounds are systematically compared to measurements for four different materials. These comparisons show a particularly good agreement in terms of frequencies and damping factors despite the relatively low number of damping parameters inserted in the model.

The present model is limited to rectangular plates and to plates with rectangular corners (see Fig. 4), for which the use of finite differences is particularly well-suited. Extension of this model to more complicated shapes makes it necessary to use other numerical methods, such as finite elements based on a variational formulation of the problem.[18]

APPENDIX A: FINITE-DIFFERENCE OPERATORS

The discrete operators given below are obtained by linear combinations of Taylor expansions of the field variable $v(x,y,t)$. In this paper, $(D_{t^p}^{(i)}v)_{l,m}^n$ denotes the approximation of the pth time derivative of v at point $(l\Delta x, m\Delta y, n\Delta t)$, where the subscripts (i) is the order of the truncation error in Δt.

1. First order decentered operators

The following operators are used to approximate the time-derivatives in the damping terms.

$$(D_t^{(1)}v)^n=\frac{1}{\Delta t}(v^n-v^{n-1})=(v_{l,m}^n)_{,t}+O(\Delta t)$$

$$(D_{tt}^{(1)}v)^n=(D_t^{(1)}oD_t^{(1)}v)^n$$

$$=\frac{1}{\Delta t^2}(v^n-2v^{n-1}+v_{n-2})=(v_{l,m}^n)_{,tt}+O(\Delta t) \tag{A1}$$

$$(D_{t^3}^{(1)}v)^n=(D_t^{(1)}oD_t^{(1)}oD_t^{(1)}v)^n$$

$$=\frac{1}{\Delta t^3}(v^n-3v^{n-1}+3v^{n-2}-v^{n-3})$$

$$=(v_{l,m}^n)_{,ttt}+O(\Delta t).$$

$$\vdots$$

2. Second order centered operators

$$(D_{tt}^{(2)}v)_{l,m}^n=\frac{1}{\Delta t^2}(v_{l,m}^{n+1}+v_{l,m}^{n-1}-2v_{l,m}^n)$$

$$=(v_{l,m}^n)_{,tt}+O(\Delta t^2), \tag{A2}$$

$$(D_{xx}^{(2)}v)_{l,m}^n=\frac{1}{\Delta x^2}(v_{l+1,m}^n+v_{l-1,m}^n-2v_{l,m}^n)$$

$$=(v_{l,m}^n)_{,xx}+O(\Delta x^2), \tag{A3}$$

$$(D_{yy}^{(2)}v)_{l,m}^n=\frac{1}{\Delta y^2}(v_{l,m+1}^n+v_{l,m-1}^n-2v_{l,m}^n)$$

$$=(v_{l,m}^n)_{,yy}+O(\Delta y^2), \tag{A4}$$

$$(D_{xy}^{(2)}v)_{l+1/2,m+1/2}^n=\frac{1}{\Delta x\Delta y}(v_{l+1,m+1}^n+v_{l,m}^n$$

$$-v_{l+1,m}^n-v_{l,m+1}^n),$$

$$=(v_{l+1/2,m+1/2}^n)_{,xy}+O(\Delta x^2)+O(\Delta y^2), \tag{A5}$$

$$(\tilde{D}_{xy}^{(2)}v)_{l,m}^n=\frac{1}{\Delta x\Delta y}(v_{l+1/2,m+1/2}^n+v_{l-1/2,m-1/2}^n$$

$$-v_{l+1/2,m-1/2}^n-v_{l-1/2,m+1/2}^n)$$

$$=(v_{l,m}^n)_{,xy}+O(\Delta x^2)+O(\Delta y^2). \tag{A6}$$

3. Fourth order centered operators

$$(D_{xx}^{(4)}v)_{l,m}^{n}=\frac{1}{12\Delta x^2}[-(v_{l+2,m}^{n}+v_{l-2,m}^{n})+16(v_{l+1,m}^{n}+v_{l-1,m}^{n})-30v_{l,m}^{n}]=(v_{l,m}^{n})_{,xx}+O(\Delta x^4), \tag{A7}$$

$$(D_{yy}^{(4)}v)_{l,m}^{n}=\frac{1}{12\Delta y^2}[-(v_{l,m+2}^{n}+v_{l,m-2}^{n})+16(v_{l,m+1}^{n}+v_{l,m-1}^{n})-30v_{l,m}^{n}]=(v_{l,m}^{n})_{,yy}+O(\Delta y^4) \tag{A8}$$

$$(D_{xy}^{(4)}v)_{l+1/2,m+1/2}^{n}=\frac{1}{24^2\Delta x\Delta y}[(v_{l-1,m-1}^{n}+v_{l+2,m+2}^{n}-v_{l-1,m+2}^{n}-v_{l+1,m-1}^{n})+27(v_{l-1,m+1}^{n}+v_{l,m+2}^{n}+v_{l+1,m-1}^{n}+v_{l+2,m}^{n})-27(v_{l-1,m}^{n}+v_{l,m-1}^{n}+v_{l+1,m+2}^{n}+v_{l+2,m+1}^{n})+27^2(v_{l,m}^{n}+v_{l+1,m+1}^{n}-v_{l+1,m}^{n}-v_{l,m+1}^{n})]=(v_{l+1/2,m+1/2}^{n})_{,xy}+O(\Delta x^4)+O(\Delta y^4) \tag{A9}$$

$$(\tilde{D}_{xy}^{(4)}v)_{l,m}^{n}=\frac{1}{24^2\Delta x\Delta y}[(v_{l-3/2,m-3/2}^{n}+v_{l+3/2,m+3/2}^{n}-v_{l+3/2,m-3/2}^{n}-v_{l-3/2,m+3/2}^{n})+27(v_{l-3/2,m+1/2}^{n}+v_{l-1/2,m+3/2}^{n}+v_{l+1/2,m-3/2}^{n}+v_{l+3/2,m-1/2}^{n})-27(v_{l-3/2,m-1/2}^{n}+v_{l-1/2,m-3/2}^{n}+v_{l+1/2,m+3/2}^{n}+v_{l+3/2,m+1/2}^{n})+27^2(v_{l-1/2,m-1/2}^{n}+v_{l+1/2,m+1/2}^{n}-v_{l+1/2,m-1/2}^{n}-v_{l-1/2,m+1/2}^{n})]=(v_{l,m}^{n})_{,xy}+O(\Delta x^4)+O(\Delta y^4). \tag{A10}$$

APPENDIX B: STABILITY

1. 1D problem

By introducing the spatial Fourier transforms of both the displacement $w_n \exp(jkl\Delta x)$ and bending moment $(m_x)_n \exp(jkl\Delta x)$, Eq. (19) can be rewritten in the following matrix form:

$$\mathbf{v_{n+1}}=\tilde{\mathbf{G}}(\mathbf{\Delta t},\mathbf{k})\mathbf{v_n}, \tag{B1}$$

with

$$\mathbf{v_{n+1}}=[w_{n+1},w_n,\cdots,w_{n-N+1},(n_x)_n,\cdots,(n_x)_{n-N+1}]^T, \tag{B2}$$

where $(n_x)_n=\Delta x^2(m_x)_n$. The amplification matrix $\tilde{G}(\Delta t,k)$ is given by:

$$\begin{pmatrix} A & -BC\chi_2 & \cdots & \cdots & -BC\chi_N & C\psi_1 & \cdots & \cdots & \cdots & C\psi_N \\ 1 & 0 & \cdots & \cdots & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & \vdots & & & & \vdots \\ \vdots & & \ddots & & \vdots & \vdots & & & & \vdots \\ 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ B\chi_0 & \cdots & \cdots & \cdots & B\chi_N & -\psi_1 & \cdots & \cdots & \cdots & -\psi_N \\ 0 & \cdots & \cdots & \cdots & 0 & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & & & & \vdots & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & \vdots & \vdots & & \ddots & & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & 0 & \cdots & 0 & 1 & 0 \end{pmatrix}, \tag{B3}$$

with

$$P=2-R_f\Delta t,\quad B=4IEX^2,\quad C=\frac{4X^2\Delta t^2}{\rho S\Delta x^4},\quad F=R_f\Delta t-1,\quad X=\sin\frac{k\Delta x}{2},\quad A=(P-BC\chi_0)(F-BC\chi_1). \tag{B4}$$

The discrete system in Eq. (B1) is called *weakly stable* if there exists a real constant $K(\tau,T)$ such that:[14]

$$\|\tilde{\mathbf{G}}(\mathbf{\Delta t},\mathbf{k})^{\mathbf{n}}\| \leqslant K(\tau,T), \quad \text{for all} \quad \begin{cases} 0<\Delta t<\tau \\ 0\leqslant n\Delta t\leqslant T, \\ k \quad \text{real} \end{cases} \tag{B5}$$

where $\|\tilde{G}(\Delta t,k)\|$ denotes the spectral norm of the matrix $\tilde{G}(\Delta t,k)$. A condition for strong stability is obtained by replacing $K(\tau,T)$ by 1 in Eq. (B5). The following criterion has been applied here:[14]

> ''If $\tilde{G}(\Delta t,k)$ is uniformly Lipschitz continuous for $\Delta t=0$, so far as $\tilde{G}(\Delta t,k)=\tilde{G}(0,k)+O(\Delta t)$, as $\Delta t\to 0$, where the constant implied by the expression $O(\Delta t)$ does not depend on k, then the scheme is weakly stable if and only if $\tilde{G}(0,k)$ is stable.''

To apply this criterion, the first step consists of studying the dependence of $\tilde{G}(\Delta t,k)$ on Δt. This parameter appears in the coefficients A, C and F defined in Eq. (B4) and in the expressions of the coefficients χ_r and ψ_r:

$$\begin{aligned} &\chi_r=(-1)^r\frac{p_N}{q_N}+O(\Delta t) \\ &\psi_r=(-1)^r+O(\Delta t). \end{aligned} \tag{B6}$$

All the coefficients of the amplification matrix can be expressed as $g_{ij}+O(\Delta t)$ and thus $\tilde{G}$ is uniformly Lipschitz continuous. The determination of a stability condition now consists of the study of $\tilde{G}(0,k)$. It must be checked whether the spectral radius of this matrix remains smaller than or equal to unity for all k or not. The eigenvalues λ_i of $\tilde{G}(0,k)$ are such that the vector:

$$\begin{pmatrix} (n_x)_n \\ w_n \end{pmatrix}=\lambda_i\begin{pmatrix} (n_x)_{n-1} \\ w_{n-1} \end{pmatrix} \tag{B7}$$

is a solution of the discrete system in Eq. (19) for $\Delta t=0$, given by:

$$\begin{pmatrix} \displaystyle\sum_{r=1}^{N}(-1)^r\lambda_i^{-r} & \displaystyle -B\frac{p_N}{q_N}\sum_{r=0}^{N}(-1)^r\lambda_i^{-r} \\ C & \lambda_i-2+\lambda_i^{-1} \end{pmatrix}\begin{pmatrix} (n_x)_n \\ w_n \end{pmatrix}=\begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{B8}$$

The nontrivial solutions of this system are obtained by zeroing the determinant of the matrix:

$$\left(\sum_{r=0}^{N}(-1)^r\lambda_i^r\right)\left(\lambda_i^2+\left(\frac{X^4}{\xi_s}\frac{IEp_N}{\rho Sq_N}-2\right)\lambda_i+1\right)=0 \tag{B9}$$

from which the following necessary condition for weak stability is obtained:

$$\Delta x^2\geqslant 2\Delta t\sqrt{\frac{IEp_N}{\rho Sq_N}}. \tag{B10}$$

2. 2D problem

A necessary condition for the system Eq. (9) to be stable is now determined in the same way as for the 1D case. It is assumed that the dissipativity conditions on the coefficients p_{iN} and q_N are fulfilled (see Ref. 1). As Δt tends to zero, χ_{ir} and ψ_{ir} can be expanded as follows:

$$\begin{aligned} &\chi_{ir}=(-1)^r\frac{p_{iN}}{q_N}+O(\Delta t), \quad 1\leqslant i\leqslant 4 \\ &\psi_{ir}=(-1)^r+O(\Delta t), \quad i=[1,4]. \end{aligned} \tag{B11}$$

As for the 1D case, it can be shown that the amplification matrix is Lipschitz-continuous for $\Delta t=0$. The eigenvalues λ_i of $\tilde{G}(0,k_x,k_y)$ are now solutions of the following system:

$$\begin{pmatrix} \mathcal{Q}(\lambda_i) & 0 & 0 & -(\mathcal{P}_1(\lambda_i)-\mathcal{P}_{21}(\lambda_i)) \\ 0 & \mathcal{Q}(\lambda_i) & 0 & -(\mathcal{P}_3(\lambda_i)-\mathcal{P}_{23}(\lambda_i)) \\ 0 & 0 & \mathcal{Q}(\lambda_i) & -\mathcal{P}_4(\lambda_i) \\ C_1 & C_3 & C_4 & \mathcal{P}(\lambda_i) \end{pmatrix} \cdot\begin{pmatrix} (n_x)_n \\ (n_y)_n \\ (n_{xy})_n \\ w_n \end{pmatrix}=\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{B12}$$

with:

$$\begin{aligned} &\mathcal{Q}(\lambda_i)=\sum_{r=0}^{N}(-1)^r\lambda_i^{-r}, \quad \mathcal{P}_{23}(\lambda_i)=2h^3D_2X^2\frac{p_{2N}}{q_N}\mathcal{Q}(\lambda_i), \\ &C_1=\frac{4X^2}{\rho h\xi_s^2}, \quad \mathcal{P}(\lambda_i)=\lambda_i-2+\lambda_i^{-1}, \\ &\mathcal{P}_3(\lambda_i)=\frac{4h^3D_3Y^2}{r_1^2}\frac{p_{3N}}{q_N}\mathcal{Q}(\lambda_i), \quad C_3=\frac{4Y^2}{\rho h\xi_s^2r_1^2}, \\ &\mathcal{P}_1(\lambda_i)=4h^3D_1X^2\frac{p_{1N}}{q_N}\mathcal{Q}(\lambda_i), \\ &\mathcal{P}_4(\lambda_i)=\frac{2h^3D_4XY}{r_1}\frac{p_{4N}}{q_N}\mathcal{Q}(\lambda_i), \quad C_4=\frac{8XY}{\rho h\xi_s^2r_1}, \\ &\mathcal{P}_{21}(\lambda_i)=\frac{2h^3D_2Y^2}{r_1^2}\frac{p_{2N}}{q_N}\mathcal{Q}(\lambda_i), \quad \xi_s=\frac{\Delta x^2}{\Delta t}, \quad r_1=\frac{\Delta y}{\Delta x}. \end{aligned} \tag{B13}$$

As for the 1D case, the determination of the stability condition is obtained by zeroing the determinant of the matrix Eq. (B12) which amounts to studying the roots of the polynomial:

$$\begin{aligned} &\mathcal{P}(\lambda)+C_1\left(A_1\frac{p_{1N}}{q_N}+A_{21}\frac{p_{2N}}{q_N}\right)+C_3\left(A_3\frac{p_{3N}}{q_N}+A_{23}\frac{p_{2N}}{q_N}\right) \\ &\quad +C_4A_4\frac{p_{4N}}{q_N}=0, \end{aligned} \tag{B14}$$

with

$$A_1=h^3D_1X, \quad Y=\sin\frac{k\Delta y}{2}, \quad A_{21}=\frac{h^3D_2Y^2}{2r_1^2}, \tag{B15}$$

$$A_3=\frac{h^3D_3Y^2}{r_1^2},\quad A_{23}=\frac{h^3D_2X^2}{2},\quad A_4=\frac{h_3D_4XY}{2r_1}.$$

This yields a stability condition similar to Eq. (17), where the D_i are replaced by $D_i p_{iN}/q_N$.

[1] A. Chaigne and C. Lambourg, ''Time-domain simulation of damped impacted plates. Part I. Theory and experiments,'' J. Acoust. Soc. Am. **109**, 1422–1432 (2001).

[2] A. Frendi, L. Maestrello, and A. Bayliss, ''Coupling between plate vibration and acoustic radiation,'' J. Sound Vib. **177**, 207–226 (1994).

[3] S. Schedin, C. Lambourg, and A. Chaigne, ''Transient sound fields from impacted plates: Comparison between numerical simulations and experiments,'' J. Sound Vib. **221**, 471–490 (1999).

[4] V. Doutaut, D. Matignon, and A. Chaigne, ''Numerical simulations of xylophones. II. Time-domain modeling of the resonator and of the radiated sound pressure,'' J. Acoust. Soc. Am. **104**, 1633–1647 (1998).

[5] A. W. Leissa, *Vibrations of Plates*, NASA SP-160 (NASA, Washington, D.C., 1969).

[6] K. F. Graff, *Wave Motion in Elastics Solids* (Dover, New York, 1991).

[7] A. E. H. Love, *A Treatise on the Mathematical Theory of Elasticity* (Dover, New York, 1944).

[8] A. Chaigne and V. Doutaut, ''Numerical simulations of xylophones. I. Time-domain modeling of the vibrating bars,'' J. Acoust. Soc. Am. **101**, 539–557 (1996).

[9] A. J. McMillan, ''A Theoretical Investigation of the Role of Vibration in Elastic Impact,'' Ph.D. thesis, Staffordshire Polytechnic, 1992.

[10] W. F. Ames, *Numerical Methods for Partial Differential Equations*, 3rd ed. (Academic, New York, 1992).

[11] S. J. Fenves, N. Perrone, A. R. Robinson, and W. C. Schnobrich, *Numerical and Computer Methods in Structural Mechanics* (Academic, New York, 1973).

[12] C. J. Moore, *An Introduction to the Psychology of Hearing*, 2nd ed. (Academic, New York, 1982).

[13] C. Lambourg, ''Modèle temporel pour la simulation numérique de plaques vibrantes. Application à la synthèse sonore,'' Ph.D. thesis, Université du Maine, 1997.

[14] R. D. Richtmyer and K. W. Morton, *Difference Methods for Initial-Value Problems*, 2nd ed. (Interscience, New York, 1967).

[15] G. Cohen, Ed., *Ecole des Ondes Inria: Méthodes numériques d'ordre élevé pour les ondes en régime transitoire* (Collection Didactique, INRIA, 1994).

[16] R. F. S. Hearmon, *An Introduction to Applied Anisotropic Elasticity* (Oxford University Press, New York, 1961).

[17] J. Laroche, ''The use of the matrix pencil method for the spectrum analysis of musical signals,'' J. Acoust. Soc. Am. **94**, 1958–1965 (1993).

[18] L. Rhaouti, A. Chaigne, and P. Joly, ''Time-domain modeling and numerical simulation of a kettledrum,'' J. Acoust. Soc. Am. **105**, 3545–3562 (1999).

The influence of substructure modeling on the structural-acoustic response of a plate system

W. Steve Shepard, Jr.[a)]
Department of Mechanical Engineering, The University of Alabama, Box 870276, Tuscaloosa, Alabama 35487

Kenneth A. Cunefare
George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 24 April 2000; revised 8 January 2001; accepted 10 January 2001)

Changes in the vibro-acoustic response of a fluid-loaded plate due to variations in some of the modeling details associated with an attached substructure are examined. The attached substructure consists of a smaller plate supported by springs along each edge. To examine the important modeling issues, three studies are performed. In the first study, discrete changes in the system response due to discrete changes in the size of the region over which the spring elasticity is distributed are examined. In the second study, substructure modeling issues are examined by varying the number of degrees-of-freedom included in the substructure model. Finally, sensitivity relationships that express changes in the system response to changes in the scale of the spring elements are developed. These relationships are used to examine changes in the system response due to small variations in the scale of the distributed elasticity. Both the combined system response and acoustic radiation are computed using the Acoustic Surface Variational Principle and Hamilton's Principle. For the example cases considered, it is shown that details associated with the scale of the spring are only important for frequencies near or below the resonances of the isolated subsystem. Furthermore, only the dynamics of the substructure including rigid-body type motions are important. [DOI: 10.1121/1.1353595]

PACS numbers: 43.40.Dx, 43.40.Rj [CBB]

I. INTRODUCTION

When predicting the vibro-acoustic behavior of complex systems, one must decide how much effort to spend on modeling the structural features with dimensions smaller than those of the primary structure. An example of such a system is one that contains a wave-bearing substructure as shown in Fig. 1. The isolator has an attachment width w, which is smaller than the dimensions of the primary and supported structures. Even for a linear isolator, one must decide if the size of the isolator is large enough to justify a greater modeling effort. That is, can the isolator be modeled using a lumped representation or must it be treated using some distributed representation? If a spatial representation is required, then the accuracy of the representation necessary to obtain a valid structural-acoustic prediction must be considered. Once the appropriate isolator modeling is determined, one must then decide how much effort to spend on modeling the supported substructure.

Several works have examined the impact of features attached to the primary structure at a mathematical point (see for example, Refs. 1–3). The impact of a feature distributed over some region of the primary structure has also been considered.[4–6] Although these works did consider a distributed feature, the purpose of these works was not to examine the impact of various distribution widths (i.e., feature scales) on the system response. In more recent work, the impact of inertial scales for distributed masses[7,8] and elastic scales for distributed springs supporting a mass[9] has been addressed. The sole purpose of these works was to examine the impact of feature scales on the response of the system. Our goal in the work presented here is to extend these latter works to address issues associated with modeling a more complicated substructure. To accomplish this goal, three basic studies are performed.

As in the previous works, the system considered here is a fluid-loaded semi-infinite thin plate in a rigid infinite baffle. The semi-infinite plate has an infinite dimension in one coordinate and a finite dimension in the other coordinate. Unlike the previous works, the attached substructure considered here is a wave-bearing plate suspended by springs along its edges. The first study is concerned with scale effects of the spring elasticities. To examine scale effects, the springs are treated first as lumped elements then as distributed elasticities. By varying the distribution width, changes in the system response are examined. The basis for comparison is the response of the system when the springs are represented as lumped elements. In the second study, the impact of substructure modeling on the system response is examined by altering the dynamics included in the substructure model. The final analysis is concerned solely with the spatial characterization of the springs. Sensitivity relationships that express changes in system response to variations in the coupling scale are developed and used to determine when the response of the system is most sensitive to changes in the feature scales.

[a)]Electronic mail: sshepard@coe.eng.ua.edu

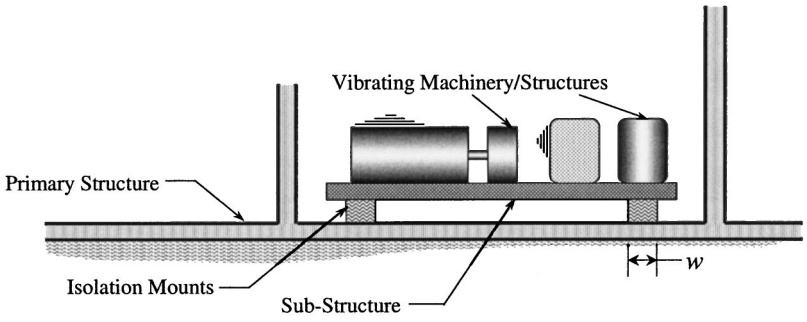


FIG. 1. Primary structure with wave-bearing substructure—Isolator of width *w*.

Although analytical solutions exist for semi-infinite plate systems carrying lumped elements, our purpose in this research is to investigate the effects of system parameters that are distributed over some finite region. Furthermore, it is desired to observe these effects within the context of a fluid-loaded system. As no closed-form solutions are readily available for systems of this type, the interrogation technique is principally numerical in nature. The techniques used to determine the response of the system are the acoustic surface variational principle (SVP)[10,11] and Hamilton's principle. The SVP is ideally suited to this research since it does not implement a surface discretization that introduces an additional set of spatial scales. Only an overview of the clean system equations similar to that presented by Cunefare *et al.*[12] is given here since detailed derivations can be found in another work.[7] Modifications of these equations to include the supported substructure are presented in a later section. Then, the development of the sensitivity relationships is presented. Finally, results for two example system configurations are discussed.

II. CLEAN SYSTEM MODEL

A section of the clean semi-infinite plate system is shown in Fig. 2. Note that first only the modeling for the main plate is considered. The nondimensional surface displacement for the main plate is expressed using an *M* term Ritz expansion,

$$w(\xi,t)=a\sum_{j=1}^{M}\phi_j(\xi)q_j(t),\quad \phi_i=\sin[i\pi(\xi+1/2)],$$
$$-\tfrac{1}{2}\leqslant\xi\leqslant\tfrac{1}{2}, \tag{1}$$

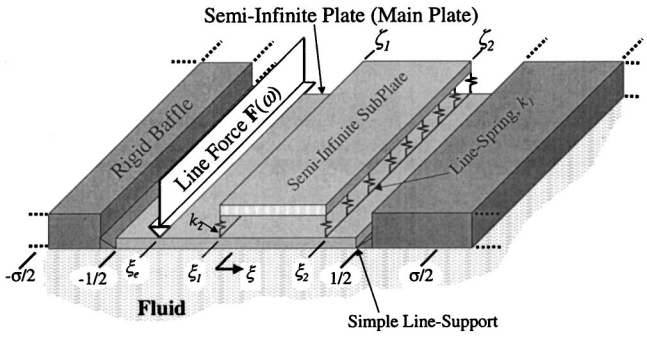


FIG. 2. Section of semi-infinite plate with elastically suspended wave-bearing plate.

where $\xi=x/a$ is the nondimensional position on the plate of width *a*. The nondimensional normal surface velocity is $v_n(\xi,t)=-ikaw(\xi,t)$, where $ka=a\omega/c$ is the nondimensional wavenumber in the fluid and *c* is the sound speed in the fluid. Note that the harmonic time dependence, $\exp(-i\omega t)$, has been dropped to simplify notation. The surface pressure is expressed using an *N* term expansion,

$$p(\xi)=\sum_{i=1}^{N}\psi_i(\xi)p_i,\quad \psi_j=\sin[j\pi(\xi/\sigma+1/2)],$$
$$-\frac{\sigma}{2}\leqslant\xi\leqslant\frac{\sigma}{2}, \tag{2}$$

where it is assumed the pressure vanishes beyond the distance $\xi=\sigma/2$ (see Refs. 13, 14) with $\sigma=5$. By applying the variational principle (SVP), a system of *N* equations is obtained relating the pressure coefficients, p_i, to the displacement coefficients, w_j. Furthermore, by applying Lagrange's equation to the potential energy, kinetic energy, and work expressions for the plate, *M* coupled equations relating the structural displacement to the surface pressure and excitation forces can be obtained.[7] The resulting $M+N$ square system of coupled equations for the bare plate is

$$\begin{bmatrix}[A] & -2\pi(ka)^2[R]^T\\ [R] & [D]\end{bmatrix}\begin{Bmatrix}\{p\}\\ \{w\}\end{Bmatrix}=\begin{Bmatrix}\{0\}\\ \{F^e\}\end{Bmatrix}, \tag{3}$$

where the superscript *T* refers to the matrix transpose. The $[A]$ matrix contains complex terms describing the fluid and the $[R]$ matrix includes purely real terms describing the fluid–structure coupling. Detailed expressions for these terms can be found in Ref. 7. The $[D]$ matrix describes the structure,[7]

$$D_{jk}=\frac{\rho_s h}{\rho a}\left(\left(\frac{c_s}{c}\right)^2\kappa_{jk}-(ka)^2\mu_{jk}\right),$$
$$\mu_{jl}=\int_{-1/2}^{1/2}\phi_j\phi_l\,d\xi,\quad \kappa_{jl}=\int_{-1/2}^{1/2}\phi_j''\phi_l''\,d\xi, \tag{4}$$

where ρ is the fluid density. Using the subscript *s* to refer to steel, the mass per unit surface area of the plate is $\rho_s h$, where *h* is the plate thickness and ρ_s is the plate density. The characteristic wave speed for the structure, $c_s^2=D/\rho_s h a^2$, depends on the bending stiffness of the plate, $D=Eh^3/(12(1-\nu^2))$, where *E* is the elastic modulus and ν is Poisson's ratio. The properties of the steel plate, which will be considered in the examples presented below, are those of a plate submerged in seawater,[15]

$$a/h=25, \quad \rho_s/\rho=7.58, \quad c_s/c=3.84(h^2/12)^{1/2}.$$

For the work presented here, the harmonic excitation is a line force of magnitude f located at ξ_e, that is, $f(\xi)=f\delta(\xi-\xi_e)$. The excitation is incorporated into Eq. (3) as

$$F_j=\int_{-1/2}^{1/2} f(\xi)\phi_j\, d\xi. \tag{5}$$

By solving Eq. (3) for the modal coefficients, p_i and w_j, the surface displacement and pressure can then be constructed from Eqs. (1) and (2), respectively. Furthermore, the nondimensional radiated acoustic power per unit depth can be computed as

$$\wp_{\mathrm{rad}}=\tfrac{1}{2}\,\mathrm{Re}\int_{-1/2}^{1/2} p(\xi)\,v(\xi)^*\,d\xi$$

$$=\tfrac{1}{2}\,\mathrm{Re}(-ika\{p\}^T[R]^T\{q\}^*), \tag{6}$$

where * denotes the complex conjugate. The energy in the evanescent field can be found by taking the imaginary part instead of the real part (see Ref. 16, p. 133). If the structure has no internal structural damping mechanisms, the radiated acoustic power is equal to the mechanical power input,

$$\wp_{\mathrm{rad}}=\tfrac{1}{2}\,\mathrm{Re}(\{F^e\}^T\{v^*\}). \tag{7}$$

Before describing the substructure modeling, it should be noted that the only relationships that depend directly on the presence of any substructures are Eqs. (3) and (4).

III. SUBSTRUCTURE MODELING

As discussed above, the attached substructure consists of a semi-infinite wave-bearing plate suspended by two line-springs as shown in Fig. 2. By varying the size, or scale, of the distribution for the springs, the impact of coupling scales can be examined. The term ''coupling scales'' is derived from the fact that two separate regions of the main plate are coupled via the springs and the suspended plate, which is hereafter referred to as the subplate. The modeling of the subplate, which is always smaller than the main plate, is incorporated using Hamilton's principle under a unified structure approach.

The motion of the subplate, z_{pl}, is represented using the expansion

$$z_{\mathrm{pl}}(\zeta,t)=\sum_{i=1}^{M_P}\Omega_i(\zeta)z_i(t), \tag{8}$$

where M_P is the truncated number of assumed modes and ζ is the nondimensional coordinate of the subplate, $-1/2\leqslant\zeta\leqslant 1/2$. The added degrees of freedom are denoted by $\{z\}$. At this point, it is convenient to note that the number of terms used in Eq. (8) defines the modeling details associated with the dynamics of the subplate. The width of the subplate is L_p and the thickness is h_p, and the subscript p is used to denote properties related to the subplate.

Since there are no geometric boundary conditions for the subplate, the free-free normalized orthogonal modes[17] are used as the basis functions:

$$\Omega_1(\zeta)=1,$$

$$\Omega_2(\zeta)=\sqrt{3}\zeta, \tag{9}$$

$$\Omega_i(\zeta)=\cos(\beta_i(\zeta+1/2))+\cosh(\beta_i(\zeta+1/2))$$

$$-C[\sin(\beta_i(\zeta+1/2))+\sinh(\beta_i(\zeta+1/2))],$$

$$i\geqslant 3,$$

where

$$C=\frac{\cos(\beta_i)-\cosh(\beta_i)}{\sin(\beta_i)-\sinh(\beta_i)}.$$

The values of β_i are the solutions to the eigenvalue equation, $\cos\beta_i\cosh\beta_i=1$, noting that $\beta_1=\beta_2=0$. The functions Ω_1 and Ω_2 are for pure translation and pure rotation, respectively. As a result, if only two terms are used in Eq. (8), then only rigid-body type motions are allowed in the subsystem.

While the springs are only attached to the endpoints of the subplate, they may be distributed over the main plate. For such a case, the two distributed springs are represented by the spatial distributions,

$$k_1(\xi)=k_{\mathrm{rat1}}\varphi_1(\xi) \quad \text{and} \quad k_2(\xi)=k_{\mathrm{rat2}}\varphi_2(\xi), \tag{10}$$

where $\varphi_i(\xi)$ is a spatial distribution defined over some region of the main plate and k_{rat} is the total added stiffness divided by the bending stiffness of the main plate,

$$k_{\mathrm{rat}}=\frac{a^4k_{\mathrm{added}}}{D}. \tag{11}$$

A detailed description of the spatial distributions for the springs, $\varphi_i(\xi)$, is provided in the next section.

Taking into account these attached structural features and applying Hamilton's Principle to the appropriately modified potential and kinetic energy expressions, the system of equations given in Eq. (3) can be modified. The combined system equations describing both the plate and subplate response are given by

$$\begin{bmatrix}[A]-2\pi(ka)^2[R]^T & [0] & \\ [R] & [D_E] & [K]^T \\ [0] & [K] & [S]\end{bmatrix}\begin{Bmatrix}\{p\}\\ \{w\}\\ \{z\}\end{Bmatrix}=\begin{Bmatrix}\{0\}\\ \{F^e\}\\ \{0\}\end{Bmatrix}. \tag{12}$$

Since the springs are included in the potential energy terms for the system, the relationship for D_{jk} in Eq. (4) must be modified as

$$D_{Ejk}=\frac{\rho_s h}{\rho a}\left(\left(\frac{c_s}{c}\right)^2\kappa_{jk}-(ka)^2\mu_{jk}+\left(\frac{c_s}{c}\right)^2\alpha_{jk}^A\right), \tag{13}$$

where

$$\alpha_{jk}^A=\int_{-1/2}^{1/2}(k_1(\xi)+k_2(\xi))\phi_j(\xi)\phi_k(\xi)d\xi. \tag{14}$$

Furthermore, the structural coupling terms of $[K]$ are

$$K_j=-\frac{\rho_s h}{\rho a}\left(\frac{c_s}{c}\right)^2\alpha_j^B, \tag{15}$$

where

$$\alpha_{jl}^{B}=\int_{-1/2}^{1/2}[k_1(\xi)\Omega_l(\zeta_1)+k_2(\xi)\Omega_l(\zeta_2)]\phi_j(\xi)d\xi. \quad (16)$$

The remaining terms are

$$S_{jl}=\frac{\rho_s h}{\rho a}\left(\frac{c_s}{c}\right)^2\alpha_{jl}^{c}+\frac{\rho_s h}{\rho a}\frac{\rho_p}{\rho_s}\frac{h_p}{h}\times\left(\left(\frac{c_s}{c}\frac{c_p}{c_s}\right)^2\frac{h_p^2}{12(1-\nu^2)}\tau_{jl}-(ka)^2\gamma_{jl}\right), \quad (17)$$

where

$$\alpha_{jk}^{C}=\int_{-1/2}^{1/2}(k_1(\xi)\Omega_j(\zeta_1)\Omega_k(\zeta_1)+k_2(\xi)\Omega_j(\zeta_2)\Omega_k(\zeta_2))d\xi, \quad (18)$$

$$\tau_{jl}=\int_{-1/2}^{1/2}\Omega_j''\Omega_l''d(\zeta), \quad (19)$$

$$\gamma_{jl}=\frac{L_p}{a}\int_{-1/2}^{1/2}\Omega_j\Omega_l d(\zeta)+I_p[\Omega_j(\zeta_2)-\Omega_j(\zeta_1)]\times\Omega_l(\zeta_2)-\Omega_l(\zeta_1)], \quad (20)$$

and

$$I_p=\frac{1}{12}\frac{L_p}{a}\left[1+\left(\frac{h_p}{h}\frac{h}{a}\frac{a}{L_p}\right)^2\right]. \quad (21)$$

The ratios ρ_p/ρ_s, h_p/h, and c_p/c_s in the above equations relate the subplate properties (p) to the main plate properties (s). The length ratio L_p/a is fixed with regard to the spring attachment locations on the main plate, $L_p/a=\xi_2-\xi_1$, with $\xi_2>\xi_1$. While the second term of Eq. (20) accounts for the rotational inertia of the subplate, translational motion is accounted for in the first term of that equation. It should be noted that the values for $\Omega(\zeta)$ in Eqs. (16) and (18) are evaluated with

$$\zeta_1=\frac{-1}{2} \quad \text{and} \quad \zeta_2=\frac{1}{2}. \quad (22)$$

Once the system in Eq. (12) has been solved, the response of both the main plate and the subplate can be found by using Eqs. (1) and (8), respectively. The surface pressure on the main plate can similarly be found using Eq. (2) and the radiated power can be found using Eq. (6).

IV. SPRING SPATIAL REPRESENTATIONS

In order to provide a baseline for comparison purposes, the springs are initially modeled using lumped parameter representations. To that end, the lumped spring representations are[14]

$$k_1(\xi)=k_{\text{rat1}}\delta(\xi-\xi_1) \quad \text{and} \quad k_2(\xi)=k_{\text{rat2}}\delta(\xi-\xi_2). \quad (23)$$

As stated above, coupling scales are introduced by distributing the elasticity over a region of the main plate. A representation that has a physical counterpart and is used in this study is the chapeau representation given by

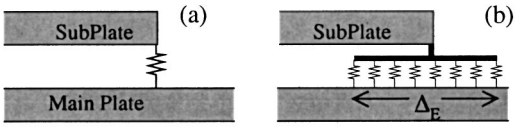


FIG. 3. Spring distributions: (a) Lumped representation, and (b) Chapeau representation.

$$k_1(\xi)=\frac{k_{\text{rat1}}}{\Delta_{E1}}[u(\xi-(\xi_1-\Delta_{E1}/2))-u(\xi-(\xi_1+\Delta_{E1}/2))], \quad (24)$$

where the width of the uniform distribution for spring number 1 is Δ_{E1} and $u(\xi-\xi_1)$ is the unit step function at ξ_1. The distribution for spring number 2 has a width Δ_{E2} that is centered about $\xi=\xi_2$. Note that when the scale (i.e., Δ_E) is changed, the total elasticity remains constant since it is only desired to examine scale effects in this work. An example of both the lumped representation and the chapeau representation is shown in Fig. 3. In using the chapeau representation, the distributed spring element is treated as a locally reacting component and only normal forces can be exerted on the plate by the distributed elasticity. Since the spring forces do not explicitly depend on rotations in the plates, the transmission of moments has been neglected. It is anticipated that future research will consider the effect of moments. It is also noted that another work considered a series representation of the spring elasticity.[18] The series representation, however, has physical limitations in its application and as such is not considered here.

V. SENSITIVITY OF SYSTEM RESPONSE TO COUPLING SCALES

In the first analysis presented in the next section, changes in the radiated acoustic power due to discrete changes in the feature scales are investigated. Before discussing those results, however, it is advantageous to discuss an alternate means of examining the impact of feature scales. In this approach, the sensitivity of the system response to changes in the coupling scale (i.e., Δ_E) is examined. In using the chapeau representation discussed above, it is possible to derive expressions for these sensitivities. A similar approach was used in a previous work to investigate the impact of inertial scales.[8] In addition to examining the problem at hand, the approach is extended here to demonstrate its applicability to more complicated systems. Before deriving the relationships, it is first beneficial to rearrange Eq. (12) in an impedance matrix form as

$$([Z_{\text{fl}}]+[D_E]-[K]^T[S]^{-1}[K])\{w\}=[Z_{\text{sys}}]\{w\}=\{F^e\}, \quad (25)$$

where

$$[Z_{\text{fl}}]=2\pi(ka)^2[R][A]^{-1}[R]^T, \quad (26)$$

$$[Z_{\text{sys}}]=[Z_{\text{fl}}]+[Z_{\text{st}}], \quad (27)$$

and

$$[Z_{\text{st}}]=[D_E]-[K]^T[S]^{-1}[K]. \quad (28)$$

TABLE I. Properties of subplate, springs, and system parameters.

Case No.	Stiffnesses and locations			Sub-plate properties				$M_p=2$ Res.	
	$k_{\text{rat1,2}}$	ξ_1	ξ_2	ρ_p/ρ_s	c_p/c_s	h_p/h	L_p/a	$(ka)_R$	$(ka)_T$
1	116	−0.373	0.027	1.0	1.0	0.625	0.40	0.7	1.4
2	1430	−0.123	0.207	0.35	1.0	2.157	0.33	2.4	5.0

To find the sensitivity of the various system variables to changes in the feature scale, it is first necessary to compute the partial derivative of $[Z_{\text{sys}}]$ with respect to Δ_E. Since the matrix $[Z_{\text{fl}}]$ does not depend on Δ_E, only the partial derivative of the structural impedance matrix $[Z_{\text{st}}]$ is needed. Since $\partial[C]^{-1}/\partial x=-[C]^{-1}(\partial[C]/\partial x)[C]^{-1}$ (see Ref. 19), it can be shown that

$$\frac{\partial[Z_{\text{st}}]}{\partial\Delta_e}=\frac{\partial[D_E]}{\partial\Delta_e}-\frac{\partial[K]^T}{\partial\Delta_e}[S]^{-1}[K]-[K]^T[S]^{-1}\frac{\partial[K]}{\partial\Delta_e}+[K]^T[S]^{-1}\frac{\partial[S]}{\partial\Delta_e}[S]^{-1}[K]. \tag{29}$$

The individual terms of this relation are

$$\frac{\partial[D_E]}{\partial\Delta_e}=\frac{\rho_s h}{\rho a}\left(\frac{c_s}{c}\right)^2\frac{\partial[\alpha^A]}{\partial\Delta_e}, \tag{30}$$

$$\frac{\partial[K]}{\partial\Delta_e}=-\frac{\rho_s h}{\rho a}\left(\frac{c_s}{c}\right)^2\frac{\partial[\alpha^B]}{\partial\Delta_e}, \tag{31}$$

and

$$\frac{\partial[S]}{\partial\Delta_e}=\frac{\rho_s h}{\rho a}\left(\frac{c_s}{c}\right)^2\frac{\partial[\alpha^C]}{\partial\Delta_e}. \tag{32}$$

Given the definition of α^C in Eq. (18), it can be shown that $\partial[S]/\partial\Delta_E=0$. Consequently, the last term of Eq. (29) has a value of zero. It should also be noted that since there are two springs in the system, there are two distribution widths. The term for α^B, for example, would be computed as

$$\frac{\partial\alpha^B_{jl}}{\partial\Delta_E}=\Omega_l(\zeta_1)\frac{\partial}{\partial\Delta_{E1}}\left(\int_{-1/2}^{1/2}k_1(\xi)\phi_j(\xi)d\xi\right)+\Omega_l(\zeta_2)\frac{\partial}{\partial\Delta_{E2}}\left(\int_{-1/2}^{1/2}k_2(\xi)\phi_j(\xi)d\xi\right). \tag{33}$$

In the following studies which use this technique, it is assumed that $\Delta_{E1}=\Delta_{E2}$. As a result, there is no need to distinguish between the two widths when taking derivatives.

In summary, the resulting relations that express the sensitivity of surface pressure and radiated acoustic power to changes in the elastic scale are

$$\frac{\partial\{p\}}{\partial\Delta_E}=2\pi\frac{\rho_s h}{\rho a}(ka)^4[A]^{-1}[R]^T[Z_{\text{sys}}]^{-1}\times\frac{\partial[Z_{\text{st}}]}{\partial\Delta_E}[Z_{\text{sys}}]^{-1}\{F^e\}, \tag{34}$$

and

$$\frac{\partial\wp_{\text{rad}}}{\partial\Delta_E}=-\frac{\rho_s h}{\rho a}\frac{(ka)^3}{2}\times\text{Re}\left[i\{F^e\}^T[Z_{\text{sys}}]^{-1}\frac{\partial[Z_{\text{st}}]}{\partial\Delta_E}[Z_{\text{sys}}]^{-1}\{F^e\}^*\right], \tag{35}$$

respectively. In the following section, the impact of coupling scales and substructure modeling on the system response is investigated for two example system configurations.

VI. NUMERICAL EXAMPLES

The properties of the subplate and springs as well as the configurations for the system that are considered are given in Table I. Note that L_p/a is equal to $\xi_2-\xi_1$ and $k_{\text{rat1}}=k_{\text{rat2}}$. The properties for Case 1 correspond to a subplate with a density equal to that of the main plate and a mass equal to one-quarter that of the main plate. The properties for Case 2 roughly correspond to a plate with a density of aluminum and a mass equal to one-quarter that of the main plate. The two rigid-body resonances for the subsystem with $M_p=2$ may be calculated analytically. These values are given in the table where $(ka)_R$ is for the rotational mode and $(ka)_T$ is for the translational mode. Although numerous metrics could be used to characterize the system response, the radiated acoustic power is used here.

In the examples presented below, twenty structural modes [$M=20$ in Eq. (1)] and one-hundred fluid modes [$N=100$ in Eq. (2)] are used for the range of excitation frequencies considered. For a discussion on numerical convergence, see Ref. 9. Unless otherwise stated, the location of the line-force is $\xi_e=-0.25$.

The radiated acoustic power for Case 1 with $M_p=2$ is shown in Fig. 4 for a few coupling scales. It can be seen that at $ka=1.0$, near the first subsystem resonance, some sort of vibration absorption is occurring. Also, near $ka=1.4$ a combined system fluid-loaded resonance occurs. It should be noted that the issues related to finding exact eigenvalues for

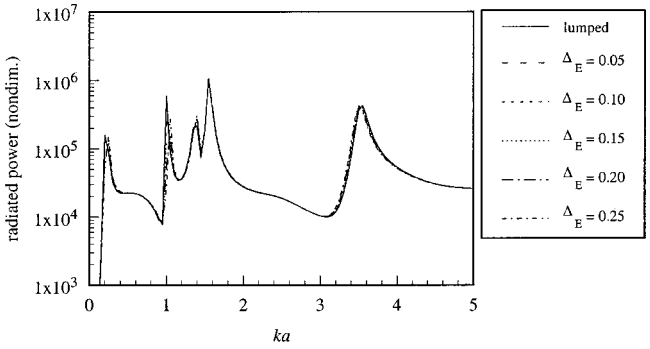


FIG. 4. Radiated acoustic power for Case 1 with $M_p=2$.

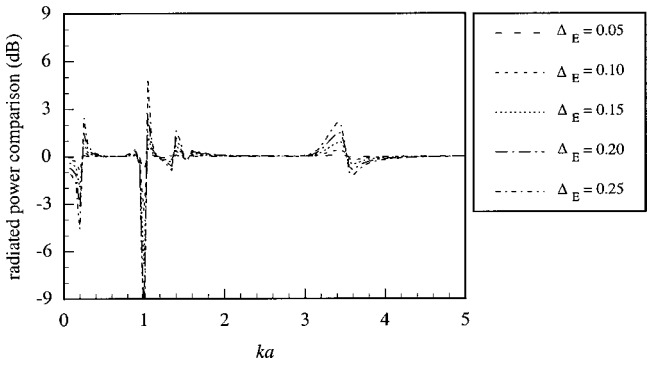


FIG. 5. A comparison of acoustic power for Case 1 with $M_p = 2$.

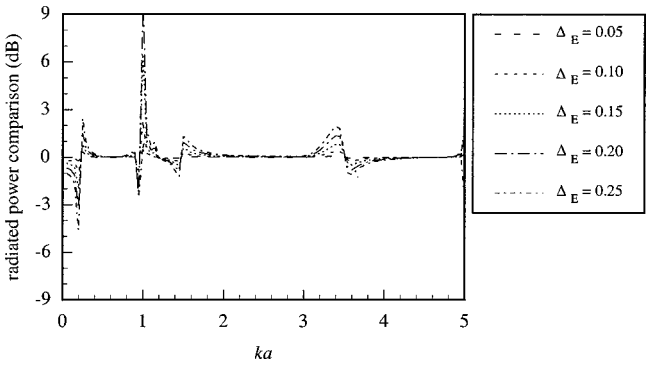


FIG. 7. A comparison of acoustic power for Case 1 with $M_p = 12$.

this system[12] have not been addressed here. As a result, the fluid-loaded resonances noted in this work should be taken as approximate. To bring out the effects of feature scales, a comparison of the radiated power is shown in Fig. 5. The power comparison is the increase in radiated power due to replacing the lumped spring model with a spatially represented spring of scale Δ_E. The impact at frequencies in the range $5 < ka \leqslant 10$ is not shown because it is negligible for the considered frequency spacing of 0.05. The impact of scales is very noticeable near the first, second, and fifth fluid-loaded resonances ($ka = 0.2$, 1.0, and 3.2). It is also noticeable near $ka = 1.4$, which is close to the second subsystem resonance.

The radiated power is shown for the same coupling scales, except with $M_p = 12$, in Fig. 6. The radiated power has changed since structural vibrations are now allowed to occur in the subplate. Nevertheless, the impact of scales on the radiated power is not much different than for the case with $M_p = 2$ as can be seen in a comparison of the radiated power in Fig. 7. The most significant impact occurs near the combined system resonances. The impact near $ka = 5$ has a maximum value of about 3 dB. Above $ka = 5$, though, the impact of coupling scales is negligible at frequencies which are integral multiples of 0.05.

The radiated acoustic power and the power comparison for Case 2 are shown in Fig. 8 and Fig. 9, respectively. These figures correspond to the analysis where only translational and rotational motion of the plate are allowed, $M_p = 2$. First note that a vibration absorbing condition is encountered near $ka = 2.6$. This is due mainly to the rotational motion of the subplate. When only translation is allowed, $M_p = 1$, the large decrease in power at $ka = 2.6$ is not observed. Although the impact of scales is certainly large around the fluid-loaded resonances, it is apparent that the impact is very noticeable near this absorption frequency. Unlike the behavior at the resonances, the frequency range of impact near the absorption frequency is fairly broad. For example, with two springs, each at a scale of $\Delta_E = 0.10$, the error in the radiated power is 3 dB or more for $2.7 < ka < 3.1$. Even though the impact near $ka = 7$ does not appear to be significant, it cannot be stated that this frequency is high enough to not be impacted by coupling scales. In fact, if the location of the excitation is moved away from the node in the fourth assumed mode for the main plate, the radiated power near $ka = 7$ increases significantly. Nevertheless, the maximum change in power there is less than that obtained for the lower resonance frequencies. If the allowed number of subplate modes is increased to $M_p = 12$, the radiated power in the frequency range $3.6 < ka < 7.5$ is only slightly different from that shown in Fig. 9 for $M_p = 2$. At the other frequencies, the results are basically identical as will be discussed below. As a result, the impact of scales on the power for $M_p = 12$ is the same as for the case where $M_p = 2$. Therefore, the power comparison is not shown for $M_p = 12$.

The impact of coupling scales due to the subplate modeling details is imbedded in the number of subplate assumed modes that are allowed. The more complex model has two rigid body modes and ten vibrational modes, $M_p = 12$, while the simpler model considered has only two rigid-body modes, $M_p = 2$. The difference in the radiated power for these two models is shown in Fig. 10. Even though coupling scales are very important for frequencies below $ka = 2.5$, the details of the subplate model are not important. That is, the simpler subplate model is adequate to obtain an accurate pre-

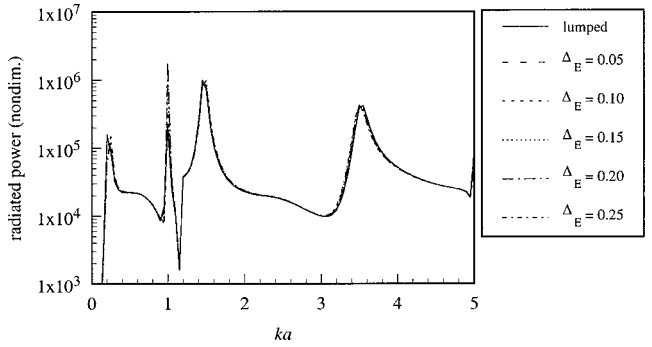


FIG. 6. Radiated acoustic power for Case 1 with $M_p = 12$.

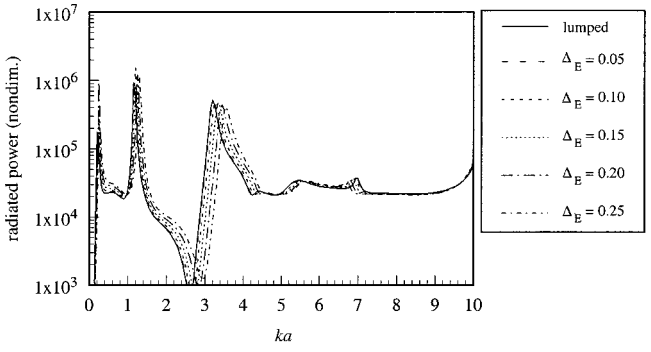


FIG. 8. Radiated acoustic power for Case 2 with $M_p = 2$.

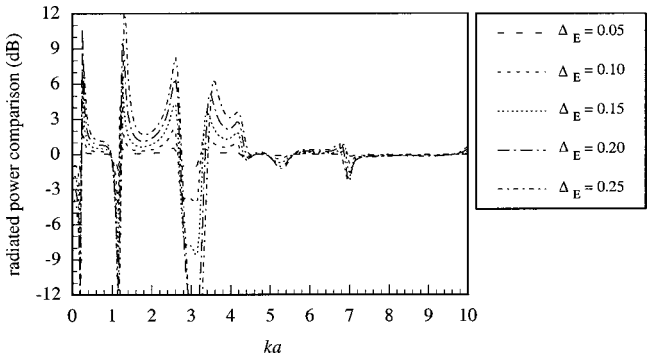


FIG. 9. A comparison of acoustic power for Case 2 with $M_p=2$.

diction of the radiated power in this range. For frequencies in the range $2.5<ka<4.2$, the effect of the subplate detail is much larger than for the previous range. Nevertheless, compared to the overall impact of coupling scales in this range, the detail of the subplate is still not very important. In the range $4.2<ka<5.5$, the subplate details play a more important role in the response of the system. However, the overall impact of coupling scales is not significant in this range. Therefore, for the entire frequency range considered, the impact of coupling scales is not significantly influenced when the simpler subplate model is implemented and structural vibrations in the subplate are neglected (i.e., $M_p=2$).

Consider the results shown in Fig. 11 for Case 1 at $ka=3.4$. The change in radiated acoustic power and normalized acoustic power sensitivity (N.A.P.S.) are shown for a range of subplate locations. The normalized acoustic power sensitivity is found using the relationship

$$\text{N.A.P.S.}=\frac{(\partial \wp_{\text{rad}}/\partial \Delta_E)}{\wp_{\text{rad}}}. \tag{36}$$

The width of the subplate is always equal to that specified in Table I. However, the attachment locations of the springs are varied and the value for the leftmost spring, ξ_1, is shown in the figures. It is trivial to locate ξ_2 since L_p/a is known.

A very low sensitivity to changes in coupling scales is observed when one of the springs is located between extrema in the magnitude of the plate surface response (not shown). Conversely, when one spring is located near a maximum or minimum in the surface response of the plate, the system response is very sensitive to changes in feature scales.

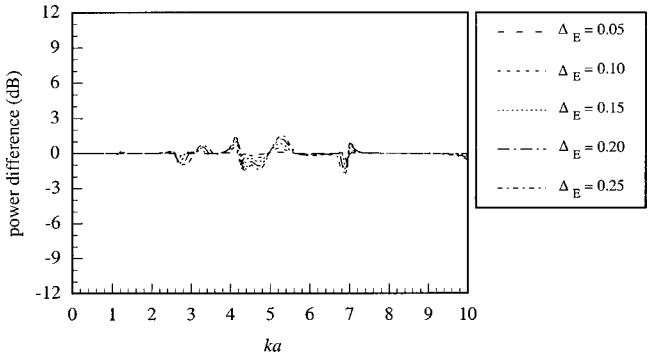


FIG. 10. The power difference between the twelve mode solution and the two mode solution, Case 2.

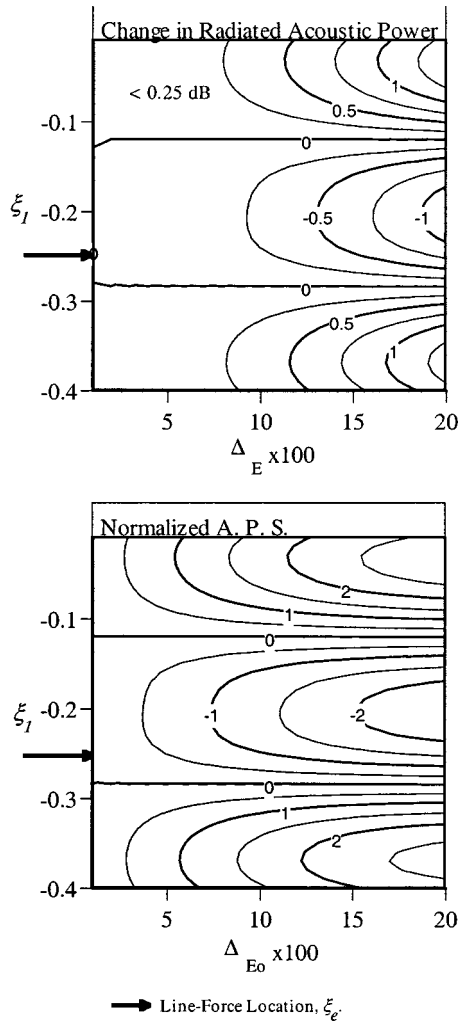


FIG. 11. Acoustic power comparison and normalized acoustic power sensitivity for a range of spring locations, ξ_1, and coupling scales, Δ_E, with $ka=3.40$, Case 1, and $M_p=12$.

The same form of results is provided in Fig. 12 for Case 2 at $ka=3.0$. Since at this frequency the springs are relatively stiff, the sensitivity to changes in scale is quite dramatic. Furthermore, the changes in the radiated power are noticeable. When the stiffness is relatively high with regard to the excitation frequency, the impact of scales does not occur in a uniform manner. Furthermore, there is no single location where the system is insensitive to changes in coupling scales.

VII. CONCLUDING REMARKS

To examine the impact of coupling scales and substructure modeling details on the response of the semi-infinite fluid-loaded plate, a wave-bearing subsystem was suspended by distributed springs from a fluid-loaded plate. By varying the size of the spring distributions and comparing the radiated acoustic power to that found using a lumped representation, the impact of coupling scales was investigated. Sensitivity relationships that relate changes in the feature scale to changes in the system response were also developed. These relationships were previously developed for a simpler system containing a single degree-of-freedom substructure. Like the results presented in that work, it was shown that the

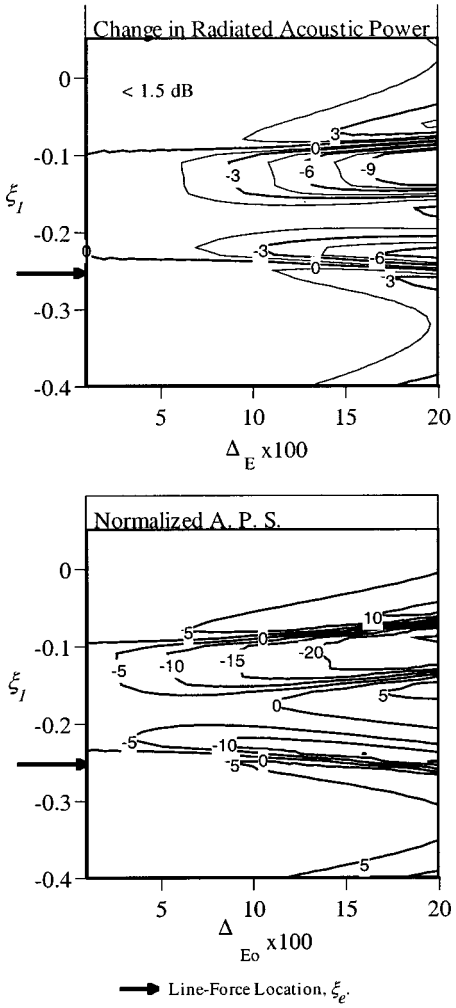


FIG. 12. Acoustic power comparison and normalized acoustic power sensitivity for a range of spring locations, ξ_1, and coupling scales, Δ_E, with $ka=3.0$, Case 2, and $M_p=12$.

method applied equally well to the more complicated system examined here. The increased complexity of the relationships is not significant given the information provided by the expressions. Two example configurations were examined to demonstrate the impact of coupling scales.

In the course of the study, it was possible to investigate effects associated with the modeling of the internal substructure waves. This was a natural step given the nature of substructure modeling in relation to other modeling issues. For the cases considered, it was shown that when the subsystem has a maximum subsystem resonance that is low relative to the excitation frequency, the impact of coupling scales has the characteristics of being quite small when an attachment location is between extrema in the surface response of the plate. As long as the two rigid-body modes of the subplate were included, allowing vibrations to occur within the subplate did not significantly alter the impact of coupling scales. Therefore, the modeling details of the subplate were not very important in the cases considered. Consequently, when the feature is located near a maxima or minima in the surface response, such as what would occur when the spring is located near a structural discontinuity, the system may be very sensitive to changes in the feature scale. Of course, many real structures contain discontinuities in the vicinity of structural stiffeners or bulkheads. A more thorough consideration of the subsystem parameters would definitely provide a more complete account of the impact of scales. Nevertheless, the results given here show that (1) the largest impact of scales is around the combined system resonances, and (2) for frequencies near and below the maximum subsystem resonance, the impact of scales can be significant between the resonances.

ACKNOWLEDGMENTS

The author would like to acknowledge the Georgia Institute of Technology and The University of Alabama for supporting this work.

[1] E. H. Dowell, ''On some general properties of combined dynamical systems,'' J. Appl. Mech. **46**, 206–209 (1979).
[2] Y. C. Das and D. R. Navaratna, ''Vibrations of a rectangular plate with concentrated mass, spring, and dashpot,'' J. Appl. Mech. **30**, 31–36 (1963).
[3] A. É. Ekimov and A. V. Lebedev, ''Vibrations of complex mechanical systems with lumped inhomogeneities,'' Sov. Phys. Acoust. **34**, 487–489 (1988).
[4] D. Feit and J. M. Cuschieri, ''Scattering of sound by a fluid-loaded plate with a distributed mass inhomogeneity,'' J. Acoust. Soc. Am. **99**, 2686–2700 (1996).
[5] B. A. T. Petersson, ''Efficiency of annularly distributed moment and force excitation regarding structural acoustic power transmission to plate-like structures,'' J. Sound Vib. **176**, 625–639 (1994).
[6] R. L. St. Pierre, Jr. and G. H. Koopmann, ''Minimization of radiated sound power from plates using distributed point masses,'' in *Proceedings of the Winter Annual Meeting*, ASME, New Orleans, LA, 1993, 93-WA/NCA-11, pp. 1–9.
[7] J. H. Ginsberg, K. A. Cunefare, and H. Pham, ''Spectral description of inertial effects in fluid-loaded plates,'' J. Vibr. Acoust. **117**, 206–212 (1995).
[8] W. S. Shepard, Jr. and K. A. Cunefare, ''Sensitivity of structural acoustic response to attachment feature scales,'' J. Acoust. Soc. Am. **102**, 1612–1619 (1997).
[9] W. S. Shepard, Jr., K. A. Cunefare, and J. H. Ginsberg, ''Identifying critical elastic scales in structural-acoustic models,'' J. Vibr. Acoust. **120**, 455–460 (1998).
[10] J. H. Ginsberg, A. D. Pierce, and X. F. Wu, ''A variational principle for sound radiation from vibrating bodies,'' Georgia Institute of Technology, 1986, No. GTADRL-86-101.
[11] A. D. Pierce, ''Stationary variational expressions for radiated and scattered acoustic power and related quantities,'' IEEE J. Ocean Eng. **OE-12**, 404–411 (1986).
[12] K. A. Cunefare and S. De Rosa, ''An improved state-space method for coupled fluid–structure interaction analysis,'' J. Acoust. Soc. Am. **105**, 206–210 (1999).
[13] J. H. Ginsberg and P. Chu, ''Asymmetric vibration of a heavily fluid-loaded circular plate using variational principles,'' J. Acoust. Soc. Am. **91**, 894–906 (1992).
[14] W. S. Shepard, Jr., ''The impact of attached feature scales and spatial distributions on the response of structural-acoustic systems,'' Ph.D. thesis, Georgia Institute of Technology, 1996.
[15] S. Alper and E. B. Magrab, ''Radiation from the forced harmonic vibrations of a clamped circular plate in an acoustic fluid,'' J. Acoust. Soc. Am. **48**, 681–691 (1970).
[16] F. Fahy, *Sound and Structural Vibration* (Academic, San Diego, 1985), 309 pp.
[17] R. E. D. Bishop and D. C. Johnson, *Vibration Analysis Tables* (Cambridge University Press, Cambridge, 1956).
[18] K. A. Cunefare, W. S. Shepard, Jr., and J. H. Ginsberg, ''Spectral techniques for assessing critical scales in models of fluid loaded structures,'' in *Proceedings of the 15th International Congress on Acoustics*, Trondheim, Norway, 26–30 June 1995, 4 of 4, edited by Mike Newman, pp. 405–408.
[19] R. A. Frazer, W. J. Duncan, and A. R. Collar, *Elementary Matrices and Some Applications to Dynamics and Differential Equations* (Cambridge University Press, 1938), 416 pp.

Noise transmission from a curved panel into a cylindrical enclosure: Analysis of structural acoustic coupling

James K. Henry and Robert L. Clark
Department of Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina 27708-0302

(Received 29 September 1999; accepted for publication 29 September 2000)

Much of the research on sound transmission through the aircraft fuselage into the interior of aircraft has considered coupling of the entire cylinder to the acoustic modes of the enclosure. Yet, much of the work on structural acoustic control of sound radiation has focused on reducing sound radiation from individual panels into an acoustic space. Research by the authors seeks to bridge this gap by considering the transmission of sound from individual panels on the fuselage to the interior of the aircraft. As part of this research, an analytical model of a curved panel, with attached piezoelectric actuators, subjected to a static pressure load was previously developed. In the present work, the analytical model is extended to consider the coupling of a curved panel to the interior acoustics of a rigid-walled cylinder. Insight gained from an accurate analytical model of the dynamics of the noise transmission from the curved panels of the fuselage into the cylindrical enclosure of an aircraft is essential to the development of feedback control systems for the control of stochastic inputs, such as turbulent boundary layer excitation. The criteria for maximal structural acoustic coupling between the modes of the curved panel and the modes of the cylindrical enclosure are studied. For panels with aspect ratios typical of those found in aircraft, results indicate that predominately axial structural modes couple most efficiently to the acoustic modes of the enclosure. The effects of the position of the curved panel on the cylinder are also studied. Structural acoustic coupling is found to not be significantly affected by varying panel position. The impact of the findings of this study on structural acoustic control design is discussed.
[DOI: 10.1121/1.1331114]

PACS numbers: 43.40.Vn [PJR]

I. INTRODUCTION

Previous research has shown that it is possible to use structural acoustic control with actuators mounted on the fuselage of the aircraft to reduce noise transmission into the interior of an aircraft. The research originated from studies conducted on the coupling between the vibration of the structure and the acoustics of an enclosure. Investigations of structural acoustic coupling relevant to the noise transmission problem were first done on plates coupled to the interior acoustics of a rectangular box.[1–8] The principles of modal coupling detailed in these works were later applied to the analysis of structural acoustic coupling between a cylinder and its interior.[9,10] As the field of active control developed, structural acoustic control using feedforward techniques aimed at the reduction of noise transmission into the interior of cylinders was heavily researched.[11–18] Much of this research focused on propellor induced low-frequency excitation of the entire cylinder. Researchers concluded that the structural modes that radiated efficiently into the enclosure were most effectively controlled using a structural acoustic performance metric in the cost function. It was also widely noted that care needed to be taken to not increase the overall vibration of the cylinder. Strong control results were demonstrated for tonal disturbances.

Recently, attention has turned to the reduction of noise transmission due to turbulent boundary layer excitation of the fuselage, presenting new design challenges. Unfortunately, the feedforward techniques used for the propellor noise problem require a reference signal correlated with the disturbance source. Since turbulent boundary layer excitation is stochastic, a reference signal is not readily available and direct, feedforward control of the panel is less practical. As a result, researchers have begun to use feedback techniques for structural acoustic control of sound transmission into enclosures. This research seeks to adapt many of the control techniques developed to reduce the radiation of flat plates into an acoustic half-space[19–26] to the problem of noise transmission into the fuselage of an aircraft. Most of the research conducted to date has examined the relatively simple case of a flat panel coupled to a rectangular enclosure.[27–31]

In response, recent research by the authors has sought to capture the application physics of fuselage panels in service conditions. Efforts in this direction have resulted in the development of an analytical model of a curved panel with attached piezoelectric transducers subjected to a static pressure load.[32,33] Continuing on this path, this paper focuses on the coupling of the curved panel model to the interior acoustics of a rigid-walled cylindrical shell utilizing a modal interaction model. The structural acoustic coupling is analyzed based on the spatial correspondence between structural and acoustic modes, as well as their frequency proximity. Results show that, for panels with aspect ratios typical of that required in service, predominately axial structural modes couple most efficiently to the acoustic modes of the enclosure. The effects of panel position relative to the cylinder on structural acoustic coupling is also investigated. Results

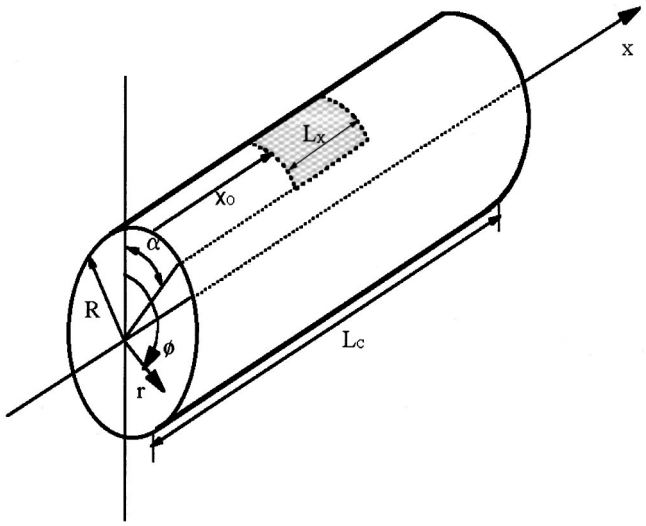


FIG. 1. Coordinate system of rigid-walled cylinder with endcaps (not shown) and curved panel mounted just inside boundary.

show that panel position does not significantly affect the coupling of the curved panel to the acoustics of the cylindrical enclosure. The impact of this work on structural acoustic control design is discussed.

II. MODEL DEVELOPMENT

The model development follows work by Fahy on the coupling of structures and fluids.[34] First, the dynamics of the acoustic enclosure are derived. Then, the equations of motion of the curved panel with attached piezoelectric transducers are discussed. Finally, the interaction of the two models is introduced and simulation parameters are defined.

A. Acoustic enclosure model development

Consider a rigid-walled cylindrical enclosure with endcaps and with coordinate system as shown in Fig. 1. The inhomogeneous wave equation describes the pressure within the fluid field:

$$\nabla^2 p - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = -\rho_0 \frac{\partial q}{\partial t}, \tag{1}$$

where p is the pressure, c is the speed of sound, ρ_0 is the density of the fluid, and q is the distribution of source volume velocity per unit volume. Now, consider a vibrating bounding surface, representative of the curved panel, covering a surface S_0 just inside the rigid boundary of the cylinder. Expressing the vibrating surface as a distributed volume velocity source, the wave equation can be rewritten in terms of the out-of-plane displacement of the curved panel, w, and the pressure within the fluid field:

$$\nabla^2 p(r,\phi,x,t) - \frac{1}{c^2}\frac{\partial^2 p(r,\phi,x,t)}{\partial t^2} = 2\rho_0 \frac{\partial^2 w}{\partial t^2}\,\delta(\xi-\xi_0), \tag{2}$$

where the spatial dirac delta function, $\delta(\xi-\xi_0)$, defines the position of the vibrating panel just inside the rigid-walled cylinder. The pressure is then expanded into a series sum of generalized coordinates and acoustic modes of the rigid-walled cylindrical enclosure with length L_c and radius R:

$$p(r,\phi,x,t) = \sum_{n_p=0}^{N_p} \Gamma_{n_p}(r,\phi,x)\,\vartheta_{n_p}(t), \tag{3}$$

where the acoustic mode shape functions are:[9]

$$\Gamma_{n_p}(r,\phi,x) = \sin\left[\beta\phi + \gamma\left(\frac{\pi}{2}\right)\right] J_\beta(\lambda_{\beta_\tau} r)\cos\left(\frac{\mu\pi x}{L_c}\right). \tag{4}$$

The index n_p corresponds to four different modal indices: β, γ, τ, and μ. The factor λ_{β_τ} represents the τth root of the derivative of the Bessel function:

$$J'_\beta(\lambda_{\beta_\tau} R) = 0. \tag{5}$$

Substituting Eq. (3) into Eq. (2), applying the identity from the homogeneous wave equation, multiplying by an arbitrary mode shape function Γ_m, integrating over the volume of the enclosure and applying the orthogonality condition yields the equations of motion of the enclosure:[34]

$$\Lambda_{n_p}\ddot{\vartheta}_{n_p}(t) + \Lambda_{n_p}\omega_{n_p}^2\vartheta_{n_p}(t) = -\rho_0 c^2 S_0 \sum_{k=1}^{K} C_{n_p k}\ddot{q}_k(t), \tag{6}$$

where the modal volume, Λ_{n_p}, is

$$\Lambda_{n_p} = \int_{V_c} \Gamma_{n_p}^T \Gamma_{n_p}\, dV, \tag{7}$$

and the coupling coefficient, $C_{n_p k}$, is

$$C_{n_p k} = \frac{1}{S_0}\int_{S_0} \Psi_k(x,\phi)\Gamma_{n_p}(R,\phi,x)\,dS. \tag{8}$$

To facilitate the development of the fully coupled structural/acoustic model, the acoustic pressure can be expressed in terms of the velocity potential, Φ:

$$p(r,\phi,x,t) = -\rho_c \frac{\partial \Phi(r,\phi,x,t)}{\partial t}. \tag{9}$$

The velocity potential can also be expanded into a series sum of generalized coordinates and acoustic modes of the rigid-walled cylindrical enclosure such that

$$\Phi(r,\phi,x,t) = \sum_{n_p=1}^{M} \Gamma_{n_p}(r,\phi,x)\,\phi_{n_p}^c(t), \tag{10}$$

where $\phi_{n_p}^c(t)$ is the generalized coordinate associated with the n_p^{th} acoustic mode, Γ_{n_p}. The equations of motion for the acoustic enclosure can then be expressed in terms of the cavity velocity potential as

$$\Lambda_{n_p}\ddot{\phi}_{n_p}^c(t) + \Lambda_{n_p}\omega_{n_p}^2\phi_{n_p}^c(t) = c^2 S_0 \sum_{n_p=1}^{N_p} C_{n_p k}\dot{q}_k(t). \tag{11}$$

The equation of motion of the enclosure can be expressed in matrix form:

$$\{\ddot{\phi}^c\} + [\mathbf{\Omega^2}]\{\phi^c\} = c^2 S_0 [\mathbf{\Lambda}]^{-1}[\mathbf{C}]\{\dot{\mathbf{q}}\}, \tag{12}$$

where $[\Omega^2]$ is a $N_p \times N_p$ diagonal matrix of the natural frequencies of the cavity squared, $\omega_{n_p}^2$, given by

$$\omega_{n_p}^2 = c^2\left[\lambda_{\beta_\tau}^2 + \left(\frac{\mu\pi}{L_c}\right)^2\right]. \tag{13}$$

$[\mathbf{\Lambda}]$ is an $N_p \times N_p$ diagonal matrix of the modal volumes defined by Eq. (7) and $[\mathbf{C}]$ is the $N_p \times K$ matrix of structural acoustic coupling coefficients as defined in Eq. (8). State–space formulation is easily accomplished.

B. Structural model development

The curved panel model development was presented in Refs. 32, 33 and will not be repeated here. The equations of motion of the curved panel now include an acoustic pressure loading term, $\mathbf{P^a}$,

$$[\mathbf{M_a}+\mathbf{M_p}]\{\ddot{\mathbf{q}}\}+[\mathbf{K_s}+\mathbf{K_p}+\mathbf{K_{tf}}]\{\mathbf{q}\} = [\mathbf{\Theta}]^T\{\mathbf{v}\}+[\mathbf{P^a}]+[\mathbf{B_f}+\mathbf{P_f}]\{\mathbf{f}\}. \tag{14}$$

The acoustic loading is expressed in terms of the derivative of the velocity potential generalized coordinate, $\dot{\phi}_{n_p}^c(t)$, such that

$$\mathbf{P^a} = -\rho_0 S_0[\mathbf{C}^T]\{\dot{\phi}_{n_p}^c(t)\}. \tag{15}$$

Simple mode shape functions satisfying shear diaphragm boundary conditions are used in the development:

$$\mathbf{\Psi}_k^d(x,\phi) = \begin{bmatrix} \sin\dfrac{n\pi\phi}{\alpha}\cos\dfrac{p\pi\hat{x}}{L_x} & 0 & 0 \\ 0 & \cos\dfrac{n\pi\phi}{\alpha}\sin\dfrac{p\pi\hat{x}}{L_x} & 0 \\ 0 & 0 & \sin\dfrac{n\pi\phi}{\alpha}\sin\dfrac{p\pi\hat{x}}{L_x} \end{bmatrix}, \tag{16}$$

where $\hat{x} = x - x_0$. The structural model can be cast in state–space form with ease.

C. Coupled structural acoustic model

The coupled structural acoustic model is formed by merging the models of the curved panel and the acoustic enclosure. Since both models are expressed in state–space form and are implemented in Matlab, coupling the models is relatively simple. Figure 2 shows the feedback loop used to create the coupled system. The curved panel model has three inputs: transverse loading from the piezoelectric actuator, a generalized force input from the static pressure load, and a coupled input from the acoustic cavity. As Eq. (15) indicates, the plate input is coupled to the derivative of the velocity potential. As Fig. 2 also shows, these states can be used to calculate an acoustic pressure output according to Eq. (9). The input to the cylindrical cavity is coupled to the structural velocity states, as described in Eq. (14).

The modal interaction model possesses certain limitations inherent to its development. By assuming that the panel is situated just inside the cylinder, the rigid-wall boundary conditions are no longer satisfied in the proximity of the panel. Therefore, the model will not accurately describe the acoustic pressure distribution near the surface of the panel. Still, the accuracy of the model away from the panel surface is sufficient for the purposes of this investigation.

D. Simulation parameters

The dimensions of the curved panel considered in this investigation have been selected to approximate fuselage subpanels found on a Boeing 737. The following parameters for the curved panel are held fixed throughout the investigation: $L_x = 0.508\,\text{m}$, $R = 1.75\,\text{m}$, $\alpha = 0.131\,\text{rad}$, $h_s = 1.8\,\text{mm}$, $E_s = 69.3\times10^9\,\text{N/m}^2$, $\rho_s = 2680\,\text{kg/m}^3$, $\nu_s = 0.33$. The panel is centered 30% down the length of the cylinder, such that $x_0 = 0.3L_c - (L_x/2)$. It is placed circumferentially starting at $\phi = 0$.

The dimensions of the rigid-walled cylinder have also been selected to approximate the size of the fuselage of a Boeing 737. The cylinder and the contained air have the following parameters: $L_c = 43.18\,\text{m}$, $R = 1.75\,\text{m}$, $\rho_0 = 1.21\,\text{kg/m}^3$, $c = 343\,\text{m/s}$. The cylinder is assumed to have internal pressure corresponding to atmospheric pressure at 10 000 ft, and to be in a medium with external pressure corresponding to atmospheric pressure at 40 000 ft. Thus the curved panel is subjected to a static pressure load resulting from the differential.

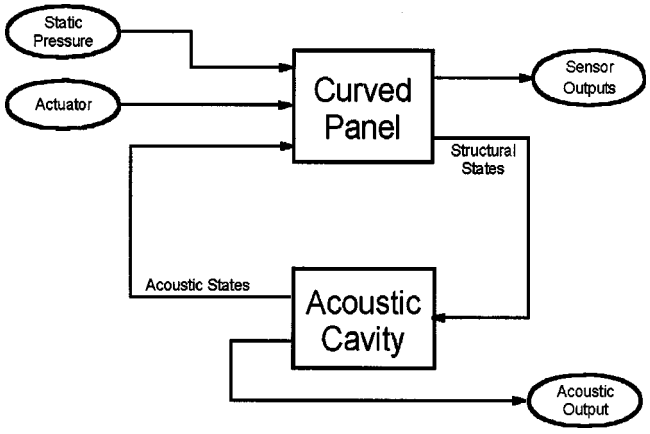


FIG. 2. Block diagram of coupled system.

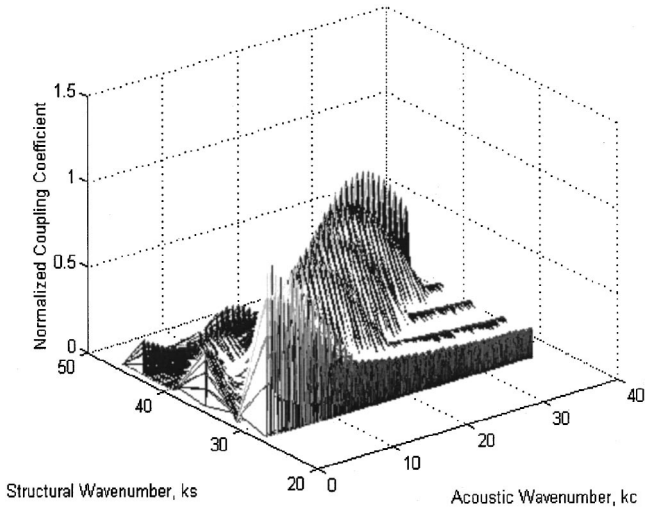


FIG. 3. Coupling coefficients for varying structural and acoustic wave numbers.

III. STRUCTURAL ACOUSTIC COUPLING

An understanding of the coupling between the structural modes and the acoustic modes is critical to the design of control systems for the reduction of the transmission of sound from the curved panel into the cylindrical enclosure. The coupling between the structural and acoustic modes is determined by two factors: spatial correspondence and frequency proximity. The spatial correspondence between structural modes and acoustic modes is defined by the structural acoustic coupling coefficients given by Eq. (8). Wave number analysis can be used to determine which acoustic modes possess a high degree of spatial correspondence to a given structural mode. Frequency proximity of acoustic modes to structural modes is determined by the natural frequencies of the modes in question. A structural mode will couple more efficiently to acoustic modes which have resonances proximal in frequency to the structural resonance. Maximal sound transmission occurs when a structural mode and an acoustic mode possess a high degree of spatial correspondence and have proximal natural frequencies.

Structural acoustic control for reduction of the transmission of sound from fuselage panels into the interior of aircraft is predominantly concerned with low frequency transmission, namely that of panel modes with resonances below 1000 Hz. Above 1000 Hz, passive treatments afford effective noise reduction. Therefore, this analysis will be confined to structural modes with natural frequencies below 1000 Hz.

A. Spatial correspondence

The structural acoustic coupling coefficients given by Eq. (8) provide a measure of the spatial correspondence between a particular structural and acoustic mode. By normalizing the coupling coefficients by the maximum coefficient in the bandwidth of consideration, the relative magnitudes of the coefficients can be considered. Further insight into the relative magnitudes of the structural acoustic coefficients can be gained by wave number analysis.

The structural wave number, k_s, for a tensioned curved panel with mode shapes given by Eq. (16) is[35]

$$k_s=\left[\left(\left(\frac{p\pi}{Lx}\right)^2+\left(\frac{n\pi}{R\alpha}\right)^2\right)^2+\frac{\sigma_{xx}h_s\left(\frac{p\pi}{Lx}\right)^2}{D_s}+\frac{\sigma_{\phi\phi}h_s\left(\frac{n\pi}{R\alpha}\right)^2}{D_s}+\frac{12\left(\frac{p\pi}{Lx}\right)^4(1-\nu^2)}{h_s^2R^2\left(\left(\frac{p\pi}{Lx}\right)^2+\left(\frac{n\pi}{R\alpha}\right)^2\right)^2}\right]^{1/4}, \tag{17}$$

where σ_{xx} and $\sigma_{\phi\phi}$ are tensile forces in the axial and circumferential directions, respectively, and $D_s=E_sh_s^3/12(1-\nu_s^2)$. The acoustic wave number, k_c, for the cylindrical enclosure with mode shapes given in Eq. (4) is[9]

$$k_c=\left[\lambda_{\beta_\tau}^2+\left(\frac{\mu\pi}{Lc}\right)^2\right]^{1/2}. \tag{18}$$

The structural and acoustic wave numbers can be calculated for a given set of modal indices.

The variation of the structural acoustic coupling coefficient, normalized by the maximum coefficient, with respect to the structural and acoustic wave numbers of the curved panel/cylindrical shell system is shown graphically in Fig. 3. The 3D plot indicates that the coupling coefficient for a particular structural wave number reaches a maximum at some value of acoustic wave number. The acoustic wave number at which the maximum is reached corresponds to the acoustic mode which demonstrates maximal spatial correspondence to the structural mode in question.

TABLE I. Maxima of the normalized coupling coefficients for structural modes.

Structural mode (p,n)	Structural wave number $(1/m)$	Max of normalized coup. coeff.	Acoustic mode (β,τ,μ)	Acoustic wave number $(1/m)$	Structural natural freq. (Hz)	Acoustic natural freq. (Hz)
(1,1)	27.7	1.0	(0,1,2)	1.5	338.3	79.4
(2,1)	31.5	0.818	(0,1,145)	10.5	437.0	575.9
(3,1)	35.7	0.799	(0,1,230)	16.7	561.2	913.5
(4,1)	39.7	0.790	(0,1,335)	24.4	695.1	1330.5
(1,2)	39.9	0.137	(20,1,0)	12.7	699.2	693.1
(2,2)	41.2	0.112	(20,1,145)	16.5	747.9	901.1
(3,2)	43.5	0.109	(20,1,230)	21.0	833.0	1146.7
(5,1)	43.9	0.789	(0,1,420)	30.6	847.6	1668.1
(4,2)	46.5	0.108	(20,1,335)	27.5	952.4	1500.2

Table I presents the maxima of the normalized coupling coefficients between each of the structural modes under 1000 Hz and all of the acoustic modes up to 2000 Hz. Examination of the maxima of the coupling coefficients indicates that structural modes which are predominately axial in orientation, namely the (1,1), (2,1), (3,1), (4,1), and (5,1) modes, have a much higher spatial correspondence to an acoustic mode than structural modes with circumferential components, namely the (1,2), (2,2), (3,2), and (4,2) modes. This phenomenon is partially explained by the fact that the panel is more than 2× as long in the axial direction as the circumferential direction, increasing the spatial aperture in the axial direction. Furthermore, the analysis is limited to cavity modes with natural frequencies less than 2000 Hz, making it entirely possible that the modes with circumferential components show a greater level of spatial correspondence to acoustic modes with higher natural frequencies and wave numbers. However, these higher order modes are well out of the desired bandwidth of control and their contribution to the acoustics at low frequency is limited. Regardless, predominately axial structural modes have spatial correspondence nearly 8× greater than structural modes with circumferential components, making it possible to virtually ignore circumferential modes as a source of sound transmission into the cavity.

Table I also presents the acoustic mode that demonstrates maximal spatial correspondence to each structural mode. Interestingly, the predominately axial modes maximally correspond to acoustic modes with axial indices, μ, which appear to be linearly related to the ratio of the length of the cavity to the length of panel, $L_c/L_x=85$. Furthermore, the structural modes with circumferential components maximally correspond to acoustic modes with these same axial indices. Analysis of the wave number spectra of the axial components of the expression for the coupling coefficients will be used to provide insight.

The coupling coefficient, C_{n_pk}, can be broken down into its spatial components:

$$C_{n_pk}=\frac{1}{S_0}\int_{S_0}\Psi_x(x)\Gamma_x(x)\Psi_\phi(\phi)\Gamma_\phi(\phi)\Gamma_R(r=R)dS. \quad (19)$$

The axial components can then be separated and the wave number transformation can be performed on the structural and acoustic components, respectively:

$$\Psi_x(k_{s_x})=\int_{x_0}^{x_0+L_x}\Psi_k(x)e^{-jk_{c_x}x}\,dx, \quad (20)$$

$$\Gamma_x(k_{c_x})=\int_{x_0}^{x_0+L_x}\Gamma_x(x)e^{-jk_{c_x}x}\,dx, \quad (21)$$

where $\Psi_x(x)=\sin(p\pi(x-x_0)/L_x)$ and $\Gamma_x(x)=\cos(\mu\pi x/L_c)$, from the expressions for the mode shape functions. Figure 4 shows the wave number spectrum of the axial structural component of the coupling coefficient for axial modal index $p=4$, and the wave number spectrum of the axial acoustic component of the coupling coefficient for various axial modal indices, μ. When $\mu=335$, the spectra nearly coincide, indicating strong spatial correspondence in the axial direction. When $\mu<335$ or $\mu>335$, as shown for the other two values of μ in Fig. 4, the overlap between the structural and acoustic wave number spectra is not as significant. When the spectra coincide, the axial component of the structural wave number at which the maximum magnitude of the structural spectrum occurs is equivalent to the axial component of the acoustic wave number at which the maximum magnitude of the acoustic spectrum occurs. Referring back to the expression for the acoustic wave number in Eq. (18), the expression for the axial component of the acoustic wave number is easily seen to be $k_{c_x}=\mu\pi/L_c$. Examining the expression for structural wave number in Eq. (17), a simple expression for the axial component of the structural wave number, k_{s_x}, is not readily apparent. Therefore, the relationship between the axial modal indices of the structural and acoustic modes with maximal spatial correspondence unfortunately cannot be stated by a simple analytical expression. Still, the wave number spectra can be used to predict the values of the axial component of the wave number at which maximal correspondence occurs.

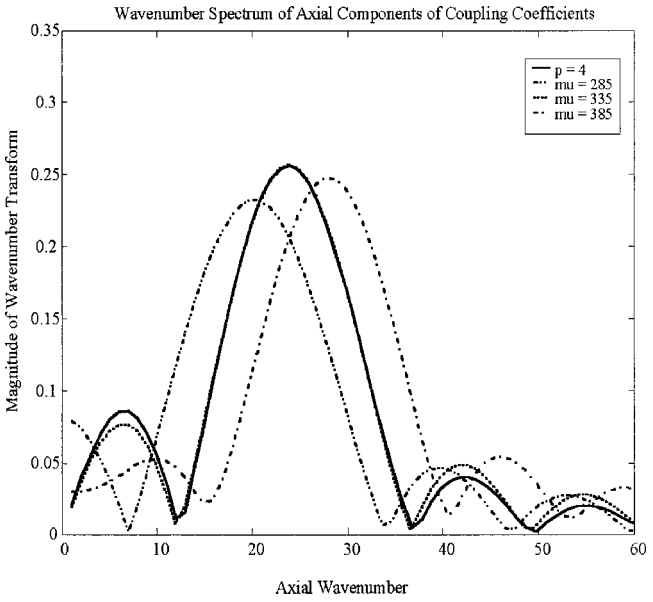


FIG. 4. Wave number spectra of axial components of coupling coefficient.

B. Frequency proximity

For maximal sound transmission, an acoustic and structural mode pair must have maximal spatial correspondence and must be proximal in natural frequency. As discussed in the previous section, Table I presents the acoustic modes which demonstrate maximal spatial correspondence to each structural mode in the bandwidth of consideration. However, as Table I also shows, the natural frequencies of the acoustic modes which have maximal spatial correspondence are not proximate to the natural frequencies of the structural modes. The criterion for frequency proximity was defined by Fahy[36] as

$$|\omega_c-\omega_s|<\xi_c\omega_c+\xi_s\omega_s, \quad (22)$$

where ξ_c and ξ_s are the damping ratios of the acoustic and structural systems, respectively. The values of the damping

TABLE II. Maxima of the normalized coupling coefficients for structural modes.

Structural mode (p,n)	Structural wave number $(1/m)$	Max of normalized coup. coeff.	Acoustic mode (β,τ,μ)	Acoustic wave number $(1/m)$	Structural natural freq. (Hz)	Acoustic natural freq. (Hz)
(1,1)	27.7	0.764	(0,1,81)	5.9	338.3	321.7
(2,1)	31.5	0.783	(0,1,115)	8.4	437.0	456.8
(3,1)	35.7	0.406	(0,1,140)	10.2	561.2	556.0
(4,1)	39.7	0.067	(0,1,183)	13.3	695.1	726.8
(1,2)	39.9	0.137	(20,1,0)	12.7	699.2	693.1
(2,2)	41.2	0.099	(19,1,105)	14.3	747.9	781.4
(3,2)	43.5	0.064	(17,1,160)	16.0	833.0	871.2
(5,1)	43.9	0.195	(0,1,210)	15.3	847.6	834.1
(4,2)	46.5	0.029	(14,1,215)	18.1	952.4	988.7

ratios assumed here are: $\xi_c = 0.05$, $\xi_s = 0.05$. Table II presents the maxima of the normalized coupling coefficients between each of the structural modes under 1000 Hz and acoustic modes which satisfy the frequency proximity criterion. With the exception of the (1,2) structural mode, the maximum coupling coefficients between the structural modes and the acoustic modes in the bandwidth proximal to the structural resonances, presented in Table II, are less than the maximum coupling coefficients between the structural modes considered and the acoustic modes in the entire bandwidth, presented in Table I. That is, the maximum coupling coefficient of acoustic modes meeting the frequency criterion in Eq. (22) is less than that associated with acoustic modes with maximal spatial correspondence to the structural mode in question. Still, the coupling coefficients associated with the low order axial modes, namely the (1,1), (2,1), and (3,1) modes, are quite high, especially in comparison to the other structural modes in the bandwidth. Clearly, the relative magnitudes of the coupling coefficients indicate which modes contribute significantly to sound radiation into the cylinder. The low order axial modes demonstrate the highest degree of spatial correspondence to acoustic modes satisfying the frequency proximity criterion. Thus the low order axial modes of the structure couple most efficiently to the modes of the enclosure. Other modes, such as the (1,2) and (5,1) structural modes, also contribute to the sound transmission into the enclosure. But structural modes whose maximum coupling coefficient in Table II is less than 10% of the normalizing factor, namely the (4,1), (2,2), (3,2), and (4,2) modes, are relatively insignificant, since the coefficients are one full order of magnitude less than the maximum coupling coefficient.

C. Panel position on cylinder

The previous sections examined the coupling of a typical curved panel to the cylindrical enclosure. To confirm the observations made about structural acoustic coupling from this typical example, other panel positions must be considered. Assume curved panels are positioned end to end inside the cylinder, starting at one end of the cylindrical enclosure and extending to the opposite end. Given the ratio of the length of the cylinder to the length of the curved panel, $Lc/Lx = 85$, there are 85 possible panel positions on the cylinder. The coupling between the curved panel and the enclosure at each of the 85 possible panel positions was examined. It was found that at each possible panel position, there were at least 4, and as many as 35, acoustic modes that demonstrated spatial correspondence within 5% of the maximum for each structural mode, as given in Table I. For example, Fig. 5 shows the number of acoustic modes demonstrating spatial correspondence within 5% of the maximum spatial correspondence to the (2,1) structural mode. As seen in the figure, at least nine acoustic modes demonstrate high spatial correspondence at each possible position of the curved panel on the cylinder. Figure 5 is typical of the other structural modes in the structural bandwidth up to 1000 Hz. No matter where the curved panel is placed on the cylinder, the structural modes of the panel will correspond spatially to many acoustic modes. The insignificance of panel position to the magnitude of the structural acoustic coupling coefficient is due in large part to the high acoustic modal density within the bandwidth of consideration and the relative length of the panel to the length of the enclosure.

Cross referencing the spatial correspondence with the frequency constraints presented in Eq. (22) allows investigation of structural acoustic coupling within the bandwidth of the structural resonances. Figure 6 shows the variation with position of the maximum coupling coefficient for acoustic

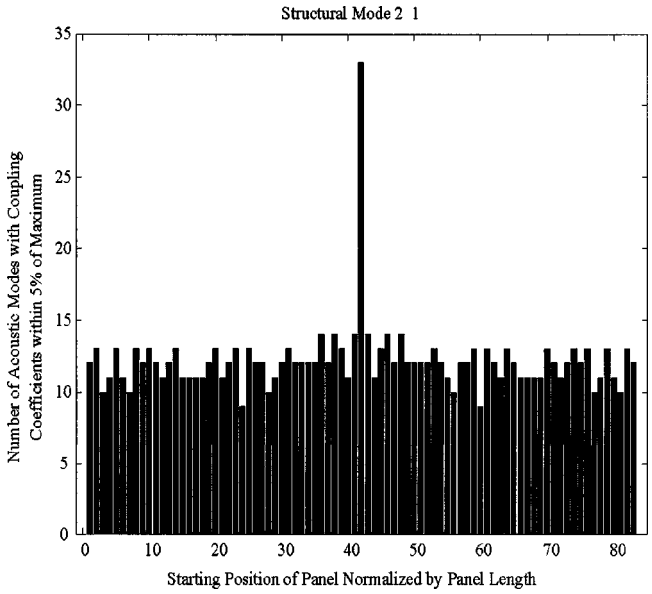


FIG. 5. Effects of position on the number of acoustic modes demonstrating maximal spatial correspondence to the 2, 1 structural mode.

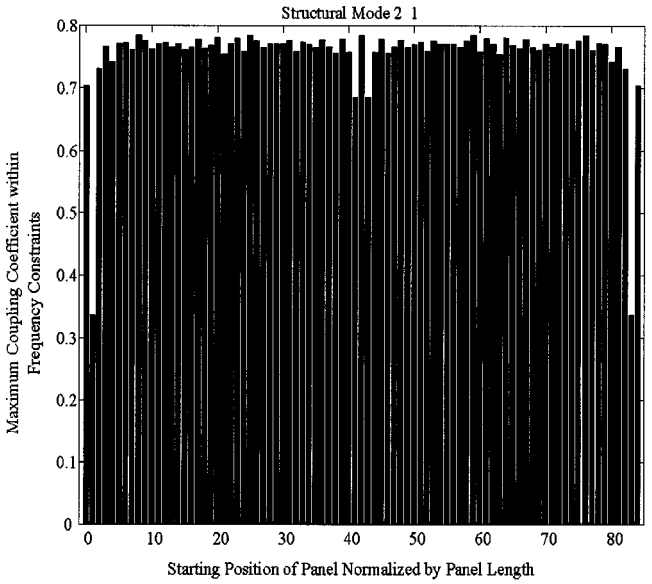


FIG. 6. Effects of position on the maximum coupling coefficient of acoustic modes proximal in natural frequency to the 2, 1 structural mode.

modes proximal in natural frequency to the (2,1) structural mode. As the figure shows, the maximum coupling coefficient at each position on the cylinder nears the maximum coefficient, $C=0.783$, from Table II, with the exception of one position at each of the ends where the coefficient drops off considerably. The acoustic mode to which the (2,1) structural mode is maximally coupled varies with position, but the magnitude of the coefficient remains high. Figure 6 is typical of the other structural modes in the bandwidth considered. With the exception of a few isolated positions, the placement of the panel on the cylinder does not affect the coupling between the structural modes of the panel and the acoustic modes of the enclosure proximal in natural frequency to the structural resonances.

D. Impact on structural acoustic control

Analysis of structural acoustic coupling, for a typical aircraft panel, has shown that predominately axial structural modes couple most efficiently to the acoustic modes of the cylindrical enclosure. As a result, structural acoustic control systems should focus primarily on reducing the radiation of the predominately axial modes of the curved panel. Using actuator placement optimization, piezoelectric transducers can be positioned on the curved panel to couple efficiently to the axial structural modes. Circumferential structural modes can largely be ignored. Furthermore, transducers designed for spatial compensation and dynamic compensators can be employed to focus control energy on the axial structural modes.

The investigation of the variation of structural acoustic coupling with panel position has shown that position on the cylinder does not significantly affect the magnitude of the structural acoustic coupling coefficients. Therefore, the structural acoustic control applied to each panel can be directed at the same structural modes.

IV. CONCLUSIONS

An analytical model of a curved panel coupled to the interior acoustics of a rigid-walled cylindrical enclosure was developed. The structural acoustic coupling between structural and acoustic modes was investigated with respect to spatial correspondence and frequency proximity. It was found that, for a typical aircraft panel, predominately axial structural modes coupled most efficiently to the acoustic modes of the enclosure. Therefore, structural acoustic control design should focus on low-order axial structural modes in order to reduce sound transmission into the enclosure. It was also found that panel position on the cylinder did not significantly affect structural acoustic coupling. Therefore, a similar control strategy can be applied to each panel on the cylinder.

ACKNOWLEDGMENTS

This work was supported by NASA Langley Research Center under Grant No. NCC 1 250 and the direction of Dr. Richard J. Silcox.

[1] R. H. Lyon, ''Noise reduction of rectangular enclosures with one flexible wall,'' J. Acoust. Soc. Am. **35**, 1791–1797 (1963).
[2] A. J. Pretlove, ''Forced vibrations of a rectangular panel backed by a closed rectangular cavity,'' J. Sound Vib. **3**, 252–261 (1966).
[3] W. R. Graham, ''Boundary layer induced noise in aircraft, Part I: The flat plate model,'' J. Sound Vib. **192**, 101–120 (1996).
[4] R. W. Guy and M. C. Bhattacharya, ''The transmission of sound through a cavity-backed finite plate,'' J. Sound Vib. **27**, 207–223 (1973).
[5] J. Pan, C. H. Hansen, and D. A. Bies, ''Active control of noise transmission through a panel into a cavity: I. Analytical study,'' J. Acoust. Soc. Am. **87**, 2098–2108 (1990).
[6] S. Narayanan and R. L. Shanbhag, ''Acoustoelasticity of a damped sandwich panel backed by a cavity,'' J. Sound Vib. **78**, 453–473 (1981).
[7] E. H. Dowell and H. M. Voss, ''The effect of a cavity on panel vibration,'' AIAA J. **1**, 476–477 (1962).
[8] E. H. Dowell, G. F. Gorman, and D. A. Smith, ''Acoustoelasticity: General theory, acoustic natural modes and forced sinusoidal excitation, including comparisons with experiment,'' J. Sound Vib. **52**, 519–542 (1977).
[9] Li Cheng and Jean Nicolas, ''Radiation of sound into a cylindrical enclosure from a point-driven end plate with general boundary conditions,'' J. Acoust. Soc. Am. **91**(3), 1504–1513 (1992).
[10] J. F. Unruh and S. A. Dobosz, ''Fuselage structural modeling for structure-borne interior noise transmission,'' ASME J. Vib., Acoust., Stress, Reliab. Des. **110**, 226–233 (1988).
[11] C. R. Fuller and J. D. Jones, ''Experiments on reduction of propellor induced interior noise by active control of cylinder vibration,'' J. Sound Vib. **112**, 389–395 (1987).
[12] R. J. Silcox, C. R. Fuller, and H. C. Lester, ''Mechanisms of active control in cylindrical fuselage structures,'' in AIAA Conference Paper, AIAA-87-2707 (1987).
[13] J. D. Jones and C. R. Fuller, ''Active control of sound fields in elastic cylinders by vibrational inputs,'' in Proceedings of the NoiseCON, pp. 413–418 (1987).
[14] S. D. Snyder and C. H. Hansen, ''The design of systems to control actively periodic sound transmission into enclosed spaces, Parts I and II,'' J. Sound Vib. **170**, 433–472 (1994).
[15] P. A. Nelson D. R. Thomas, and S. J. Elliot, ''Active control of the transmission of sound through a thin cylindrical shell, Part I: The minimization of vibrational energy,'' J. Sound Vib. **167**, 91–111 (1993).
[16] P. A. Nelson, D. R. Thomas, and S. J. Elliot, ''Active control of the transmission of sound through a thin cylindrical shell, Part II: The minimization of acoustic potential energy;'' J. Sound Vib. **167**, 113–128 (1993).
[17] J. Q. Sun, M. A. Norris, D. J. Rossetti, and J. H. Highfill, ''Distributed piezoelectric actuators for shell interior noise control,'' Trans. ASME, J. Vib. Acoust. **118**, 676–681 (1996).

[18] B. S. Cazzolato and C. H. Hansen, ''Active control of sound transmission using structural error sensing,'' J. Acoust. Soc. Am. **104**, 2878–2889 (1998).
[19] W. T. Baumann, W. R. Saunders, and H. H. Robertshaw, ''Active structural acoustic control of broadband disturbances,'' J. Acoust. Soc. Am. **92**, 1998–2005 (1992).
[20] Yi Gu, R. L. Clark, C. R. Fuller, and A. C. Zander, ''Experiments on active control of plate vibration using piezoelectric actuators and polyvinylidene fluoride modal sensors,'' Trans. ASME, J. Vib. Acoust. **116**, 303–308 (1994).
[21] R. L. Clark and C. R. Fuller, ''Optimal placement of piezoelectric actuators and polyvinylidene fluoride (PVDF) error sensors in active structural acoustic control approaches,'' J. Acoust. Soc. Am. **92**, 1521–1533 (1992).
[22] R. L. Clark and C. R. Fuller, ''Active structural acoustic control with adaptive structures and wave number considerations,'' J. Intell. Mater. Syst. Struct. **3**, 296–315 (1992).
[23] R. L. Clark and C. R. Fuller, ''Experiments on active control of structurally radiated sound using multiple piezoceramic actuators,'' J. Acoust. Soc. Am. **91**, 3313–3320 (1992).
[24] R. L. Clark and C. R. Fuller, ''Control of sound radiation with adaptive structures,'' J. Intell. Mater. Syst. Struct. **2**, 431–452 (1991).
[25] J. S. Vipperman and R. L. Clark, ''Multivariable feedback active structural acoustic control with adaptive piezoelectric sensoriactuators,'' J. Acoust. Soc. Am. **105**, 219–225 (1997).
[26] J. S. Vipperman, ''Adaptive piezoelectric sensoriactuators for active structural acoustic control,'' Ph.D. thesis, Duke University, 1996.
[27] S. Koshigoe, J. T. Gillis, and E. T. Falangas, ''A new approach for active control of sound transmission through an elastic plate backed by a rectangular cavity,'' J. Acoust. Soc. Am. **94**, 900–907 (1993).
[28] K. D. Frampton and R. L. Clark, ''Active control of noise transmission through an aeroelastic plate,'' in Proceedings of the 38th Structures, Structural Dynamics and Materials Conference, Kissimmee, FL, April 1997.
[29] K. D. Frampton and R. L. Clark, ''Power flow in an aeroelastic plate backed by a reverberant cavity,'' J. Acoust. Soc. Am. **102**, 1620 (1997).
[30] K. D. Frampton and R. L. Clark, ''Sound transmission through an aeroelastic plate into a cavity,'' AIAA J. **35**, 1113–1118 (1997).
[31] K. D. Frampton, ''Active control of noise transmission through an aeroelastic plate into an acoustic enclosure,'' Ph.D. thesis, Duke University, 1996.
[32] James K. Henry and Robert L. Clark, ''A curved piezo-structure model: Implications on active structural acoustic control,'' J. Acoust. Soc. Am. **106**, 1400–1407 (1999).
[33] J. K. Henry and Robert L. Clark, ''Smart aircraft panels: The effects of internal pressure loading on panel dynamics,'' in Proceedings of SPIE's 6th Annual International Symposium on Smart Structures and Materials, Newport Beach, CA, Paper 3668-09, March 1999.
[34] F. Fahy, *Sound and Structural Vibration* (Academic, London, 1985).
[35] E. Szechenyi, ''Approximate methods for the determination of the natural frequencies of stiffened and curved plates,'' J. Sound Vib. **14**, 401–418 (1971).
[36] F. J. Fahy, ''Vibration of containing structures by sound in the contained fluid,'' J. Sound Vib. **10**, 490–512 (1969).

Psychoacoustic correlates of individual noise sensitivity[a)]

Wolfgang Ellermeier,[b)] Monika Eigenstetter, and Karin Zimmer[c)]
Institut für Psychologie, Universität Regensburg, 93040 Regensburg, Germany

(Received 22 February 2000; accepted for publication 2 January 2001)

In environmental noise surveys, self-reported noise sensitivity, a stable personality trait covering attitudes toward a wide range of environmental sounds, is a major predictor of individual noise-annoyance reactions. Its relationship to basic measures of auditory functioning, however, has not been systematically explored. Therefore, in the present investigation, a sample of 61 unselected listeners was subjected to a battery of psychoacoustic procedures ranging from threshold determinations to loudness scaling tasks. No significant differences in absolute thresholds, intensity discrimination, simple auditory reaction time, or power-function exponents for loudness emerged, when the sample was split along the median into two groups of ''low'' vs ''high'' noise sensitivity on the basis of scores obtained from a psychometrically evaluated questionnaire [Zimmer and Ellermeier, Diagnostica **44**, 11–20 (1998)]. Small, but systematic differences were found in verbal loudness estimates, and in ratings of the unpleasantness of natural sounds, thus suggesting that self-reported noise sensitivity captures evaluative rather than sensory aspects of auditory processing.
 [DOI: 10.1121/1.1350402]

PACS numbers: 43.50.Qp, 43.66.Cb, 43.66.Fe [RVS]

I. INTRODUCTION

A. The concept of noise sensitivity

Individual noise sensitivity is a personality trait covering attitudes towards a wide range of environmental sounds, and is typically assessed by obtaining responses to one or several rating-scale items.

Noise sensitivity is a major antecedent of individual noise annoyance, as has been demonstrated in field surveys and laboratory experiments alike (see Taylor, 1984; Stansfeld *et al.*, 1985; Stansfeld, 1992; Job, 1988, 1999; Staples, 1996). In a review of 27 studies pertaining to railway, aircraft, and construction noise, Job (1988) reported annoyance reactions to be determined most strongly by noise exposure. With noise exposure controlled for, the second most powerful predictor was individual noise sensitivity, with correlations ranging from $r=0.25$ to $r=0.45$. On average, individual noise sensitivity explained 10.2% of variation in noise-annoyance reactions towards a given sound source, as compared to 17.6% of variation explained by noise-exposure measures. Taylor (1984) found an even stronger influence of noise sensitivity on annoyance reactions. Using a path-modeling approach to investigate the impact of noise-exposure measures, attitudes towards aircraft operation, and several personal-background characteristics on annoyance by aircraft noise, Taylor (1984) found noise sensitivity to have the largest single effect overall (as did Langdon, 1976).

It should be emphasized that, conceptually, *noise sensitivity* is clearly distinguishable from *noise annoyance*, as can be seen by their different correlation structures: Whereas measures of annoyance show a clear positive correlation with indices of noise exposure ($r=0.30$), noise-sensitivity measures are independent of exposure ($r=-0.02$; see Taylor, 1984; Job, 1988, 1999). Nevertheless, noise sensitivity has been demonstrated to have direct or indirect effects on health by (a) constituting a stressful psychological condition in its own right (Job, 1999); (b) increasing physiological reactivity of the cardiovascular system (Ising *et al.*, 1980; Stansfeld and Shine, 1993); and (c) being found to covary with the degree of psychopathology (Stansfeld, 1992).

The use of the noise-sensitivity concept is not restricted to clinical or subclinical populations, but refers to a property prevalent in the population at large. There is strong consensus that noise sensitivity constitutes a personality trait that is stable over time (Stansfeld, 1992; Langdon, 1976; Weinstein, 1978; Zimmer and Ellermeier, 1998a, 1999). A definition encompassing all facets of noise sensitivity discussed in the literature has been proposed only recently: Job (1999) defined noise sensitivity as referring to ''*...the internal states (be they physiological, psychological [including attitudinal], or related to life style or activities conducted) of any individual which increase their degree of reactivity to noise in general*'' (p. 59).

B. Evidence for a sensory component in noise sensitivity

While this definition distinguishes several levels at which noise sensitivity might operate, many investigators have conceptualized a crucial component of noise sensitivity to be *perceptual* in nature. Reason (1972) hypothesized a physiologically based disposition that might lead individuals to ''...transduce... input more effectively so that the subjective experience it evokes is more intense than that produced within less 'receptive' individuals by the same level of stimulus energy'' (p. 306). Taylor (1984) conceptualized

[a)]Portions of the data were presented at the joint meeting of the Acoustical Society of America (ASA) and the European Acoustics Association (EAA), Berlin, March 1999.

[b)]Electronic mail: wolfgang.ellermeier@psychologie.uni-regensburg.de
[c)]Currently at Institut für Kognitionsforschung, Universität Oldenburg, FB5/A6, P.O. Box 2503, 26111 Oldenburg, Germany. Electronic mail: karin.zimmer@uni-oldenburg.de

personal sensitivity as related to how individuals ''perceive noisy events'' (p. 259), as opposed to how they evaluate them, and Job (1999) explicitly includes greater ''hearing acuity'' (p. 59) as a possible contributor to noise sensitivity.

The evidence for a sensory component in noise sensitivity is weak, however. We found only four studies addressing the issue, and we will demonstrate that methodological and conceptual shortcomings prevent unequivocal conclusions from the present state of research.

Investigating physiological and psychophysical correlates of noise sensitivity in a female community sample ($n=72$), Stansfeld *et al.* (1985) found no relationship between noise sensitivity and absolute hearing threshold or uncomfortable loudness level. A small effect of noise sensitivity was found in graphic ratings of the loudness of 2-kHz tones (see their Fig. 1), indicating that highly noise-sensitive individuals exhibit steeper psychophysical functions. This finding, however, did not hold up when instead of a direct self-rating of noise sensitivity the McKennell (1963) scale capturing potential disturbance by seven everyday sounds, was employed. Furthermore, the authors acknowledge the procedure to be flawed, since reversing the polarity of the graphic rating scale was occasionally misunderstood by the participants.

In a laboratory experiment using a student sample of 93 persons, Öhrström, Björkman, and Rylander (1988) had subjects determine the levels at which three artificial sounds became ''unpleasant.'' Noise-sensitive subjects tended to report discomfort at lower sound-pressure levels, resulting in a correlation of $r=-0.25$ (see their Appendix) between discomfort thresholds and noise sensitivity as determined either by a Swedish translation of Weinstein's (1978) questionnaire or a single-item self-rating. Likewise, Dornic, Laaksonen, and Ekehammar (1990) found correlations ranging from $r=-0.44$ to $r=-0.65$ between noise sensitivity as assessed by Weinstein's scale and the level of three different types of noise that respondents from a student sample ($n=18$) chose as being ''clearly annoying.''

Further evidence for suprathreshold effects of noise sensitivity was presented by Moreira and Bryan (1972). In their Table I, they report a correlation of $r=0.31$ between self-assessed noise sensitivity and the slope of loudness functions determined in an earlier, unpublished study. Furthermore, they had 34 subjects rate three recorded sounds presented at various levels on a scale mixing adjectives refering to loudness and annoyance. When they contrasted their three most noise sensitive with the three least sensitive subjects, their graphs (Moreira and Bryan, 1972, Figs. 3 and 4) seem to indicate that the former assign higher ratings at low SPLs, thus producing psychophysical functions with shallower slopes, which appears to be at variance with their earlier loudness scaling data.

C. Rationale for the present investigation

In our view, conclusive assessment of the relationship between noise sensitivity and auditory functioning is precluded by two major deficiencies of previous research: (1) Unsatisfactory measurement of noise sensitivity, and (2) insufficient quality and breadth of the psychoacoustical measures employed. These criticisms shall be dealt with in turn.

1. Measurement of noise sensitivity

With the notable exception of the two Swedish studies (Öhrström *et al.*, 1988; Dornic *et al.*, 1990) that used a translation of Weinstein's (1978) 21-item noise-sensitivity questionnaire, the determination of noise sensitivity rested on measures of unknown psychometric quality. In many other studies, noise sensitivity is confounded with noise annoyance, or assessed from single-item self-ratings that have been shown to be inferior to questionnaire-based measurement (Zimmer and Ellermeier, 1999). First of all, such ratings do not meet established psychometric criteria of reliability and validity; second, they are open to *ad hoc* interpretation depending on the context in which they are presented, and finally, they fail to elicit a wide enough range of situations to qualify for the measurement of *general* noise sensitivity (Job, 1999). These properties make them unfit to measure individual differences with the precision needed for correlational analyses.

Therefore, the present investigation employed a recently developed (Zimmer and Ellermeier, 1998a) full-length, German-language noise-sensitivity questionnaire having excellent psychometric properties. Details about this questionnaire are given in Sec. II C 7.

2. Psychoacoustical measurement

A greater hindrance to conclusive assessment of the relationship between noise sensitivity and perceptual acuity is the scarcity of adequate psychophysical measurements in the pertinent literature. The studies reviewed in the previous section clearly do not cover the breadth of psychophysical methods available to assess auditory functioning. They (1) often employ methods that do not represent the state of the art; (2) collect too little data to justify the derivation of individual parameters; (3) make statements based on small numbers of subjects (e.g., selected extreme groups); (4) use stimuli or response categories that confound psychoacoustical performance with noise assessment, and (5) are subject to a host of response biases that makes it difficult to separate sensory from judgmental (or attitudinal) contributions.

Therefore, the present investigation was designed with a test-battery approach in mind: An attempt was made to cover a broad range of methodologies and phenomena including (a) absolute and difference thresholds; (b) suprathreshold reaction-time measurements; (c) a ratio-scaling procedure; and (d) direct ratings of loudness or annoyance. In our view, converging with an integrative perspective recently proposed by Baird (1997, Chaps. 1 and 15), these procedures may be ordered on a hypothetical dimension along which the sensory contribution decreases (from thresholds to ratings, for example) while the judgmental contribution increases accordingly. Furthermore, an effort was made to employ bias-free adaptive procedures which had not been used in this research area thus far, and to include signal-detection analyses which might help to disentangle sensory and judgmental effects of noise sensitivity, if present.

As a result of refining the assessment of noise sensitivity, while simultaneously broadening the collection of psychoacoustical parameters, we hope to make a contribution to the clarification of the concept of noise sensitivity, specifically by probing if it rests on a perceptual basis.

II. GENERAL METHOD

A. Subjects

An unselected sample of 61 volunteers, most of whom were students at the University of Regensburg, participated in the experiments. This sample had a median age of 24 years (range 19–37 years). Care was taken to recruit an approximately equal number of female ($N=33$), and male ($N=28$) participants.

All subjects were audiometrically tested using Békésy tracking at the standard audiometric frequencies (0.5–8 kHz). Two of the 61 subjects had a hearing loss exceeding 30 dB HL with respect to the ANSI (1996) standard for at least one of the frequencies tested; an additional eight subjects showed losses greater than 20 dB HL. These ten slightly impaired subjects were not excluded from the analyses, since restricting the range of audiometric performance might have weakened correlations with noise-sensitivity scores. Separate analyses revealed, however, that the relationships observed in the present study were not affected by the marginal impairments found in the audiometric screening test.

In order to minimize the effect of expectations, the participants were not informed about the central role of self-reported noise-sensitivity in the present investigation. Rather, they were told that the study focused on interrelations between different measures of auditory performance obtained in the laboratory.

B. Apparatus and stimuli

All stimuli—except for the natural sounds used in the annoyance rating experiment (Sec. II C 6)—were computed using a Tucker-Davis-Technologies (TDT) signal processor card (model AP2), and played from a 16-bit digital-analog converter (TDT model DD1) at a sampling rate of 50 kHz. After passing through a low-pass filter set at 10 kHz (TDT model FT5), the signal was adjusted to the proper level by means of two programmable attenuators (TDT model PA4). Upon passing through a headphone buffer (TDT model HB6) the signal was delivered to the subject via audiometric headphones (Beyerdynamic DT 48). A different set of phones (Beyerdynamic DT 550) was used in the annoyance rating experiment. The equipment was calibrated by measuring sound-pressure levels at the headphones using an artificial ear (Bruel & Kjær type 4153) and a sound level meter (Bruel & Kjær type 2610). Subjects were run individually and were seated in a double-walled sound-attenuated chamber throughout the experiments.

C. Procedure

1. Absolute thresholds

Absolute thresholds for 1-kHz tones were determined using a two-interval adaptive forced-choice method as described by Levitt (1971). On each trial, the subject had to decide in which of two observation intervals marked by the consecutive illumination of two LEDs the signal tone had occurred. Immediately after responding via a hand-held unit, the subject received visual feedback as to whether the decision was correct. At the outset of the measurement sequence, the signal—a 200-ms sinusoid having 10-ms rise/decay ramps—was presented well above threshold (at 40 dB SPL). Following two successive correct responses, the level of the signal was decreased; following a single incorrect response, it was increased again (''2-down/1-up rule,'' Levitt, 1971). Initially, level thus varied in steps of 4 dB, but after the first four reversals (changes from decreasing to increasing intensity or vice versa) the step size was reduced to 2 dB. Another eight reversals were collected at this final step size, and their mean was taken as an estimate of the 71%-correct threshold (Levitt, 1971).

Two adaptive threshold determinations were obtained for the right ear of each participant, and two for the left ear with measurements being made in a counterbalanced RLLR (or LRRL) sequence.

2. Intensity discrimination

To obtain a measure of differential sensitivity, we determined each subject's intensity discrimination performance at 1 kHz. An adaptive procedure of identical format as in Sec. II C 1 was used. This time, however, the subject had to tell which of the two observation intervals contained the tone of greater intensity. The standard tone always had a level of 54 dB SPL; the variable comparison was generated by electrically adding the same signal to the standard after passing it through a programmable attenuator. At the outset of the adaptive track, the signal was added in phase at equal level, corresponding to a relative amplitude (RA) of 0 dB, and yielding a level difference ΔL of 6 dB. Subsequently, the relative amplitude was decreased or increased following a two-down/one-up rule (see Sec. II C 1), using an initial step size of 4 dB (RA) which was reduced to 2 dB after the first four reversals. Note, however, that by varying the relative amplitude of the added signal, level differences (ΔL) between standard and comparison amounting to fractions of a dB may be generated (Green, 1988, Table 3-1; leftmost and rightmost columns). Again, the arithmetic mean of the last eight reversals was taken as an estimate of the amplitude difference that would yield 71%-correct responses. For ease of comprehension, the relative amplitudes obtained were converted [Ellermeier, 1996, Eq. (2)] to the intuitively more accessible measure ΔL, the ''just noticeable'' level difference in decibels between standard and comparison.

Again, two adaptive measurements were made for each ear of each participant, with appropriate control of order effects as in Sec. II C 1.

3. Magnitude estimation of loudness

Direct loudness judgments of 1-kHz sinusoids were obtained using the method of magnitude estimation with a fixed standard (also termed ''ratio estimation,'' Gescheider, 1997). On each trial, subjects first heard a 70-dB tone (the ''standard'') which—via instruction—was given a loudness value of ''10'' (the ''modulus''). Following a 2-s interval a second tone was presented, the loudness of which was to be numerically estimated relative to the standard. Both tones had a total duration of 500 ms, including 10-ms rise/fall times. The participants had unlimited time to note their estimate on a chart, then pressed a button to initiate the next trial. In a block of trials, nine sound-pressure levels covering the range from 50 to 90 dB SPL in 5-dB steps were presented in a random sequence. After a block of practice that was discarded, three repetitions of the stimulus set were presented to each subject, in a different random permutation each time.

4. Loudness category scaling

In order to obtain categorical judgments of loudness, the same procedure (stimulus levels, timing, number of trials) as in Sec. II C 3 was used, with the exception that no standard or modulus was presented, and that subjects had to rate their loudness impression on a five-point scale. They entered their judgment by pressing one of five response buttons on a handheld unit which were labeled with the German equivalents of ''very soft,'' ''soft,'' ''medium,'' ''loud,'' and ''very loud.''

5. Simple auditory reaction time

For reaction-time (RT) measurements, the same apparatus and stimuli as in the scaling experiments (Secs. II C 3 and 4) were used with the exception that stimulus duration was shortened to 200 ms. A simple reaction-time paradigm was employed; that is, subjects had to press a key in response to the onset of a tone as fast as they could. Each trial began with the illumination of a warning light for 200 ms. After another 300 ms, an exponentially distributed random foreperiod ranging between 0 and 3000 ms (having an expected value of 500 ms) was initiated which was followed by the presentation of the target tone. Reaction time was measured from the onset of the target tone to the closing of the response key contact. After the subject had made a response, the next trial was started following a 2-s intertrial interval. As in the scaling experiments, trials were permuted in blocks containing all nine sound-pressure levels. Ten such permutations of the nine stimulus levels were presented, yielding a total of 90 RT measurements per subject. Trials resulting in reaction times shorter than 100 ms (anticipations) or longer than 1 s (misses) were repeated at the end of each block.

6. Annoyance ratings

Since annoyance is awkward to assess in a laboratory situation without reference to a focal task with which the annoying sounds may interfere, we decided to have subjects rate the ''unpleasantness'' of ten environmental sounds which a previous study (Ellermeier, Mader, and Daniel, 1997) had shown to be measurable on a unidimensional ratio scale as specified by the ''BTL'' scaling model (Luce, 1959). The ten sounds to be rated were natural, traffic, and industrial noises ranging from ''water running from a faucet'' to the recording of a ''jackhammer.'' The sounds were stored in ''wav'' file format and were—deviating from the general description of the apparatus (see Sec. II B)—played with 16-bit resolution at 22-kHz sampling rate via a ''Soundblaster compatible'' PC sound card. After adequate amplification they were diotically delivered via Beyerdynamic DT 550 headphones. The sounds were presented ''as recorded'' and had vastly different (A-weighted, energy-equivalent $[L_{eq}]$) sound-pressure levels ranging from 60 to 81 dB SPL.

Subjects were asked to rate each sound as either ''not at all unpleasant,'' ''somewhat,'' ''medium,'' ''rather,'' or ''very unpleasant'' by pressing one of five response buttons labeled both verbally and numerically. The ten sounds were presented four times, in a different random order each time.

7. Noise sensitivity questionnaire

Noise sensitivity was assessed using a psychometrically evaluated 52-item questionnaire (''Lärm-Empfindlichkeits-Fragebogen,'' LEF) developed by Zimmer and Ellermeier (1998a). This questionnaire encompasses statements about a wide variety of environmental noises in a range of situations that affects the entire population. The material covers seven content areas: everyday life, recreation, health, sleep, communication, work, and noise in general. The 52 items presented relate to perceptual, cognitive, affective, and behavioral responses towards noise in these contexts. For every item, respondents may choose one of four response options ranging from strong disagreement to strong agreement. In order to correct for response bias, an almost equal number of items is scored in each direction. The questionnaire scores very well on psychometric indices: It has high internal consistency (Cronbach's $\alpha = 0.92$) and retest reliability ($r_{tt} = 0.91$). Though its scope is somewhat broader, it correlates well with the better-known Weinstein (1978) noise-sensitivity scale ($r = 0.79$), and is superior to single-item self-ratings of noise sensitivity (Zimmer and Ellermeier, 1999).

8. Session format

In administering the procedures detailed in Secs. C 1–7, we strictly adhered to the following sequence: In the first session lasting approximately 40 min, data were collected using (1) Békésy audiometry; (2) adaptive intensity discrimination; and (3) category scaling of loudness. The second session lasted 60–70 min including appropriate rest breaks, and involved (4) annoyance ratings of natural sounds; (5) measurement of absolute thresholds; (6) magnitude estimation; (7) simple reaction time; and (8) administering the noise-sensitivity questionnaire (LEF). The two sessions were a minimum of 1 and a maximum of 6 weeks apart. Given the high retest reliability of both noise-sensitivity and psychoacoustical measures, the time lapse between sessions was not considered problematic.

TABLE I. Absolute and difference thresholds for 1-kHz, 200-ms tones as a function of noise sensitivity. Intensity discrimination performance ΔL refers to the level increment in dB required to make a test tone distinguishable from the 54-dB SPL standard.

		N	Mean	Min	Max	s.d.
Absolute thresholds (dB SPL)						
Low noise sensit.	left ear	31	−2.70	−10.05	5.40	4.34
High noise sensit.		30	−0.99	−8.80	23.75	7.08
Low noise sensit.	right ear	31	−2.48	−11.65	6.65	4.98
High noise sensit.		30	0.81	−9.90	30.90	7.98
Intensity discrimination thresholds ΔL [dB]						
Low noise sensit.	left ear	31	1.46	0.52	3.18	0.69
High noise sensit		30	1.38	0.39	3.50	0.80
Low noise sensit.	right ear	31	1.60	0.68	4.31	0.83
High noise sensit.		30	1.57	0.60	5.54	1.10

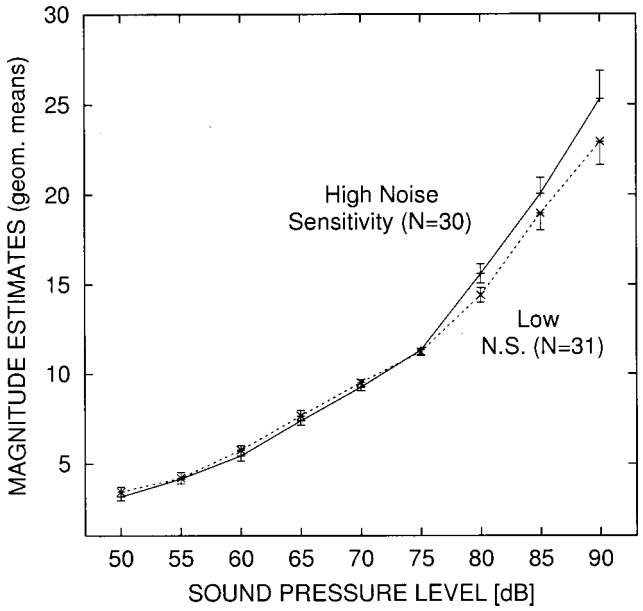


FIG. 1. Mean magnitude estimates made by the two noise-sensitivity groups. Data points are geometric means plus (or minus) one standard error of the mean. These error bars are asymmetric, since they are based on the log transformation of the responses. For legibility, some are plotted in one direction only.

III. RESULTS

A. Noise sensitivity

Overall noise sensitivity, as measured via questionnaire, was normally distributed in the present sample (Kolmogorov–Smirnov test: $z=0.664$, $p=0.77$). The data exhibited a sufficient range of scores (min=44, max=116), and the overall mean ($M=80.18$, s.d.=16.4) agreed well with the mean noise-sensitivity score found in the original student sample ($M=79.4$; Zimmer and Ellermeier, 1998a).

In order to make group-wise comparisons in the various psychoacoustic tasks, the present sample was split along the median ($med=81$) into a group exhibiting ''low noise sensitivity'' ($M_{\text{low}}=67.45$, s.d.=9.84), and one exhibiting ''high noise-sensitivity'' scores ($M_{\text{high}}=93.33$, s.d.=10.22). Interestingly, the ''high noise-sensitivity'' group was dominated by female participants (21 female, 9 male); the ''low noise-sensitivity'' group contained a majority of males (12 female, 19 male). A χ^2 test confirmed that noise sensitivity and gender may not be considered independent in the present sample [$\chi^2(1)=4.82$; $p=0.028$].

B. Absolute and difference thresholds

Mean absolute thresholds for 1-kHz tones as well as difference limens of intensity are given in Table I for the two noise-sensitivity groups, and for left and right ears, separately. Overall, absolute hearing sensitivity of our sample seems to be quite good, with an average threshold value of −1.36 dB SPL. The fact that this measurement is roughly 8 dB lower than published norms is most likely due to the more sensitive adaptive procedures used in the present experiments (Marshall and Jesteadt, 1986; Kollmeier, Gilkey, and Sieben, 1988). The difference thresholds (ΔL) given in the lower portion of Table I, on the other hand, match published values of intensity-discrimination performance quite well (Jesteadt, Wier, and Green, 1977; Florentine, Buus, and Mason, 1987; Green, 1988). In the most recent comprehensive study, employing the same (two-down, one-up) adaptive procedure as the present investigation (except for a longer stimulus duration of 500 ms), Florentine, Buus, and Mason (1987) reported a mean difference limen (ΔL) of 1.42 dB at 50 as well as at 60 dB SPL (see their Table II).

As is evident in Table I, however, there is no indication that the two noise-sensitivity groups formed on the basis of the questionnaire data differ in absolute or differential sensitivity to 1-kHz tones. In no case do the small apparent differences in mean thresholds reach statistical significance, as is confirmed by between-groups t-tests ($p>0.05$). Likewise, when individual noise-sensitivity scores are correlated with individual thresholds, weak and nonsignificant correlations emerge: $r=0.219$, $p=0.091$, for absolute thresholds (averaged across the two ears), and $r=-0.111$, $p=0.396$, for the relationship between noise sensitivity and intensity discrimination (ΔL).

C. Magnitude estimation of loudness

Suprathreshold data on intensity perception were collected by having subjects make direct numerical estimates of the loudness of sinusoids varying in level. Following Stevens' (1975, Chap. 1) recommendations, these magnitude estimates were geometrically averaged both across the three repetitions of each level, and across individuals. The resulting loudness-growth functions are depicted in Fig. 1, separately for the high and the low noise-sensitivity group. These functions are hardly distinguishable, diverging only at the three highest decibel levels with noise-sensitive individuals showing a steeper growth of loudness. The two curves are well fit by psychophysical power functions $\psi=k\phi^{\beta}$ with exponents of $\beta=0.450$; $r^2=0.9986$ for the low noise-

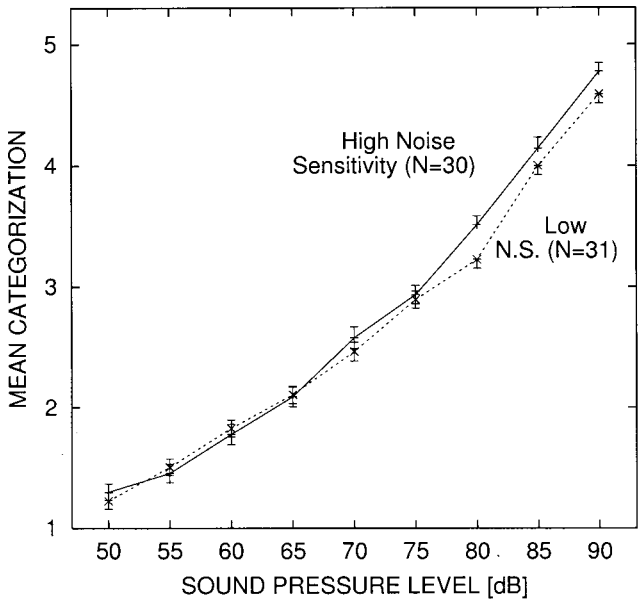


FIG. 2. Mean category ratings of the loudness of 1-kHz tones. Each data point is based on three repetitions of each stimulus level per subject. For better legibility, standard errors are plotted in one direction only at the high sound-pressure levels.

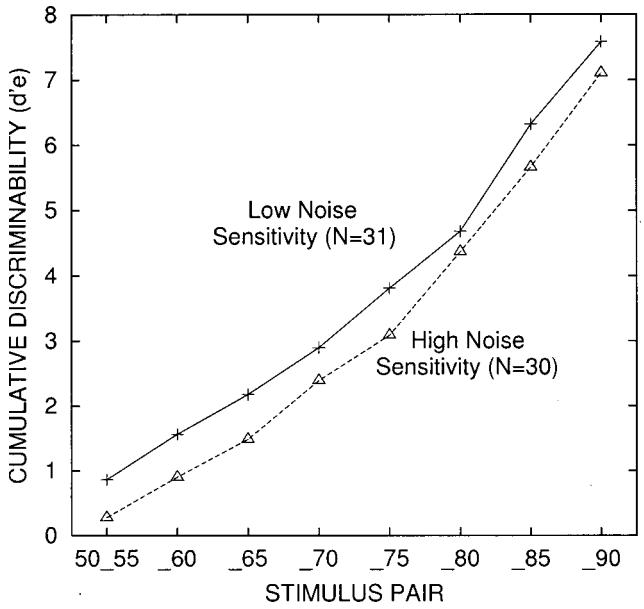


FIG. 3. Cumulative sensitivity (d'_e) computed from the pooled category ratings (see the text). Each entry on the abscissa refers to pair of adjacent stimuli contributing to the cumulative record. For better legibility, only the higher-level member of each pair is given on the abscissa, preceded by a leading underscore.

sensitivity group and $\beta = 0.414$; $r^2 = 0.9964$ for the high noise-sensitivity group. The difference in exponents is not statistically significant, however, as is evident when the individual exponents of the low noise-sensitivity group are compared to those of the high noise-sensitivity group [$t(59) = 1.02$]. More importantly, individually fitted exponents do not correlate significantly with the noise-sensitivity scores obtained by each participant in the questionnaire ($r = 0.10$).

It should be noted, though, that the present ratio estimation experiment may not differentiate groups of subjects or individuals in every respect, since all functions are forced through a fixed point in the center of the curves: the point defined by the standard (70 dB) and the agreed-upon modulus (a judgment of 10). Shifts along the ordinate may not be detected by this implementation. Therefore, other suprathreshold methods were investigated as well.

D. Loudness category scaling

Figure 2 shows the loudness categorizations made by the two groups of participants: Individuals expressing high noise sensitivity appear to assign slightly higher loudness ratings when presented with sound-pressure levels exceeding 75 dB SPL. Using the standard analysis of variance approach in order to evaluate the statistical significance of this divergence was ruled out, since the data showed significant deviations from the normality assumption, especially towards the extreme categories of the rating scale. Therefore, a nonparametric equivalent of a two-factor mixed analysis of variance (Bortz, Lienert, and Boehnke, 2000, Sec. 6.2.5.2) based on Kruskal and Wallis' (1952) *H*-statistic was performed. In the absence of a main effect of noise sensitivity, it revealed a significant (groups by SPL) interaction, $H^*_{A\times B} = 21.541$; $p < 0.01$, thus confirming the statistical significance of the divergence seen at high SPLs.

In order to assess whether the divergence in loudness ratings observed when comparing the two noise-sensitivity groups is due to a true sensory difference or merely to a judgmental artifact, the data were subjected to a signal-detection analysis (see Irwin and Whitehead, 1991; Ellermeier, 1997, for analogous applications in the psychophysics of pain). This is achieved by treating the pooled loudness ratings of the two groups like ''confidence ratings'' in a signal-detection experiment. These serve to trace out ''receiver-operating curves'' (ROCs) from which two parameters may be computed: (1) A discriminability index d'_e indicating the sensory distance between stimuli, and (2) an entirely independent ''bias'' parameter β reflecting a tendency to assign high ratings.

A maximum-likelihood method (Alf and Grossberg, 1987) was used to estimate d'_e for each adjacent pair of stimuli, separately for the two noise-sensitivity groups.[1] Figure 3 shows these measures of sensitivity, cumulated over the stimulus range. There is no indication that the highly noise-sensitive subjects show a greater growth in cumulative discriminability than do the less-sensitive subjects. If anything, the latter group exhibits a slight advantage due to an offset generated by superior discrimination of the two lowest sound-pressure levels. The slopes of the two curves, however, which may be interpreted as indicating the growth in sensation magnitude unbiased by judgmental tendencies (Irwin and Whitehead, 1991), do not seem to differ between the two noise-sensitivity groups, $t(14) = 0.308$, $p = 0.763$.

E. Simple auditory reaction time

Simple reaction time (RT) is often seen as a dependent variable that might tap the underlying sensory processes more directly than various verbal measures used in psychophysics. Furthermore, as a long research tradition has shown

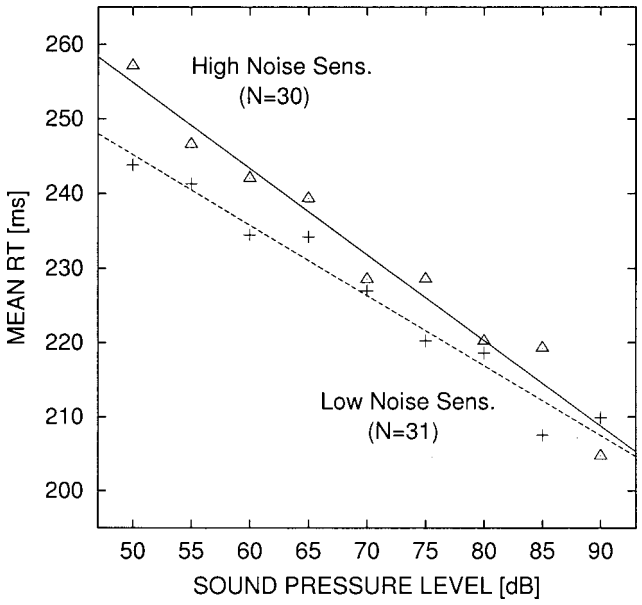


FIG. 4. Mean auditory reaction time (RT) to the onset of tones having SPLs given on the abscissa. Linear functions relating RT and sound-pressure level are fitted to the data of groups of subjects reporting high (triangles) and low (crosses) noise sensitivity. Each data point is based on at least 300 RT measurements.

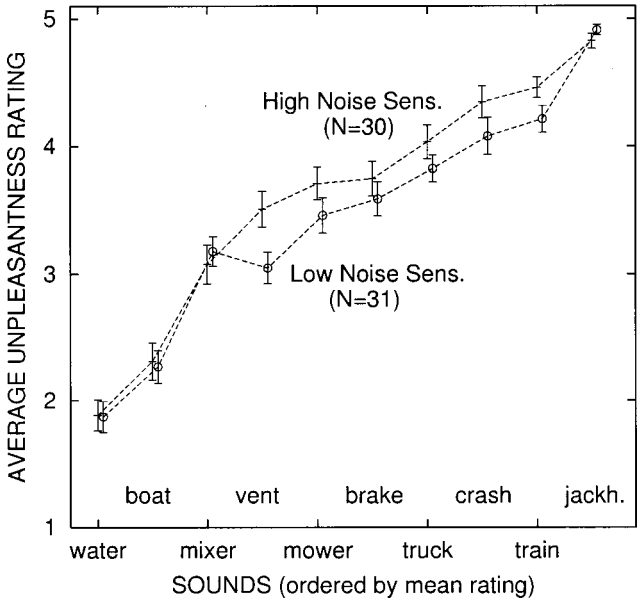


FIG. 5. Unpleasantness ratings of the ten natural sounds identified at the bottom of the figure (for details, see the text). Mean ratings plus/minus one standard error of the mean are given. Note that the abscissa is not metric; sounds are merely ordered according to the mean rating received.

(Chocholle, 1944; Luce, 1986; Kohfeld, Santée, and Wallace, 1981), it may be seen as an indirect way of scaling sensory magnitude via ratio-scale physical measurements. Therefore, reaction times were measured in response to the same stimuli as in the scaling experiments. The parameter of interest is the slope of the function relating decreasing response times to increasing sound-pressure levels.

As is evident in Fig. 4, this slope appears to be somewhat steeper for the subjects reporting high noise sensitivity in the questionnaire. Note that while in the literature (e.g., Kohfeld *et al.*, 1981) Piéron functions (power functions accounting for the steep rise in RT at very low intensities) are used to describe data like these, starting at a clearly audible 50 dB SPL takes us into the linear portion of the RT function (cf. Kohfeld *et al.*, 1981, Fig. 4). Therefore, we fitted functions that are linear over sound-pressure levels (dB) to the group data which are plotted in Fig. 4. The function describing the mean data of the low noise-sensitivity group is $\mathrm{RT} = -0.943 \times \mathrm{SPL} + 292.33$; the function accounting for the high noise-sensitivity data is $\mathrm{RT} = -1.153 \times \mathrm{SPL} + 312.54$. The difference in slope, however, is not statistically significant; either when the two sets of individual slope parameters are compared [$t(59) = 1.368$; $p = 0.177$], or when these parameters are correlated with the individual noise-sensitivity scores obtained from the questionnaires ($r = -0.138$, $p = 0.29$).

F. Unpleasantness ratings

Figure 5 shows mean unpleasantness ratings of ten natural sounds ranging from the sound of water running from a faucet to the noise of a jackhammer. The entries on the abscissa are ordered according to the mean rating given by all 61 subjects. This ordering does not correspond perfectly to the ordering obtained from a paired-comparison methodology in a previous study using the same sounds (Ellermeier, Mader, and Daniel, 1997); the rank correlation between the two orderings, however, is $\rho = 0.81$.

When the ratings of the two noise-sensitivity groups are considered separately (as in Fig. 5), it is evident that the high noise-sensitivity group tends to assign higher unpleasantness ratings once the middle category of the scale is exceeded. The convergence at the top end of the scale is most likely due to a ceiling effect: the sound of the jackhammer receives mean ratings almost indistinguishable from the maximum value of 5. Since this entails violations of the homogeneity-of-variance assumption, a nonparametric test (two-way analysis of variance on ranks; Bortz, Lienert, and Boehnke, 2000) was performed to statistically evaluate the unpleasantness ratings. Both the main effect of noise sensitivity ($H_A^* = 2.66$; $p \approx 0.10$) and its interaction with the stimuli to be rated ($H_{A \times B}^* = 13.975$; $p \approx 0.08$) just failed to reach statistical significance.

Analysis on an individual-subjects level, however, shows unpleasantness ratings to be systematically related to noise sensitivity: When each participant's mean rating (averaging across all ten sounds) is paired with his/her noise-sensitivity score from the questionnaire, a correlation of $r = 0.26$ ($p = 0.042$) emerges between the two measures.

IV. DISCUSSION

The present investigation found noise sensitivity as measured by a psychometrically evaluated questionnaire to be largely unrelated to psychoacoustic indices of auditory functioning. Small, but significant effects of noise sensitivity only emerged in loudness scaling and in ratings of the unpleasantness of sounds; that is, in those tasks most closely related to annoyance (which noise sensitivity was originally conceived to predict).

These findings shall be discussed with respect to (a) the (lacking) psychoacoustical basis of noise sensitivity; (b) the role of gender effects; and (c) the potential of combining indicators to predict increased noise sensitivity.

A. Psychoacoustic correlates of noise sensitivity

Our finding that individuals expressing high or low noise sensitivity do not differ in absolute thresholds agrees well with the two relevant earlier studies (Moreira and Bryan, 1972; Stansfeld *et al.*, 1985). Whereas these studies used classical methods for threshold measurement, the present finding is based on more rigorous, bias-free adaptive procedures. Furthermore, it extends to difference thresholds (see Table I). Thus, there is no indication that individuals expressing increased susceptibility to noise differ in absolute or differential hearing sensitivity. If anything, there is a slight trend (though statistically not significant) for subjects having higher absolute thresholds to express greater noise sensitivity.

It may be argued that most of our more sophisticated psychophysical analyses (adaptive thresholds, scaling, RT) are restricted to data collection at 1 kHz. It is of course possible that noise sensitivity is mediated by increased auditory acuity at other (e.g., higher) frequencies. An attempt was made to assess this possibility by analyzing the Békésy tracking data obtained from all subjects at the standard audiometric frequencies (0.5 to 8 kHz). Those subjects, however, that showed evidence for slight hearing losses in this screening test were roughly equally distributed across the two noise-sensitivity groups: 6 (low) vs 4 (high noise sensitivity) subjects showed hearing losses exceeding 20 dB for at least one of the audiometric frequencies. Reanalyzing the scaling data while excluding these participants did not change the general outcome.

The suprathreshold data collected over a large sound-pressure range encompass (a) magnitude estimation and (b) category scaling of loudness, as well as (c) reaction-time measurements and (d) unpleasantness ratings. Despite the breadth of methodologies used, significant effects of noise sensitivity only emerged in loudness category scaling and in the unpleasantness ratings of natural sounds. The general tendency of noise-sensitive subjects to assign higher loudness categories or greater unpleasantness ratings was largely restricted to higher sound-pressure levels, and thus contradicts the earlier finding by Moreira and Bryan (1972), who found their extreme groups (of $N=3$ each) to *converge* at high SPLs.

To further explore the nature of the small suprathreshold differences emerging in the present investigation, the loudness category ratings were subjected to a signal-detection analysis which enabled us to disentangle sensory and judgmental aspects in the data. Cumulative sensitivity (d') was shown to grow at the same rate for sensitive and nonsensitive participants (see Fig. 3), while computing a bias measure (B; McNicol, 1972) showed the same divergence at high SPLs as did the ''raw'' mean ratings. Thus, from a detection-theory perspective (MacMillan and Creelman, 1991; Irwin and Whitehead, 1991) the apparent noise-sensitivity effect in the loudness scaling data might be interpreted as judgmental rather than sensory in nature: a tendency to assign higher ratings in the absence of a true difference in auditory processing.

This interpretation is consistent with the fact that in the present investigation, it is only in the ''softer,'' more judgmental psychophysical tasks that noise-sensitivity effects are observed, while objectively measured thresholds, RT measurements, and ratio estimates do not show systematic differences. Taken together, the evidence suggests that self-reported noise sensitivity is not related to auditory acuity, but reflects a judgmental, evaluative predisposition towards the perception of sounds.

This predisposition should not be interpreted as simply reflecting a ''response style,'' e.g., a tendency to use extreme categories of a rating scale, as discussed by Job (1999). While individual differences of this sort might indeed create spurious correlations, when single-item ratings of annoyance and sensitivity are related, for example, such an explanation seems unlikely for the present results (a) since different response formats and settings (questionnaire items versus psychophysical judgments obtained in the course of a laboratory experiment) are being compared and (b) since noise sensitivity is measured using a psychometrically sound questionnaire, constructed explicitly to cancel out biases resulting from idiosyncratic response styles.

B. Gender effects

An unusual finding related to the present sample is that we found a significant majority (roughly two-thirds) of women in the noise-sensitive group (see Sec. III A). This is atypical, both for research published by other investigators, who found no effects of sex on noise sensitivity (Moreira and Bryan, 1972; Weinstein, 1978; Taylor, 1984), and for a vast amount of data collected in our own laboratory. In four different samples, three of which were drawn from a similar student population, and all of which consisted of a far greater number of participants (ranging between 117 and 213), we never found a significant effect of gender (Zimmer and Ellermeier, 1997, 1998a, 1998b). Therefore, we tend to interpret the gender imbalance found in the present investigation as a peculiarity of that particular sample.

Interestingly, however, the present data occasionally show significant effects of sex on the form of the psychophysical functions obtained. To distinguish these effects from the consequences of increased noise sensitivity which are the focus of this article, additional analyses were performed. The general strategy was to make gender another factor in the analyses of variance;[2] that is, to inspect effects of noise sensitivity, gender, and sound-pressure level, and their respective interactions. Significant main effects of, or interactions with the participant's gender emerged for only two psychophysical tasks: (1) loudness category scaling, and (2) unpleasantness ratings of natural sounds. The case of loudness category scaling is instructive, since here the effect of gender consists of a discrepancy at *low* sound-pressure levels, evident in a significant interaction [$H^*_{A\times B}=17.04$; $p<0.05$] between the effects of sex and SPL, which is qualitatively different from the divergence at *high* sound-pressure

levels found if subjects are grouped according to their noise-sensitivity scores (see Fig. 2). When rating the unpleasantness of natural sounds, female participants tended to assign higher categories to all sounds presented, leading to a significant main effect of gender [$H_A^* = 7.769$; $p < 0.01$] in that data set. In no case, however, did a two-way or three-way interaction involving both noise sensitivity and gender reach statistical significance, implying that the effects of noise sensitivity are the same in both genders with no need to consider differential effects for male and female participants.

Observing effects of gender in psychoacoustic measures is by no means unusual. Whether they are biological in nature, as the evidence compiled by McFadden (1998) suggests, or whether they reflect different judgmental styles is still a matter of debate. The fact that in the present investigation gender effects show up in the same ''soft'' psychophysical tasks as do the effects of noise sensitivity seems to suggest a similar, judgmental origin.

C. Combining psychoacoustical predictors

Since individual indices of psychoacoustic performance showed only occasional and weak relationships with noise sensitivity, one may ask whether a combination of these indices provides a better prediction. To address this question, a multiple regression analysis was performed, into which potential predictors from all psychoacoustical tasks were entered: (1) mean absolute threshold; (2) the threshold produced by the poorer ear alone; (3) the difference threshold (ΔL); (4) the individual magnitude-estimation exponent; (5) the slope parameter of the loudness category-scaling function; (6) the mean unpleasantness rating of ten sounds; and (7) the slope of the function relating reaction time to SPL. All seven variables in combination account for 15.2%(R^2) of the variance in noise-sensitivity scores. Note, however, that this value is the optimal prediction to be made from the present *sample*. If it is corrected for potential measurement errors to estimate the relationship in the *population*, a disappointingly low ''adjusted'' R^2_{adj} of approximately 0.04 results.

Furthermore, when in a ''stepwise multiple regression'' those parameters that contribute least to the prediction are successively excluded (''backward approach,'' Bernstein, Garbin, and Teng, 1988), we are left with a model according to which (a) the mean unpleasantness rating ($\beta = 0.256$); (b) the threshold in the poorer ear ($\beta = 0.197$); and (c) the category-scaling slope ($\beta = -0.129$) provide the best prediction of individual noise sensitivity. Even though the prediction provided by this model is statistically significant [$F(3,57) = 2.947$; $p = 0.04$]; given the small proportion of variance accounted for ($R^2 = 0.134$; $R^2_{\text{adj}} = 0.089$), it is of little practical relevance. The best thing that may be said for it is that it comes up with the same psychoacoustical parameters as did looking at each task in turn.

V. CONCLUSIONS

The present results clearly refute the conjecture—often traced back to Reason (1972)—that for noise-sensitive individuals ''the world is a brighter, louder, smellier, tastier, heavier, faster, and more painful place than it is for less 'receptive' people'' (Reason, 1972, p. 306). There is no indication in the present data set that noise sensitivity may be attributed to a predisposition to perceive sound events more intensely, or to discriminate between them more accurately. Nevertheless, noise-sensitive participants systematically tended to *judge* the same stimuli as louder or more unpleasant than the less sensitive group, suggesting that what is psychophysically tractable in the concept of noise sensitivity might primarily reflect attitudinal/evaluative rather than sensory components. Furthermore, as an earlier study (Ellermeier and Zimmer, 1997) employing the irrelevant speech paradigm had shown, individual noise sensitivity is only weakly related to objectively measured performance decrements under noise.

Hence, laboratory experiments may help to clarify the concept of noise sensitivity by providing some of the empirical evidence, Job (1999) had called for in his recent review. Even though the central outcome of the present study is negative, showing that self-reported noise sensitivity is *not* related to auditory acuity, the effects observed suggest it to reflect a judgmental, evaluative predisposition towards the perception of sounds. This is consistent with the vast literature relating noise sensitivity to the annoyance produced by unwanted sounds. Some of that work might have to be re-evaluated, however, in the light of the both conceptually, and psychometrically more evolved measures of noise sensitivity available today, which provide better protection against the risk of circularity involved in assessing an individual's noise sensitivity and his or her annoyance produced by an environmental source by posing two very similar questions. Furthermore, the issue of ''specificity'' will have to be addressed, in order to clarify whether noise sensitivity is specific to acoustic nuisances or represents a broader, more general tendency (e.g., Winneke, Neuf, and Steinheider, 1996) to be bothered by environmental stressors.

ACKNOWLEDGMENTS

We would like to thank Till Pfeiffer at Regensburg for providing us with a program to perform the nonparametric equivalents of analyses of variance. Furthermore, we are grateful for the comments made by two anonymous reviewers which helped to improve the quality of the paper.

[1]Actually, d'_e, a measure allowing for different variances of the signal and noise distributions [cf. Macmillan and Creelman, 1991, Eq. (3.8)], was computed. Since overall, however, the ratio of the two variances did not differ significantly from one, d'_e was taken as an index of the more familiar d'.

[2]As in Sec. III D., due to violations of the homogeneity-of-variance assumption, nonparametric equivalents of analyses of variance (Bortz, Lienert, and Boehnke, 2000, Sec. 6.2.5.2) are reported. Using parametric or nonparametric analyses, however, did not affect the statistical conclusions to be drawn from the data analyzed here.

Alf, E. F., and Grossberg, J. M. (**1987**). ''DORF2R.BAS: Analyzing signal-detection theory rating data in the BASIC programming language,'' Behav. Res. Methods Instrum. Comput. **19**, 475–482.

ANSI (**1996**). ANSI S3.6-1996, ''Specification for Audiometers'' (American National Standards Institute, New York).

Baird, J. C. (**1997**). *Sensation and Judgment: Complementarity Theory of Psychophysics* (Erlbaum, Mahwah, NJ).

Bernstein, I. H., Garbin, C. P., and Teng, G. K. (**1988**). *Applied Multivariate Analysis* (Springer, New York).
Bortz, J., Lienert, G. A., and Boehnke, K. (**2000**). *Verteilungsfreie Methoden in dér Biostatistik* [Nonparametric Methods in Biological Statistics], 2nd ed. (Springer, Berlin).
Chocholle, R. (**1944**). ''Etude de la psychophysiologie de l'audition par la méthode des temps de réaction,'' L'Année Psychologique **45**, 90–131.
Dornic, S., Laaksonen, T., and Ekehammar, B. (**1990**). *Noise sensitivity: General self-reports vs. noise effect in laboratory situations* (University of Stockholm, Reports from the Department of Psychology, No. 716).
Ellermeier, W. (**1996**). ''Detectability of increments and decrements in spectral profiles,'' J. Acoust. Soc. Am. **99**, 3119–3125.
Ellermeier, W. (**1997**). ''On separating pain from the willingness to report it,'' Behav. Brain Sci. **20**, 448–449.
Ellermeier, W., Mader, M., and Daniel, P. (**1997**). ''BTL-Skalierung der Unangenehmheit von natürlichen Geräuschen [Scaling the unpleasantness of natural sounds according to the BTL model],'' in *Fortschritte der Akustik—DAGA 97* (DEGA, Oldenburg, Germany), pp. 407–408.
Ellermeier, W., and Zimmer, K. (**1997**). ''Individual differences in susceptibility to the 'irrelevant speech effect,' '' J. Acoust. Soc. Am. **102**, 2191–2199.
Florentine, M., Buus, S., and Mason, C. R. (**1987**). ''Level discrimination as a function of level for tones from 0.25 to 16 kHz,'' J. Acoust. Soc. Am. **81**, 1528–1541.
Gescheider, G. A. (**1997**). *Psychophysics. The Fundamentals*, 3rd ed. (Erlbaum, Mahwah, NJ).
Green, D. M. (**1988**). *Profile Analysis. Auditory Intensity Discrimination* (Oxford, New York).
Irwin, R. J., and Whitehead, P. R. (**1991**). ''Towards an objective psychophysics of pain,'' Psychol. Sci. **2**, 230–235.
Ising, H., Dienel, D., Gunther, T., and Market, B. (**1980**). ''Health effects of traffic noise,'' International Archives of Occupational and Environmental Health **47**, 179–190.
Jesteadt, W., Wier, C. C., and Green, D. M. (**1977**). ''Intensity discrimination as a function of frequency and sensation level,'' J. Acoust. Soc. Am. **61**, 169–177.
Job, R. F. S. (**1988**). ''Community response to noise: A review of factors influencing the relationship between noise exposure and reaction,'' J. Acoust. Soc. Am. **83**, 991–1001.
Job, R. F. S. (**1999**). ''Noise sensitivity as a factor influencing human reaction to noise,'' Noise and Health **3**, 57–68.
Kohfeld, D. L., Santée, J. L., and Wallace, N. D. (**1981**). ''Loudness and reaction time. I,'' Percept. Psychophys. **29**, 535–549.
Kollmeier, B., Gilkey, R. H., and Sieben, U. K. (**1988**). ''Adaptive staircase techniques in psychoacoustics: A comparison of human data and a mathematical model,'' J. Acoust. Soc. Am. **83**, 1852–1862.
Kruskal, W. H., and Wallis, W. A. (**1952**). ''Use of ranks in one-criterion variance analysis,'' J. Am. Stat. Assoc. **47**, 583–621.
Langdon, F. (**1976**). ''Noise nuisance caused by road traffic in residential areas. II,'' J. Sound Vib. **47**, 265–282.
Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Luce, R. D. (**1959**). *Individual Choice Behavior* (Wiley, New York).
Luce, R. D. (**1986**). *Response Times. Their Role in Inferring Elementary Mental Organization* (Oxford, New York).
Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press, Cambridge, UK).
Marshall, L., and Jesteadt, W. (**1986**). ''Comparison of pure-tone audibility thresholds obtained with audiological and two-interval forced-choice procedures,'' J. Speech Hear. Res. **29**, 82–91.
McFadden, D. (**1998**). ''Sex differences in the auditory system,'' Dev. Neuropsychol. **14**, 261–298.
McKennell, A. C. (**1963**). *Aircraft Noise Annoyance Around London (Heathrow) Airport* (Her Majesty's Stationery Office, London).
McNicol, D. (**1972**). *A Primer of Signal Detection Theory* (Allen & Unwin, London).
Moreira, N., and Bryan, M. (**1972**). ''Noise annoyance susceptibility,'' J. Sound Vib. **21**, 449–462.
Öhrström, E., Björkman, M., and Rylander, R. (**1988**). ''Noise annoyance with regard to neurophysiological sensitivity, subjective noise sensitivity and personality variables,'' Psychol. Med. **18**, 605–613.
Reason, J. T. (**1972**). ''Some correlates of the loudness function,'' J. Sound Vib. **20**, 305–309.
Stansfeld, S. (**1992**). ''Noise, noise sensitivity and psychiatric disorder: Epidemiological and psychophysiological studies,'' Psychol. Med. Monograph Suppl. **22**.
Stansfeld, S., Clark, C., Jenkins, L., and Tarnopolsky, A. (**1985**). ''Sensitivity to noise in a community sample. I. Measurement of psychiatric disorder and personality,'' Psychol. Med. **15**, 243–254.
Stansfeld, S. A., and Shine, P. (**1993**). ''Noise sensitivity and psychophysiological responses to noise in the laboratory,'' in *Noise as a Public Health Problem. Proceedings of the 6th International Congress*, edited by M. Vallet (Institut National de Recherche sur les Transports et leur Sécurité, Arceuil, France), pp. 481–484.
Staples, S. L. (**1996**). ''Human response to environmental noise: Psychological research and public policy,'' Am. Psychol. **51**, 143–150.
Stevens, S. S. (**1975**). *Psychophysics. Introduction to its Perceptual, Neural, and Social Prospects* (Wiley, New York).
Taylor, S. M. (**1984**). ''A path model of aircraft noise annoyance,'' J. Sound Vib. **96**, 243–260.
Weinstein, N. D. (**1978**). ''Individual differences in reactions to noise: A longitudinal study in a college dormitory,'' J. Appl. Psychol. **63**, 458–466.
Winnecke, G., Neuf, M., and Steinheider, B. (**1996**). ''Separating the impact of exposure and personality in annoyance response to environmental stressors, particularly odors,'' Environ. Int. **1**, 73–81.
Zimmer, K., and Ellermeier, W. (**1997**). ''Eine deutsche Version der Lärmempfindlichkeitsskala von Weinstein [A German version of Weinstein's noise sensitivity scale],'' Zeitschrift für Lärmbekämpfung **44**, 107–110.
Zimmer, K., and Ellermeier, W. (**1998a**). ''Konstruktion und Evaluation eines Fragebogens zur Erfassung der individuellen Lärmempfindlichkeit [Construction and evaluation of a noise sensitivity questionnaire],'' Diagnostica **44**, 11–20.
Zimmer, K., and Ellermeier, W. (**1998b**). ''Ein Kurzfragebogen zur Erfassung der individuellen Lärmempfindlichkeit [Short form of a questionnaire measuring individual noise sensitivity],'' Umweltpsychologie **2**, 54–63.
Zimmer, K., and Ellermeier, W. (**1999**). ''Psychometric properties of four measures of noise sensitivity: A comparison,'' J. Environ. Psychol. **19**, 295–302.

Predicting speech metrics in a simulated classroom with varied sound absorption

Sylvio R. Bistafa[a] and John S. Bradley
Institute for Research in Construction-Acoustics, National Research Council, 1500 Montreal Road, Ottawa K1A 0R6, Canada

(Received 30 May 2000; revised 22 November 2000; accepted 19 December 2000)

By systematically varying the amount of sound absorption, and the location of the sound-absorbing material in a simulated classroom, it was possible to assess the accuracy of the prediction of speech metrics in quite simple acoustical environments. Predictions of speech level, early-to-late sound ratios (C_{50}) and speech transmission index (STI) values were obtained analytically and with two hybrid ray-based computer programs, RAYNOISE 3.0 and ODEON 4.1. The RAYNOISE predictions were accomplished with a purely specular reflection model and also with a calibrated diffuse reflection model. ODEON uses a parameter called transition order, TO, to change the reflection procedure from purely specular to diffuse for reflections that have orders higher than TO. A parametric study was conducted to determine the best transition order for the ODEON prediction of speech metrics. It was found that the analytical predictions of speech level and C_{50} were on average accurate to about 1 just-noticeable difference (jnd), whereas the analytical predictions of STI were on average within 2 jnd's. ODEON predictions of speech level, C_{50} and STI were on average within 2 jnd's. RAYNOISE predictions of C_{50} and STI with the specular model were on average within 2 jnd's. However, the RAYNOISE predictions of speech level, with both types of reflection models, and the RAYNOISE predictions of C_{50} and STI with the diffuse model had average errors greater than 2 jnd's. The effects of the sound-absorption treatments on the measured speech metric values are also discussed. [DOI: 10.1121/1.1354199]

PACS numbers: 43.55.Br, 43.55.Dt, 43.55.Fw, 43.55.Ka [JDQ]

I. INTRODUCTION

The ultimate acoustical goal in a classroom is adequate speech intelligibility. Sound reflections and ambient noise control speech intelligibility in rooms. Because of this, speech metrics have been proposed to measure the combined effects of room reflections and ambient noise on speech intelligibility. Despite the fact that most of the speech metrics that have been proposed can be measured in real rooms, classroom acoustical performance is seldom specified in terms of these quantities. Reverberation time and ambient noise level are the usual basis of classroom acoustics standards and regulations.

The accuracy of reverberation time predictions using analytical formulas and two commercially available ray-tracing type computer programs were addressed in a previous publication.[1] Equally important is knowledge of the accuracy to be expected when predicting speech levels and speech metrics using these prediction methods.

There are several speech metrics for the evaluation of speech intelligibility conditions in rooms.[2,3] Speech metrics such as the useful-to-detrimental sound ratio U_{50} and the speech transmission index STI[4] have been developed to take into account the effects of room reflections and ambient noise on speech intelligibility. (The useful-to-detrimental ratio U_{50} is a simplification[5] of an earlier proposal by Lochner and Burger,[6] in which the direct sound and early reflected speech sounds are assumed to be useful to increased intelligibility and later-arriving speech sounds and noise are assumed to be detrimental to intelligibility.) In the present study, the accuracy of predicting these speech metrics was determined by taking only the effects of the room reflections into account. In this case, U_{50} reduces to the early-to-late sound ratio C_{50}, and STI is obtained with the noise-to-signal ratio equal to zero.

As occurred with measured reverberation times in the previous work,[1] sound levels, C_{50}, and STI were measured in the simulated classroom with different sound-absorption treatments. By systematically varying the absorptive treatment by adding material of known sound absorption in an acoustically simple environment, it was hoped to address the most basic issues of predicting the suitability of a room for speech. The accuracy of the analytical and the computer predictions was assessed by comparing predictions with measured values in the simulated classroom. Formulas used for the analytical predictions are based on the assumption of a diffuse sound field and ideal exponential decays. Computer predictions were obtained with two commercially available hybrid ray-tracing type programs, RAYNOISE 3.0[7] and ODEON 4.1.[8,9]

As far as the room reflections are concerned, C_{50} and STI are speech metrics capable of taking into account the influence of earlier reflections and single echoes on speech intelligibility. Unlike computer predictions, which are based on impulse responses obtained from various reflections in the room, the analytical formulas considered here for predicting speech metrics are only an approximation to actual condi-

[a]Permanent address: Department of Mechanical Engineering, Polytechnic School, University of São Paulo, São Paulo, Brazil.

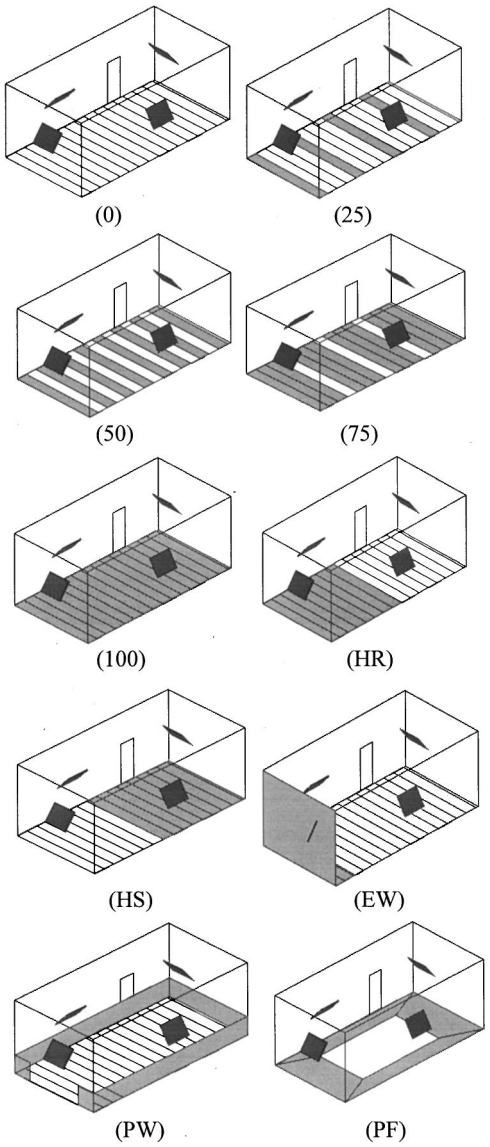


FIG. 1. Schematics depicting the application of the sound-absorbing ceiling tiles in different amounts and configurations. For ease of handling, the ceiling tiles were laid on the floor, to simulate the ceiling absorption normally found in classrooms. Configuration (0) depicts the classroom with no absorption. Configurations (25), (50), (75), and (100) depict the classroom with ceiling tiles applied, respectively, on 26.2%, 52.4%, 78.7%, and 98.3% of the floor area. Configurations (HR), (HS), (EW), (PW), and (PF) depict sound-absorbing configurations that have areas equal to 52.4% of the floor area. Also shown are the diffuser panels and the room door.

tions in real rooms because they are based on ideal exponential decays. The result is that the room reflections are represented by the reverberation time, and this is the only parameter that describes the room acoustical conditions.

The room, the materials, the sound-absorbing configurations, and the measuring system were all described in the previous publication.[1] Figure 1 shows the different sound-absorbing configurations tested in the simulated classroom. For ease of handling, the ceiling tiles were laid on the floor, to simulate the ceiling absorption normally found in classrooms. Configurations included no added absorption (0), 26.2% of floor covered (25), 52.4% of the floor covered (50), 78.7% of the floor covered (75), and 98.3% of the floor covered (100). Configurations (HR), (HS), (EW), (PW), and (PF) illustrated in Fig. 1 all had sound-absorption areas equal to 52.4% of the floor area. The absorbing material consisted of 25-mm-thick glass fibre tiles. The absorption coefficients of the tiles were obtained from a standard reverberation chamber absorption test of the same tiles.

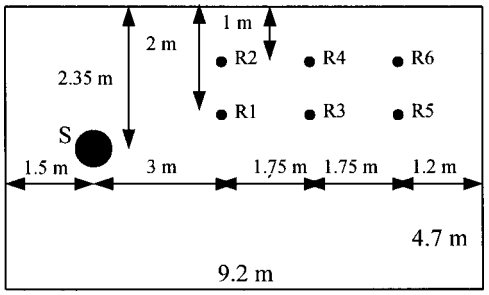


FIG. 2. Plan view of the simulated classroom showing source position (large filled circle) and the six receiver positions (small filled circles).

Figure 2 shows the sound source and the six microphone positions. In the results included here, the microphone was omnidirectional and the sound source was approximately omnidirectional. The room dimensions were 9.20 m long by 4.67 m wide and 3.56 m high. The effect of sound scattering on measured reverberation times was addressed in Ref. 1. For the measurement results reported here, sound scattering was accomplished by four diffuser panels, each with dimensions of 1.2×0.9 m, placed at an approximately 45° angle with the room walls and also illustrated in Fig. 1.

II. SPEECH INTELLIGIBILITY METRIC FORMULAS FOR THE ANALYTICAL PREDICTIONS

A. Measures excluding ambient noise

Assuming a diffuse sound field in the room and ideal exponential decays, an expression for C_{50} can be found in the form[2]

$$C_{50}=10\log\left[\frac{1+(r_h/r)^2-e^{-0.69/T}}{e^{-0.69/T}}\right], \tag{1}$$

where r is the distance from the sound source to the receiver and r_h is the reverberation distance [see Eq. (6) below].

The basis of the STI calculation scheme is the modulation transfer function $m(F)$. Assuming a diffuse sound field in the room, ideal exponential decays, and with the noise-to-signal ratio equal to zero, $m(F)$ is given by[10]

$$m(F)=\frac{\sqrt{A^2+B^2}}{C}, \tag{2}$$

where

$$A=\frac{1}{r^2}+\frac{1}{r_h^2}[1+(2\pi FT/13.8)^2]^{-1}, \tag{3}$$

$$B=\frac{2\pi FT}{13.8r_h^2}[1+(2\pi FT/13.8)^2]^{-1}, \tag{4}$$

$$C=\frac{1}{r^2}+\frac{1}{r_h^2}, \tag{5}$$

where F is the speech modulation frequency and T is the reverberation time.

The reverberation distance r_h, is the distance for which the reflected sound energy density equals the direct sound energy density. For omnidirectional sound sources, the reverberation distance squared is approximately given by[2]

$$r_h^2 = 0.0032 \frac{V}{T} e^{0.161/T}. \tag{6}$$

For a given room volume V, $m(F)$ is calculated using Eq. (2), for each of the 14 one-third-octave speech modulation-frequency values F from 0.63 to 12.5 Hz, for the reverberation time T. The corresponding STI value is then calculated using a standard procedure given elsewhere.[4]

The speech level L_s for a receiver at a distance r is given by

$$L_s = L_{splm} + 10 \log[r^{-2} + r_h^{-2}], \tag{7}$$

where L_{splm} is the long-term anechoic speech level at 1 m straight ahead of the talker. Speech levels in the classroom were based on a female speaking with a ''normal'' voice, which has been shown to correspond to an A-weighted L_{splm} value of approximately 55 dB.[11]

It should be pointed out that C_{50} and $m(F)$ as given by Eqs. (1) and (2), respectively, and the speech level as given by Eq. (7), with the reverberation distance given by Eq. (6), have all been written in terms of the reverberation time T, which under the assumption of a diffuse sound field and ideal exponential decays is the only parameter that characterizes the room acoustical conditions. Reverberation times for these analytical expressions were obtained using Eyring's formula. (Of course, the assumption of an ideal diffuse field also implies that the reverberant level will be uniform throughout the room.)

The accuracy of the predictions was assessed based on single-number values for the speech level L_s, C_{50}, and STI. That is, L_s and C_{50} values were obtained by summing, on an energy basis, A-weighted levels in the octave frequency bands from 125 Hz to 4 kHz. These will be referred to as $L_s(A)$ and $C_{50}(A)$, respectively. STI requires the use of specific octave-band weighting factors to obtain the corresponding STI value as a single number.[4] This was done for each receiver and for each sound-absorbing configuration in the simulated classroom.

B. Measures including ambient noise

Although ambient noise is known to be a major determinant of speech intelligibility,[12] the comparisons of predicted and measured speech metrics in this paper are limited to cases that do not include the effects of ambient noise.

However, the combined effects of the room reflections and ambient noise on speech intelligibility can be evaluated by means of the useful-to-detrimental sound ratio U_{50} and STI *with noise*. Once the speech level L_s has been predicted and the ambient noise L_n has been specified, U_{50} can be obtained analytically by adding to the denominator of Eq. (1) the noise-to-signal ratio $10^{(L_n - L_s)/10}$. Similarly, U_{50} can be obtained from the computer predictions of definition D by also adding to the denominator of Eq. (8) the noise-to-signal ratio $10^{(L_n - L_s)/10}$. STI with noise can be obtained analytically by multiplying the modulation transfer function $m(F)$, as given by Eq. (2), by an additional modulation-reduction factor m due to the ambient noise.[4] Here, m is given by $m = [1 + 10^{(L_n - L_s)/10}]^{-1}$. Once the ambient noise has been specified, RAYNOISE and ODEON have the capability of yielding values of STI *with noise*.

III. PROCEDURES FOR THE COMPUTER PREDICTIONS

As discussed in Ref. 1, the accuracy of reverberation time predictions using two hybrid ray-tracing type computer programs RAYNOISE 3.0 and ODEON 2.6 was determined using purely specular models and calibrated diffuse models. The essence of the calibration procedure was, for a specific room, to match predicted and measured reverberation times using the diffusion coefficient(s) as a ''calibrating tool''.[1] This calibration was achieved by systematically adjusting the surface diffusion coefficients until the best agreement between predicted and measured reverberation times was obtained for configuration (50) (see Fig. 1). The resulting diffusion coefficients were then used in all other calculations with the same computer program.

Unlike RAYNOISE, ODEON uses another diffusion-related parameter, the transition order (TO), to change the reflection procedure from purely specular to diffuse, for reflections that have orders higher than TO. For reflection orders lower than TO, the reflections are treated as purely specular, to preserve the specular influence of the room for these earlier reflections. With the possibility of changing TO, the diffusion coefficient is not the sole parameter that accounts for the effect of diffusion in this computer program.

Lam[13] investigated the dependence of diffusion parameters for ODEON 2.5 on room acoustical quantities, by comparing predicted and measured values in physical scale models of auditoriums. He found that the most significant effect of the diffusion coefficient was on the prediction of reverberation time. The effect of TO on sound level and on clarity C_{80} was relatively small—less than 0.5-dB changes in the prediction errors with the diffusion coefficient varying between 0.1 and 0.5. Lam[13] found, using ODEON, that reverberation time prediction was insensitive to variations in TO. This was expected because reverberation time in ODEON is calculated from the slope of the decay between -5 and -35 dB, and changes in TO only affect the details of particular early reflections and hence the early part of the decay. One of his conclusions was that matching predicted to measured reverberation times provided a convenient means for determining the value of the diffusion coefficient required. He then investigated the effect of TO in terms of C_{80} values. In rectangular rooms a TO value of 0 in the low frequency bands, and between 1 and 3 in the high frequency bands, was found to be appropriate.

An addition to the newer version 4.1 of ODEON is the calculation of STI. Like the early-to-late sound ratio C_{50}, STI is also sensitive to variations in the earlier reflections. A parametric study was conducted to investigate the effects of changing TO on $L_s(A)$, $C_{50}(A)$, and STI. Because ODEON

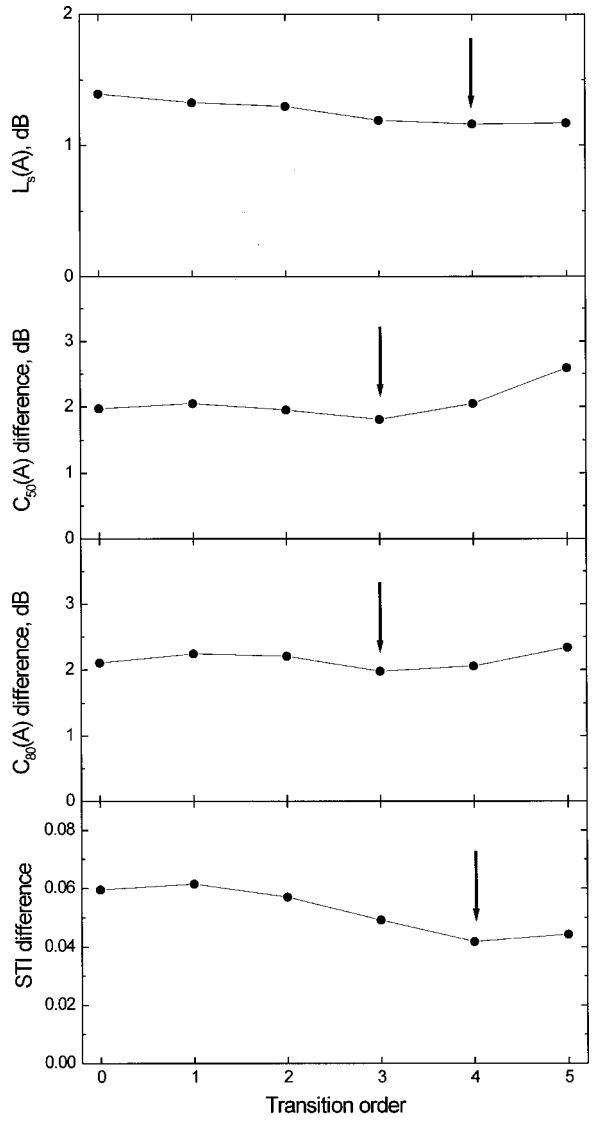


FIG. 3. Overall average absolute differences versus transition order TO for the ODEON predictions of $L_s(A)$, $C_{50}(A)$, $C_{80}(A)$, and STI. The arrows indicate the smallest difference.

predicts clarity C_{80} as well, the effect of changing TO on this acoustical quantity was also examined in terms of $C_{80}(A)$. RAYNOISE and ODEON do not give C_{50} directly. However, both programs yield results of the early-to-total sound ratio called definition D, which is related to C_{50} according to

$$C_{50}=10\log\left(\frac{D}{1-D}\right). \tag{8}$$

For each TO, and for each acoustical quantity, the error of the prediction was assessed in terms of the average of the absolute differences between measured and predicted values for the six receivers and for the ten sound-absorbing configurations in the simulated classroom. For each TO, this resulted in 60 values for determining the average errors of the prediction for each acoustical quantity.

Figure 3 shows these average errors plotted versus transition order TO for the ODEON predictions of $L_s(A)$, $C_{50}(A)$, $C_{80}(A)$, and STI. It can be seen that TO has little influence on the prediction of $L_s(A)$, with the difference reducing only slightly from 1.39 to 1.17 dB as TO increases from 0 to 5, and with a minimum of 1.16 dB with TO equal to 4. TO has similar effects on the prediction of both $C_{50}(A)$ and $C_{80}(A)$, and predictions of these two early-to-late sound ratios with TO varying from 0 to 4 have practically equal differences of approximately 2 dB. However, the differences in the predictions of $C_{50}(A)$ and $C_{80}(A)$ increase, respectively, to 2.59 and 2.34 dB with TO equal to 5. The smallest differences, with values of 1.81 dB for $C_{50}(A)$ and of 1.98 dB for $C_{80}(A)$, were obtained with TO equal to 3. Predictions of STI with the smallest difference of 0.042 were obtained with TO equal to 4.

With ODEON, the accuracy of the predictions will be assessed with TO values that, from the results of the parametric study, gave the smallest error for the respective acoustical quantity. As found above, these correspond to a TO equal to 3 for the $C_{50}(A)$ predictions, and to a TO equal to 4 for both $L_s(A)$ and STI predictions.

The accuracy of the RAYNOISE predictions of speech metrics was assessed for two different reflection models: the purely specular reflection model and the calibrated diffuse reflection model. However, only the calibrated diffuse reflection model was used with ODEON for assessing the accuracy of predicting speech metrics. This is because, for this computer program, as discussed above, the main influence of the type of reflection model is on the prediction of sound decays. Also, as shown in Ref. 1, with RAYNOISE, both types of reflection models have predicted reverberation times with similar accuracy, while the ODEON calibrated diffuse reflection model was definitely far more accurate than its purely specular model for predicting reverberation times. It was then assumed that the better accuracy of the ODEON diffuse model for predicting reverberation times would also be best for the prediction of speech metrics.

IV. ACCURACY OF THE PREDICTIONS—RESULTS AND DISCUSSION

The various predictions are compared with measured values for each absorption configuration. The accuracy of the predictions will be assessed in terms of the just-noticeable difference (jnd) for each of the acoustical quantities considered. For practical calculations, the jnd for sound levels can be approximated as 1 dB.[14] Bradley *et al.*[15] found in simulated sound fields in an anechoic room that the jnd for the early-to-late sound ratio C_{50} is 1.1 dB. In the same work, the jnd for the speech transmission index STI was found to be 0.03.

A. Comparisons between measured and predicted values

The average, over all receivers, of the differences between measured and predicted values was calculated for each sound-absorbing configuration in the simulated classroom. This gives the accuracy for the prediction of the room average value. Figures 4–6 show, for each sound-absorbing configuration and for each prediction method, the average of the differences together with the standard deviation of the differences (indicated by the vertical bars) for the prediction of

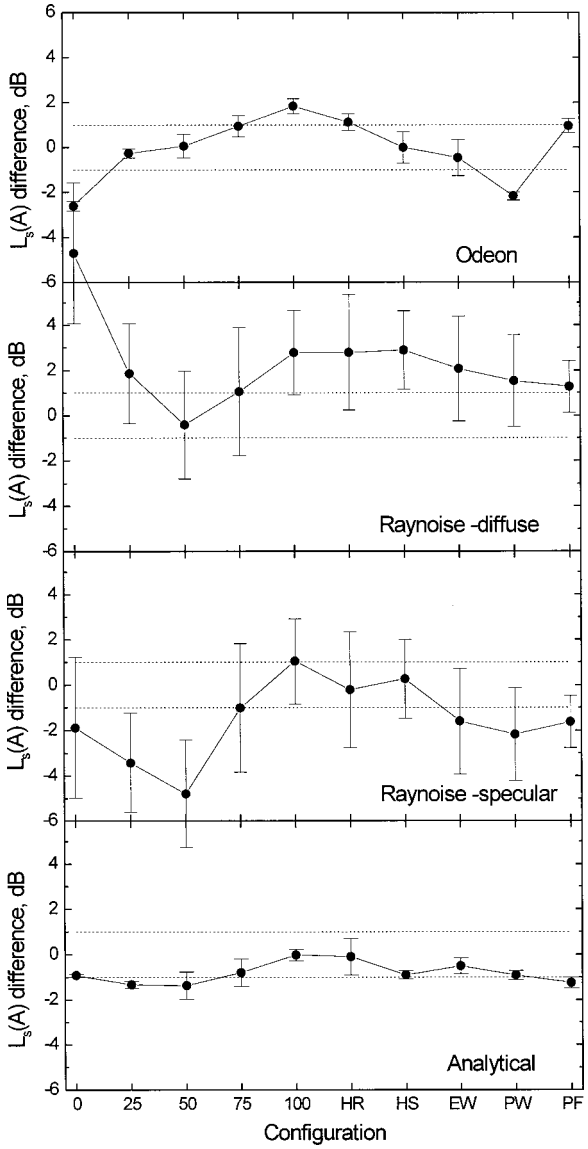


FIG. 4. Averages with respect to receivers and standard deviations (vertical bars) of the differences between measured and predicted speech levels $L_s(A)$ by each prediction method and for each sound-absorbing configuration in the simulated classroom. The just-noticeable difference, jnd, for sound levels of ± 1 dB is drawn as dotted lines across the plot area.

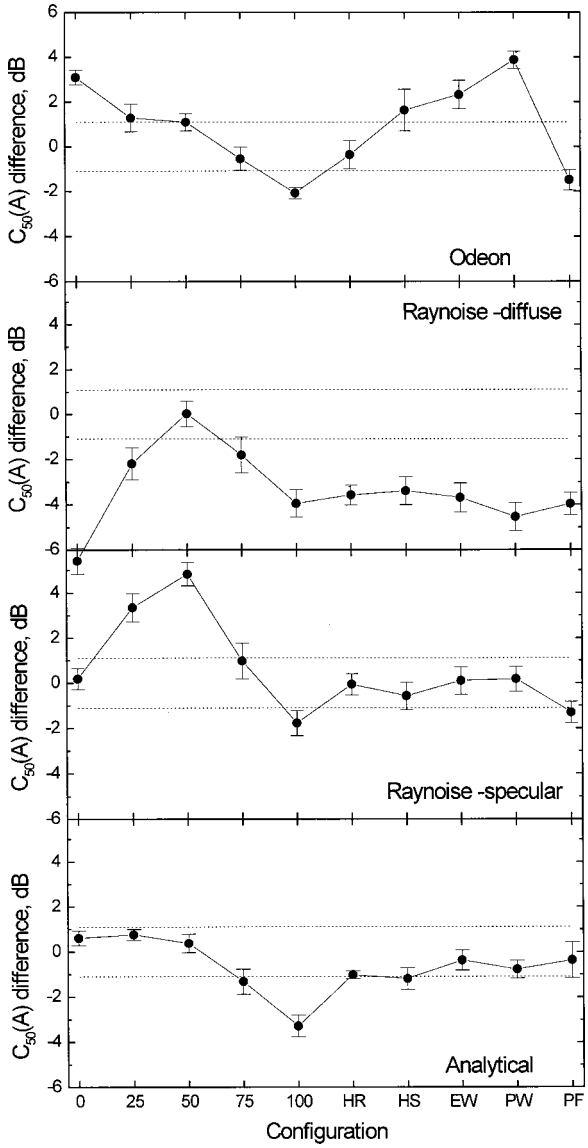


FIG. 5. Averages with respect to receivers and standard deviations (vertical bars) of the differences between measured and predicted $C_{50}(A)$ by each prediction method and for each sound-absorbing configuration in the simulated classroom. The jnd for C_{50} of ± 1.1 dB is drawn as dotted lines across the plot area.

$L_s(A)$, $C_{50}(A)$, and STI. In these figures, the jnd for the respective acoustical quantity is shown as a dotted line across the plot areas.

Figure 4 reveals that the analytical predictions of $L_s(A)$ values are within one jnd for seven out of ten configurations. Only four out of ten of the RAYNOISE predictions with the specular model are within 1 jnd, whereas for the RAYNOISE predictions with the diffuse model, all configurations, with the exception of configuration (50), have differences greater than 1 jnd. The ODEON predictions are within 1 jnd for six out of ten of the configurations. The standard deviations of the RAYNOISE predictions are much larger than the standard deviations of the analytical and the ODEON predictions.

Figure 5, which shows the differences for the $C_{50}(A)$ predictions, reveals that the analytical predictions, for all configurations, with the exception of configuration (100), are essentially within 1 jnd. The RAYNOISE predictions with the specular model, for six out of ten configurations, are within 1 jnd, whereas for the RAYNOISE predictions with the diffuse model, only for configuration (50) are the differences within 1 jnd. For seven out of ten configurations, the ODEON predictions have differences greater than 1 jnd. The standard deviations of the analytical predictions are in general smaller than the standard deviations of the computer predictions.

Figure 6, which shows the differences for the STI predictions, reveals that the analytical predictions, for most configurations, have differences greater than 1 jnd. The RAYNOISE predictions with the specular model, for most configurations, and the RAYNOISE predictions with the diffuse model, for all configurations, with the exception of configuration (50), have differences greater than 1 jnd. The ODEON predictions, for half of the ten configurations, are within 1 jnd of measured values. The standard deviations of the ana-

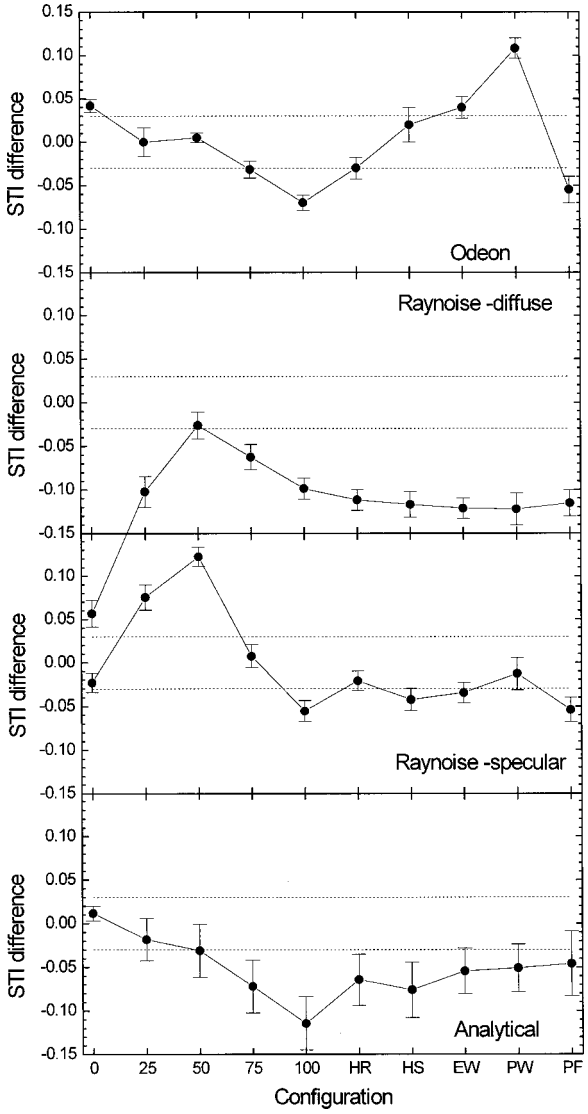


FIG. 6. Averages with respect to receivers and standard deviations (vertical bars) of the differences between measured and predicted STI by each prediction method and for each sound-absorbing configuration in the simulated classroom. The jnd for STI of ± 0.03 is drawn as dotted lines across the plot area.

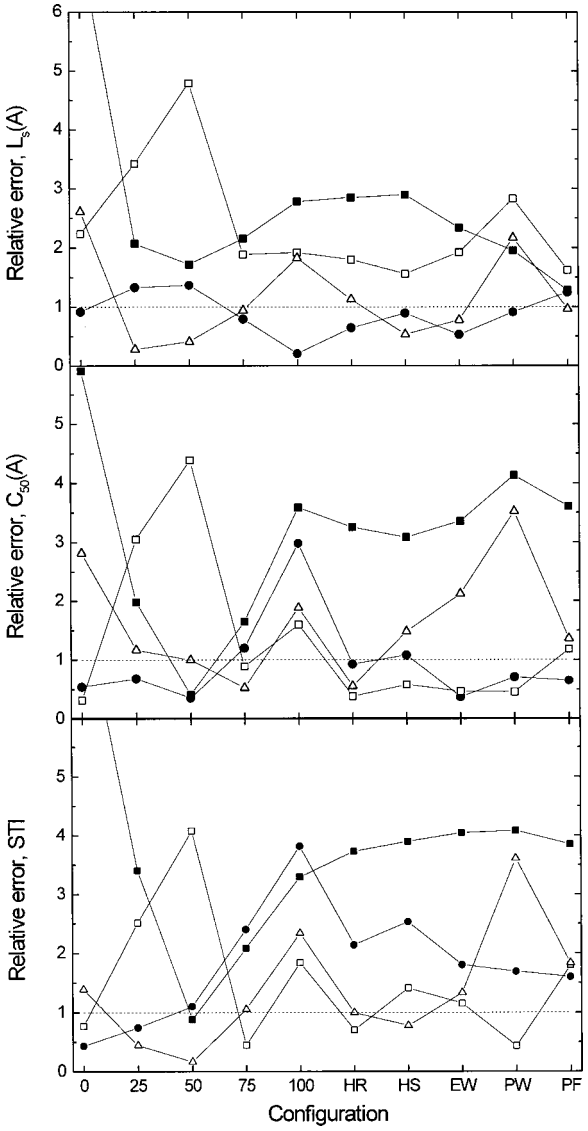


FIG. 7. Receiver-averaged relative errors in terms of the jnd for the prediction of $L_s(A)$, $C_{50}(A)$, and STI for each sound-absorbing configuration in the simulated classroom. Analytical predictions (●), RAYNOISE specular model predictions (□), RAYNOISE diffuse model predictions (■), and ODEON predictions (△).

lytical predictions are in general larger than the standard deviations of the computer predictions.

Because the plots in Figs. 4–6 are for the averages of *simple* differences between measured and predicted values over all receivers, they in fact reveal the accuracy for predicting the room average value. Another way of assessing the accuracy of the predictions is by calculating the average of the *absolute* differences. The average of the absolute differences rather than the average of simple differences gives a more adequate measure of the accuracy of the prediction method. The average of the absolute differences can then be related to the jnd for the acoustical quantity under consideration. These relative errors, for each prediction method and for each acoustical quantity, are shown in Fig. 7 for each sound-absorbing configuration.

For $L_s(A)$ Fig. 7 shows that the analytical predictions have relative errors of less than 1 for seven of the ten configurations and half of the ODEON predictions have relative errors less than 1. The RAYNOISE predictions with either reflection model and for all configurations, have relative errors greater than 1.

For $C_{50}(A)$ Fig. 7 shows that the analytical predictions, for most configurations (seven of ten), have relative errors less than 1. On the other hand, seven out of ten configurations for the ODEON predictions have relative errors greater than 1. For the RAYNOISE predictions with the specular model, six out of ten configurations have relative errors less than 1, whereas for the RAYNOISE predictions with the diffuse model, all configurations, with the exception of configuration (50), have relative errors greater than 1.

For STI Fig. 7 shows, that for all methods, most predictions have relative errors greater than 1. Three of the ODEON configurations, four of the RAYNOISE specular model configurations, one of the RAYNOISE diffuse model configurations, and two of the analytical configurations have relative errors less than 1.

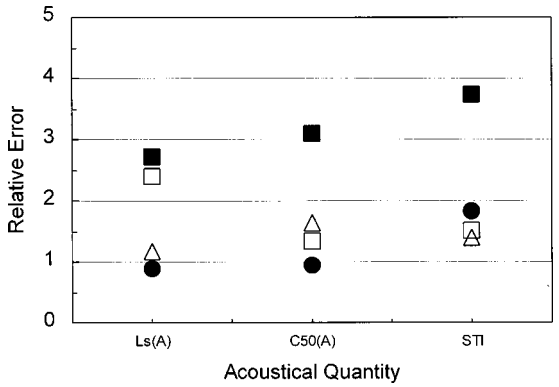


FIG. 8. Overall relative errors in terms of the jnd for the prediction of $L_s(A)$, $C_{50}(A)$, and STI in the simulated classroom. Analytical predictions (●), RAYNOISE specular model predictions (□), RAYNOISE diffuse model predictions (■), and ODEON predictions (△).

It can be seen in Fig. 7 that the relative errors of the RAYNOISE diffuse model predictions of $C_{50}(A)$ and STI were less than 1 only for configuration (50). For all the other configurations, the RAYNOISE diffuse model predictions of both metrics have relative errors much greater than 1. Configuration (50) was used for calibrating the diffuse reflection model, based on the prediction of the reverberation time with minimum relative error.[1] This result seems to suggest that more accurate predictions of $C_{50}(A)$ and STI with RAYNOISE depend on the successful prediction of the sound decays.

B. Accuracy rating

An overall accuracy rating for each of the prediction methods can now be made by averaging the relative errors of Fig. 7 over all sound-absorbing configurations. For each prediction method, these overall relative errors are shown in Fig. 8 for each acoustical quantity. For practical purposes, an overall relative error equal to or less than 2 can be considered as acceptable.[16] As can be seen in Fig. 8, the analytical and the ODEON predictions of $L_s(A)$, $C_{50}(A)$, and STI have overall relative errors less than 2. The analytical predictions of $L_s(A)$ and $C_{50}(A)$ were even more accurate, with overall relative errors of about 1. The RAYNOISE specular model predictions of $C_{50}(A)$ and STI have overall relative errors less than 2. The RAYNOISE predictions of $L_s(A)$ with both reflection models, and the RAYNOISE predictions of $C_{50}(A)$ and STI with the diffuse model have overall relative errors greater than 2. Therefore, with RAYNOISE it seems more appropriate to calculate speech metrics using the purely specular model.

V. THE EFFECTS OF THE SOUND-ABSORPTION TREATMENTS

Figure 9 shows the receiver-averaged measured values of $L_s(A)$, $C_{50}(A)$, and STI, for the ten sound-absorbing configurations in the simulated classroom. As expected, the speech levels, $L_s(A)$, drop as absorption is added, from about 63 dB(A) for the bare room [configuration (0)], to about 53 dB(A) for the most absorbing configuration (100); the latter corresponds to covering the whole ceiling area with sound-absorbing ceiling tiles. For the sound-absorbing configuration (50), which corresponds to covering approximately 50% of the ceiling area with sound-absorbing ceiling tiles evenly distributed, and for the other configurations with the same amount of absorption [configurations (HR), (HS), (EW), (PW), and (PF)], the speech levels $L_s(A)$ are all around 55/56 dB(A).

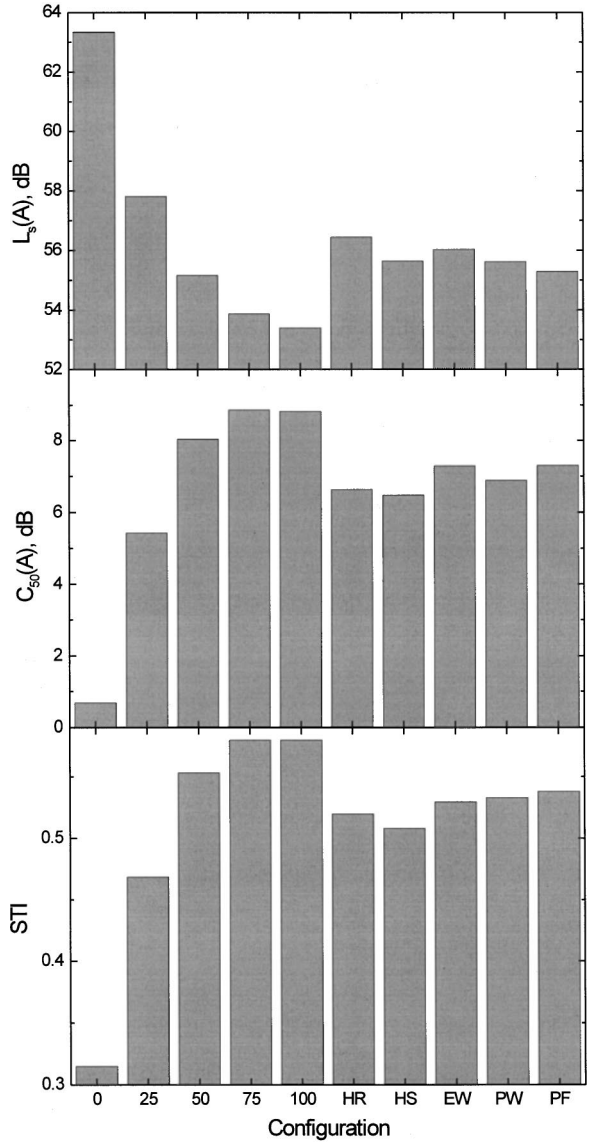


FIG. 9. Receiver-averaged measured values of $L_s(A)$, $C_{50}(A)$, and STI for each sound-absorbing configuration in the simulated classroom.

The speech metrics $C_{50}(A)$ and STI indicate that the most favorable conditions for speech transmission are attained with the most absorbing configurations (75) and (100), with approximately the same values of around 9 dB for $C_{50}(A)$ and 0.58 for STI. There is practically no increase in these speech metric values by increasing the amount of absorption from 75% to 100% of the ceiling area. For the configurations with the same amount of absorption that corresponds to covering approximately 50% of the ceiling area, the highest values of 8 dB for $C_{50}(A)$ and 0.55 for STI were obtained for the sound-absorbing configuration (50). This result indicates that evenly distributing the absorption on the ceiling is the most effective means of increasing the ratio of

early to late arriving speech sounds. Figure 9 reveals that there is a good correlation between measured values of $C_{50}(A)$ and STI.

The overall optimum configuration for good speech intelligibility would correspond to that which increased both $L_s(A)$ and $C50(A)$. Figure 9 suggests that this occurs for configuration EW, in which the absorption is located on the end wall. Recent computer model studies[17] suggested that a configuration similar to (PW) led to optimum results and that ambient noise levels also influenced the choice of optimum condition. However, as in this work, the differences with the same amount of absorbing material in different configurations were quite small.

VI. SUMMARY AND CONCLUSIONS

By varying the amount of sound absorption, and the location of the sound-absorbing material in a simulated classroom, it was possible to assess the accuracy of the prediction of speech metrics. Predictions of $L_s(A)$, $C_{50}(A)$, and STI were obtained analytically and with two hybrid ray-based computer programs, RAYNOISE 3.0 and ODEON 4.1. With RAYNOISE, two types of reflection models were used, a purely specular model and a calibrated diffuse reflection model. With ODEON, only the calibrated diffuse reflection model was used because, for this computer program, the main influence of the type of reflection model is on the prediction of sound decays.

ODEON uses a parameter called transition order, TO, to change the reflection procedure from purely specular to diffuse for reflections that have orders higher than TO. A parametric study was conducted of the effects of transition order on $L_s(A)$, $C_{50}(A)$, $C_{80}(A)$, and STI. This was accomplished by varying TO from 0 to 5 while computing the errors for the prediction of each speech metric for each TO value. It was then found in the simulated classroom that the errors for the prediction of $L_s(A)$, $C_{50}(A)$, and $C_{80}(A)$ were almost independent of the TO value, when it was varied from 0 to 4. For TO equal to 5, the errors for the prediction of $C_{50}(A)$ and $C_{80}(A)$ were a little higher than for the other TO values, while the error for the prediction of $L_s(A)$ remained practically the same for all TO values. However, a slightly smaller error for the prediction of $L_s(A)$ was obtained with TO equal to 4, while slightly smaller errors for the prediction of $C_{50}(A)$ and $C_{80}(A)$ were obtained with TO equal to 3. With respect to the STI predictions, it was found that the smallest error was obtained with TO equal to 4. Based on these results, the accuracy of the ODEON predictions of $L_s(A)$ and STI were then assessed with TO equal to 4, while that of $C_{50}(A)$ was assessed with TO equal to 3.

By averaging the absolute differences between measured and predicted values with respect to both the receiver positions and the sound-absorbing configurations tested in the simulated classroom, the overall average accuracy for the prediction of each speech metric, by each prediction method, was obtained. An accuracy equal to or less than 2 jnd's was considered acceptable for practical purposes. It was then found in the simulated classroom that $L_s(A)$, $C_{50}(A)$, and STI were predicted on average within 2 jnd's analytically and by ODEON. Analytical predictions of $L_s(A)$ and $C_{50}(A)$ were even more accurate, with an average accuracy of about 1 jnd. The RAYNOISE specular model predictions of $C_{50}(A)$ and STI were also on average within 2 jnd's. However, the RAYNOISE predictions of $L_s(A)$, with both types of reflection models, and the RAYNOISE predictions of $C_{50}(A)$ and STI with the diffuse model had average errors greater than 2 jnd's. For these simulated classrooms the simpler analytical predictions were generally more accurate than the computer model predictions. Of course, other computer models may produce quite different results.

No attempt has been made to consider results at individual measurement positions. With the small rooms typical of most school classrooms, systematic effects of distance are usually small and difficult to define because of the increased uncertainty associated with individual position measurement results. While previous studies have considered more complex spaces, the main point of this paper is that it is difficult to accurately predict even room-average speech metrics in these acoustically simple rooms.

ACKNOWLEDGMENTS

The first author is spending sabbatical leave, granted by the University of São Paulo, São Paulo, Brazil, with the Acoustic Group at the National Research Council—NRC Canada. The first author would like to acknowledge the personal financial support for the development of the classroom acoustic project currently underway, granted by Fundação de Amparo à Pesquisa do Estado de São Paulo—FAPESP. The generous donation of the sound-absorbing material by Ottawa Fibre is gratefully acknowledged.

[1] S. R. Bistafa and J. S. Bradley, ''Predicting reverberation times in a simulated classroom,'' J. Acoust. Soc. Am. **108**, 1721–1731 (2000).
[2] S. R. Bistafa and J. S. Bradley, ''Reverberation time and maximum background-noise level for classrooms from a comparative study of speech intelligibility metrics,'' J. Acoust. Soc. Am. **107**, 861–875 (2000).
[3] J. S. Bradley, ''Relationships among measures of speech intelligibility in rooms,'' J. Audio Eng. Soc. **46**, 396–405 (1998).
[4] Objective Rating of Speech Intelligibility by the Speech Transmission Index, IEC Standard 60268-16, 2nd ed., 1998–03.
[5] J. S. Bradley, ''Predictors of speech intelligibility in rooms,'' J. Acoust. Soc. Am. **80**, 837–845 (1986).
[6] J. P. A. Lochner and J. F. Burger, ''The influence of reflections on auditorium acoustics,'' J. Sound Vib. **1**, 426–454 (1964).
[7] RAYNOISE is a registered trademark of LMS Numerical Technologies NV, Leuven, Belgium.
[8] ODEON is the main part of a room acoustic modeing system developed at the Department of Acoustic Technology, Technical University of Denmark, Lyngby, Denmark.
[9] G. M. Naylor, ''ODEON—Another hybrid room acoustical model,'' Appl. Acoust. **38**, 131–143 (1993).
[10] T. Houtgast, H. J. M. Steeneken, and R. Plomp, ''Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics,'' Acustica **46**, 60–72 (1980).
[11] K. S. Pearsons, R. L. Bennett, and S. Fidell, ''Speech Levels in Various Noise Environments,'' Bolt, Beranek, and Newman Report to U.S. EPA, PB-270-053, May 1977.
[12] J. S. Bradley, R. Reich, and S. G. Norcross, ''On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility,'' J. Acoust. Soc. Am. **106**, 1820–1828 (1999).
[13] Y. M. Lam, ''The dependence of diffusion parameters in room acoustics prediction model on auditorium sizes and shapes,'' J. Acoust. Soc. Am. **100**, 2193–2203 (1996).

[14] E. Zwicker and H. Fastl, *Psychoacoustics Facts and Models* (Springer, Berlin, 1990), p. 156.
[15] J. S. Bradley, R. Reich, and S. G. Norcross, ''A just noticeable difference in C_{50} for speech,'' Appl. Acoust. **58**, 99–108 (1999).
[16] M. Vorländer, ''International Round Robin on Room Acoustics Computer Simulations,'' Proceedings of the 15th International Congress on Acoustics, Trondheim, Norway, 26–30 June 1995, pp. 26–30.
[17] R. Reich and J. S. Bradley, ''Optimizing classroom acoustics using computer model studies,'' Can. Acoust. **26**, 15–21 (1998).

Structure-borne noise reduction for an infinite, elastic cylindrical shell[a)]

Sung-Hwan Ko, Woojae Seong, and Sangwoo Pyo
Department of Naval Architecture and Ocean Engineering, College of Engineering, Seoul National University, Seoul 151-742, Korea

(Received 26 July 2000; accepted for publication 26 December 2000)

A theoretical model was developed to evaluate the reduction of structure-borne noise generated by an axially symmetric ring force which is applied on the interior of the cylindrical shell. The vibrating cylindrical shell is coated with a microvoided elastomer that is acoustically soft material designed for the reduction of the generated noise. The analytical model is a two-layer shell structure comprised of a cylindrical shell and an outer layer (coating) that is perfectly bonded to the cylindrical shell. The outer and inner surfaces of the coated shell are in contact with water and air, respectively. The analysis for this problem is based on the theory of elasticity, acoustic wave equations, and pertinent boundary conditions. Effects of various parameters such as coating thickness and material properties on the noise reductions are presented. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1349540]

PACS numbers: 43.58.Vb, 43.20.Tb, 43.40.Rj, 43.30.Nb [SLE]

I. INTRODUCTION

The purpose of this work is to develop a model for evaluating the performance of a microvoided elastomer coating (baffle) in reducing the structure-borne noise generated by an axially symmetric ring force which is applied on the interior of the cylindrical shell. Although the axially symmetric ring force is independent of the angular coordinate, the present analysis allows the angular dependence for the more general use in the future.

Some sonar functions would be facilitated if hydrophone arrays could be mounted directly on the hull of a ship. Some means of noise reduction is required using a baffle layer placed between the hull and the array. Such a baffle must reduce the pressure fluctuations in the water resulting from the hull vibration.

A simple model for evaluating the flexural wave reduction has been studied using a two-layer plane structure that consists of a vibrating plate and a fluid baffle layer by Ko *et al.*[1–3] The acoustic radiation from a coated plate has been studied by Maidanik.[4] Further studies were made by Ko[5] using a similar model that replaces a fluid baffle layer with a microvoided elastomer layer. The use of a compliant material such as a microvoided elastomer for a plane baffle layer is a step closer to reality. However, more realistic models can be developed using a cylindrical shell coated with a microvoided elastomer layer for the conformal large aperture blanket arrays mounted on a ship. The theoretical model in this study is a coated cylindrical shell, i.e., a two-layer structure comprised of a cylindrical shell and an outer layer of a microvoided elastomer (coating) which is bonded to the shell. The outer surface of the microvoided elastomer layer and the inner surface of the cylindrical shell are in contact with water and air, respectively.

This paper describes the analysis for a coated cylindrical shell. The formulation of the problem is based on the theory of elasticity, acoustic wave equations, and pertinent boundary conditions. The results of this work are the parametric studies for the reduction of noise using the assigned values of the baffle layer dimensions and material properties.

The results obtained for a coated cylindrical shell model are compared to those obtained for a plane two-layer model.

II. THEORETICAL ANALYSIS

Figure 1 shows an immersed cylindrical shell excited by an axially symmetric time harmonic ring force. The pressure field in the water region generated by the vibration of the cylindrical shell can be reduced by coating the shell with an acoustically soft baffle such as a microvoided elastomer. Figure 2 depicts a cylindrical shell coated with a microvoided elastomer layer designed for the reduction of the structure-borne noise which is generated by the ring force applied on the interior of the shell. In this study, a two-layer cylindrical structure comprised of a cylindrical shell and an outer layer of a microvoided elastomer is used for the analysis. Then, the case of a one-layer cylindrical shell is deduced from the case of the two-layer composite structure.

A. Acoustic waves in the water region

The governing wave equation in the water region is written as

$$\frac{\partial^2 p_0}{\partial r^2}+\frac{1}{r}\frac{\partial p_0}{\partial r}+\frac{1}{r^2}\frac{\partial^2 p_0}{\partial \theta^2}+\frac{\partial^2 p_0}{\partial z^2}=\frac{1}{c_0^2}\frac{\partial^2 p_0}{\partial t^2}, \tag{1}$$

where $p_0 \equiv p_0(r,\theta,z,t)$ is the pressure field, c_0 is the sound speed, t is the time, and r, θ, and z are the spatial coordinates in the radial, circumferential, and axial directions, respectively. The pressure field in the water region is written as

[a)]Paper presented at the 139th Meeting of the Acoustical Society of America, 30 May–3 June 2000, Atlanta, Georgia [J. Acoust. Soc. Am. **107**, 2839(A) (2000)].

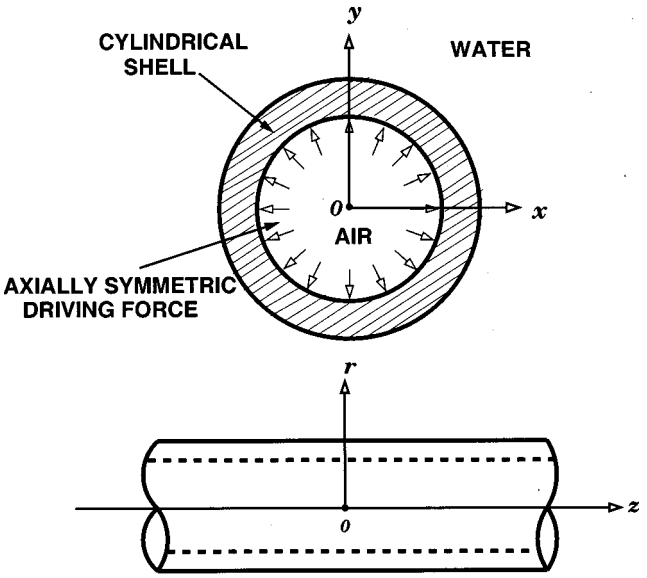


FIG. 1. Vibrating cylindrical shell excited by an axially symmetric ring force.

$$p_0(r,\theta,z,t)=\frac{1}{2\pi}\sum_{n=0}^{\infty}\varepsilon_n\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}P_0^{(n)}(r,k_z,\omega)\cos n\theta \times\exp[-i(k_z z-\omega t)]dk_z d\omega, \tag{2}$$

where $P_0^{(n)}(r,k_z,\omega)$ is the Fourier transform of the pressure field, $\omega=2\pi f$ is the angular frequency in rad/s, f is the frequency in Hz, n is an integer, and ε_n is the Neumann constant ($\varepsilon_n=1$ for $n=0$ and $\varepsilon_n=2$ for $n\geqslant 1$). Substituting Eq. (2) into Eq. (1) gives for the mode n,

$$\frac{\partial^2 P_0^{(n)}}{\partial r^2}+\frac{1}{r}\frac{\partial P_0^{(n)}}{\partial r}+\left[\alpha_0^2-\left(\frac{n}{r}\right)^2\right]P_0^{(n)}=0, \tag{3}$$

where $\alpha_0=(k_0^2-k_z^2)^{1/2}$, $k_0=\omega/c_0$ is the acoustic wave number, and the superscript (n) denotes the circumferential mode number. Equation (3) is Bessel's equation that has the solution as follows:

$$P_0^{(n)}(r)=A_0^{(n)}H_n^{(2)}(\alpha_0 r), \tag{4}$$

where $A_0^{(n)}$ is the unknown coefficient to be determined, $H_n^{(2)}(\alpha_0 r)$ is the Hankel function of the second kind with an argument $\alpha_0 r$ and is given by

$$H_n^{(2)}(\alpha_0 r)=J_n(\alpha_0 r)-iY_n(\alpha_0 r), \tag{5}$$

where $J_n(\alpha_0 r)$ and $Y_n(\alpha_0 r)$ are the Bessel and Neumann functions of order n, respectively. The requirement that the pressure field must propagate outward has been applied in obtaining Eq. (4). For large values of the argument, the Hankel function of the second kind has the following asymptotic expression (as $r\to\infty$):

$$H_n^{(2)}(\alpha_0 r)\sim\left(\frac{2}{\pi\alpha_0 r}\right)^{1/2}\exp\left[-i\left(\alpha_0 r-\frac{n\pi}{2}-\frac{\pi}{4}\right)\right]. \tag{6}$$

The Hankel function of the second kind has been chosen so that the exponential representation of the asymptotic behavior has the form $\exp[-i(\alpha_0 r-\omega t)]$ for the time dependence $\exp(+i\omega t)$. Remember that for $k_0>k_z$ the expression $\alpha_0=(k_0^2-k_z^2)^{1/2}$ is real. Then, Eq. (6) represents the wave that propagates away from the outer surface of the cylindrical shell structure in the radial direction (supersonic wave). If $k_0<k_z$, then the expression α_0 becomes purely imaginary. Then, Eq. (6) becomes the wave that grows as the radial distance increases, which violates the radiation condition. Therefore, the appropriate expression for the wave that decays away in the radial direction from the outer surface of the cylindrical shell structure (subsonic wave) is $\alpha_0=-i(k_z^2-k_0^2)^{1/2}$. Then, the pressure field in the water region is obtained as

$$p_0(r,\theta,z,t)=\frac{1}{2\pi}\sum_{n=0}^{\infty}\varepsilon_n\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}A_0^{(n)}H_n^{(2)}(\alpha_0 r) \times\cos n\theta\exp[-i(k_z z-\omega t)]dk_z d\omega, \tag{7}$$

where

$$\alpha_0=(k_0^2-k_z^2)^{1/2}\quad\text{for}\quad k_0>k_z \tag{8}$$

and

$$\alpha_0=-i(k_z^2-k_0^2)^{1/2}\quad\text{for}\quad k_0<k_z. \tag{9}$$

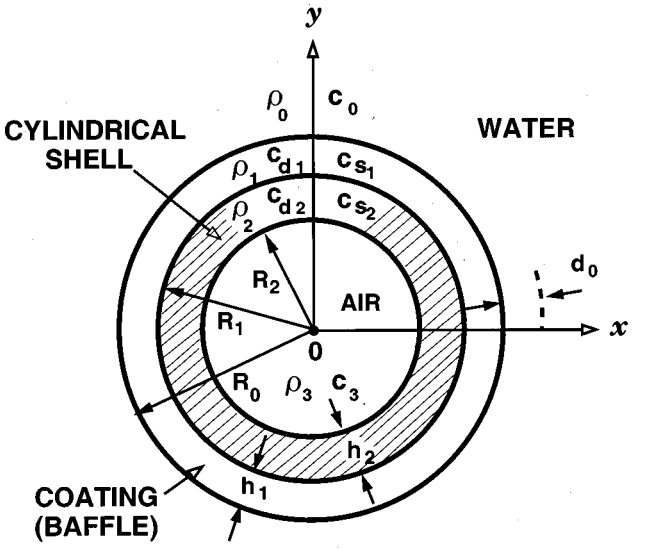


FIG. 2. Vibrating cylindrical shell coated with a microvoided elastomer (baffle layer).

B. Elastic waves in cylindrical shells

In this section, equations necessary for the formulation of the present problem are described. The vector differential equation that governs the small elastic motion in the elastic medium is written as[6–9]

$$\mu\nabla^2\mathbf{u}+(\lambda+\mu)\nabla(\nabla\cdot\mathbf{u})=\rho\frac{\partial^2\mathbf{u}}{\partial t^2}, \tag{10}$$

where λ and μ are the Lamé constants, $\mathbf{u}$ is the displacement vector, ρ is the material density, ∇ is the gradient operator, and ∇^2 is the Laplacian operator. The solutions of Eq. (10) are a combination of a vector potential $\boldsymbol{\Psi}$ and a scalar potential ϕ, so that

$$\mathbf{u}=\nabla\phi+(\nabla\times\boldsymbol{\Psi}), \tag{11}$$

where $\boldsymbol{\Psi}=(\Psi_r,\Psi_\theta,\Psi_z)$ is the vector potential and Ψ_r, Ψ_θ, and Ψ_z are the radial, circumferential, and axial components, respectively. Equation (11) is true provided that ϕ and $\boldsymbol{\Psi}$ are the solutions of the elastic wave equations,

$$\nabla^2\phi=\frac{1}{c_d^2}\frac{\partial^2\phi}{\partial t^2}, \tag{12}$$

$$\nabla^2\boldsymbol{\Psi}=\frac{1}{c_s^2}\frac{\partial^2\boldsymbol{\Psi}}{\partial t^2}, \tag{13}$$

and

$$\nabla\cdot\boldsymbol{\Psi}=F, \tag{14}$$

where F is an arbitrary function of the spatial coordinate and the time. The complex dilatational (c_d) and shear (c_s) wave speeds are given by

$$c_d=\left[\frac{\lambda+2\mu}{\rho}\right]^{1/2} \tag{15}$$

and

$$c_s=\left[\frac{\mu}{\rho}\right]^{1/2}. \tag{16}$$

If Eq. (12) is written in the cylindrical coordinate, then

$$\frac{\partial^2\phi}{\partial r^2}+\frac{1}{r}\frac{\partial\phi}{\partial r}+\frac{1}{r^2}\frac{\partial^2\phi}{\partial\theta^2}+\frac{\partial^2\phi}{\partial z^2}=\frac{1}{c_d^2}\frac{\partial^2\phi}{\partial t^2}. \tag{17}$$

If Eq. (11) is utilized, then the radial displacement (u_r) is written as

$$u_r=(\nabla\phi+\nabla\times\boldsymbol{\Psi})_r=\frac{\partial\phi}{\partial r}+\left(\frac{1}{r}\frac{\partial\Psi_z}{\partial\theta}-\frac{\partial\Psi_\theta}{\partial z}\right). \tag{18}$$

Similarly, the circumferential (u_θ) and axial (u_z) displacements are written as

$$u_\theta=(\nabla\phi+\nabla\times\boldsymbol{\Psi})_\theta=\frac{1}{r}\frac{\partial\phi}{\partial\theta}+\left(\frac{\partial\Psi_r}{\partial z}-\frac{\partial\Psi_z}{\partial r}\right) \tag{19}$$

and

$$u_z=(\nabla\phi+\nabla\times\boldsymbol{\Psi})_z=\frac{\partial\phi}{\partial z}+\left(\frac{\Psi_\theta}{r}+\frac{\partial\Psi_\theta}{\partial r}-\frac{1}{r}\frac{\partial\Psi_r}{\partial\theta}\right). \tag{20}$$

Equation (13) can be decomposed into the radial, circumferential, and axial directions, respectively. In the radial direction, one obtains

$$\left\{\left(\frac{\partial^2\Psi_r}{\partial r^2}+\frac{1}{r}\frac{\partial\Psi_r}{\partial r}+\frac{1}{r^2}\frac{\partial^2\Psi_r}{\partial\theta^2}+\frac{\partial^2\Psi_r}{\partial z^2}\right)-\frac{\Psi_r}{r^2}\right\}-\frac{2}{r^2}\frac{\partial\Psi_\theta}{\partial\theta}=\frac{1}{c_s^2}\frac{\partial^2\Psi_r}{\partial t^2}. \tag{21}$$

In the circumferential direction, one obtains

$$\left\{\left(\frac{\partial^2\Psi_\theta}{\partial r^2}+\frac{1}{r}\frac{\partial\Psi_\theta}{\partial r}+\frac{1}{r^2}+\frac{\partial^2\Psi_\theta}{\partial\theta^2}+\frac{\partial^2\Psi_\theta}{\partial z^2}\right)-\frac{\Psi_\theta}{r^2}\right\}+\frac{2}{r^2}\frac{\partial\Psi_r}{\partial\theta}=\frac{1}{c_s^2}\frac{\partial^2\Psi_\theta}{\partial t^2}. \tag{22}$$

In the axial direction, one obtains

$$\frac{\partial^2\Psi_z}{\partial r^2}+\frac{1}{r}\frac{\partial\Psi_z}{\partial r}+\frac{1}{r^2}\frac{\partial^2\Psi_z}{\partial\theta^2}+\frac{\partial^2\Psi_z}{\partial z^2}=\frac{1}{c_s^2}\frac{\partial^2\Psi_z}{\partial t^2}. \tag{23}$$

The solutions of elastic wave equations given by Eqs. (17), (21), (22), and (23) are written as[8–10]

$$\phi(r,\theta,z,t)=\frac{1}{2\pi}\sum_{n=0}^{\infty}\varepsilon_n\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[A_1^{(n)}J_n(\alpha r)+B_1^{(n)}Y_n(\alpha r)]\times\cos n\theta\exp[-i(k_z z-\omega t)]dk_z d\omega, \tag{24}$$

$$\Psi_r(r,\theta,z,t)=\frac{1}{2\pi}\sum_{n=0}^{\infty}\varepsilon_n\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[A_2^{(n)}J_{n+1}(\alpha r)+B_2^{(n)}Y_{n+1}(\alpha r)]\times\sin n\theta\exp[-i(k_z z-\omega t)]dk_z d\omega, \tag{25}$$

$$\Psi_\theta(r,\theta,z,t)=\frac{1}{2\pi}\sum_{n=0}^{\infty}\varepsilon_n\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[-A_2^{(n)}J_{n+1}(\alpha r)-B_2^{(n)}Y_{n+1}(\alpha r)]\times\cos n\theta\exp[-i(k_z z-\omega t)]dk_z d\omega, \tag{26}$$

and

$$\Psi_z(r,\theta,z,t)=\frac{1}{2\pi}\sum_{n=0}^{\infty}\varepsilon_n\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[A_3^{(n)}J_n(\beta r)+B_3^{(n)}Y_n(\beta r)]\times\sin n\theta\exp[-i(k_z z-\omega t)]dk_z d\omega, \tag{27}$$

where $\alpha=(k_d^2-k_z^2)^{1/2}$, $k_d=\omega/c_d$, $\beta=(k_s^2-k_z^2)^{1/2}$, $k_s=\omega/c_s$, and $A_1^{(n)}$, $B_1^{(n)}$, $A_2^{(n)}$, $B_2^{(n)}$, $A_3^{(n)}$, and $B_3^{(n)}$ are the unknown coefficients to be determined. Note that the superscript (n) denotes the circumferential mode number. The radial, circumferential, and axial stresses are written as

$$\tau_{rr}=\lambda\left(\frac{\partial u_r}{\partial r}+\frac{1}{r}\frac{\partial u_\theta}{\partial\theta}+\frac{u_r}{r}+\frac{\partial u_z}{\partial z}\right)+2\mu\frac{\partial u_r}{\partial r}=(\lambda+2\mu)\left(\frac{\partial u_r}{\partial r}\right)+[(\lambda+2\mu)-2\mu]\times\left(\frac{1}{r}\frac{\partial u_\theta}{\partial\theta}+\frac{u_r}{r}+\frac{\partial u_z}{\partial z}\right), \tag{28}$$

$$\tau_{r_\theta}=\mu\left(\frac{1}{r}\frac{\partial u_r}{\partial\theta}+\frac{\partial u_\theta}{\partial r}-\frac{u_\theta}{r}\right), \tag{29}$$

and

$$\tau_{rz}=\mu\left(\frac{\partial u_z}{\partial r}+\frac{\partial u_r}{\partial z}\right). \tag{30}$$

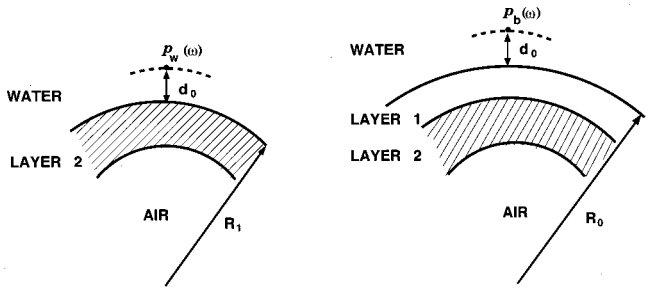


FIG. 3. Geometry used for noise reduction.

C. Acoustic waves and excitation function in the air region

The pressure field inside the cylindrical shell structure is written as

$$p_3(r,\theta,z,t)=\frac{1}{2\pi}\sum_{n=0}^{\infty}\varepsilon_n\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}B_0^{(n)}J_n(\beta_0 r)\cos n\theta \times\exp[-i(k_z z-\omega t)]dk_z d\omega, \tag{31}$$

where $B_0^{(n)}$ is the unknown coefficient to be determined, $\beta_0=(k_3^2-k_z^2)^{1/2}$, $k_3=\omega/c_3$, and c_3 is the speed of sound in air. Note that $\beta_0=(k_3^2-k_z^2)^{1/2}$ for $k_3>k_z$, $\beta_0=i(k_z^2-k_3^2)^{1/2}$ for $k_3<k_z$, $J_n(i\sqrt{k_z^2-k_3^2}r)=i^n I_n(\sqrt{k_z^2-k_3^2}r)$, and I_n is the third kind of the Bessel function of order n.

The axially symmetric ring force applied on the interior surface $(r=R_2)$ is written as

$$f(z,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}F(k_z,\omega) \times\exp[-i(k_z z-\omega t)]dk_z d\omega, \tag{32}$$

where $F(k_z,\omega)$ is the Fourier transform of $f(z,t)$. If the driving force is assumed to be

$$f(z,t)=F_0\delta(z-z_0)\exp(i\omega_0 t), \tag{33}$$

then the Fourier transform of Eq. (33) is obtained as

$$\begin{aligned}F(k_z,\omega)&=\frac{F_0}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(z-z_0)\exp[i(k_z z-\omega t)] \times\exp(i\omega_0 t)dz dt\\&=\frac{F_0}{2\pi}\exp(ik_z z_0)2\pi\delta(\omega_0-\omega)\\&=F_0\delta(\omega_0-\omega)\exp(ik_z z_0),\end{aligned} \tag{34}$$

where F_0 is a constant force per unit area, δ is the delta function, z_0 is the location of the ring force application, and ω_0 is an excitation frequency.

III. FORMULATION OF THE PROBLEM

The reduction of noise is measured by the insertion loss which is defined as the ratio of the pressure at the standoff distance in the presence of the baffle layer to the pressure at the standoff distance in the absence of baffle layer (Fig. 3). Therefore, in the present study, both one- and two-layer models are presented.

A. Two-layer shell model

The geometry used in this section is depicted in Fig. 2. The boundary conditions for order n to be satisfied at the interface between the outer fluid (water) and the coating surface, i.e., $r=R_0$, are written as

$$[\tau_{rr,1}^{(n)}]_{r=R_0}=[-p_0^{(n)}]_{r=R_0}, \tag{35}$$

$$\left[\frac{\partial^2 u_{r,1}^{(n)}}{\partial t^2}\right]_{r=R_0}=-\frac{1}{\rho_0}\left[\frac{\partial p_0^{(n)}}{\partial r}\right]_{r=R_0}, \tag{36}$$

$$[\tau_{r\theta,1}^{(n)}]_{r=R_0}=0, \tag{37}$$

and

$$[\tau_{rz,1}^{(n)}]_{r=R_0}=0, \tag{38}$$

where $[\tau_{rr,1}^{(n)}]_{r=R_0}$ is the radial stress of order n at the surface $r=R_0$ and the subscript 1 refers to the coating (layer 1). Similar expressions are used for the radiated wave pressure, the radial and circumferential displacements, and the tangential and axial shear stresses. Note that ρ_0 is the density of the outer fluid. The boundary conditions to be satisfied at the interface between the cylindrical shell (layer 2) and the coating; i.e., $r=R_1$, are written as

$$[\tau_{rr,1}^{(n)}]_{r=R_1}=[\tau_{rr,2}^{(n)}]_{r=R_1}, \tag{39}$$

$$[\tau_{r\theta,1}^{(n)}]_{r=R_1}=[\tau_{r\theta,2}^{(n)}]_{r=R_1}, \tag{40}$$

$$[\tau_{rz,1}^{(n)}]_{r=R_1}=[\tau_{rz,2}^{(n)}]_{r=R_1}, \tag{41}$$

$$[u_{r,1}^{(n)}]_{r=R_1}=[u_{r,2}^{(n)}]_{r=R_1}, \tag{42}$$

$$[u_{\theta,1}^{(n)}]_{r=R_1}=[u_{\theta,2}^{(n)}]_{r=R_1}, \tag{43}$$

and

$$[u_{z,1}^{(n)}]_{r=R_1}=[u_{z,2}^{(n)}]_{r=R_1}. \tag{44}$$

The boundary conditions to be satisfied at the interface between the cylindrical shell and the inner fluid (air), i.e., $r=R_2$, are written as

$$[\tau_{r\theta,2}^{(n)}]_{r=R_2}=0, \tag{45}$$

$$[\tau_{rz,2}^{(n)}]_{r=R_2}=0, \tag{46}$$

$$\left[\frac{\partial^2 u_{r,2}^{(n)}}{\partial t^2}\right]_{r=R_2}=-\frac{1}{\rho_3}\left[\frac{\partial p_3^{(n)}}{\partial r}\right]_{r=R_2}, \tag{47}$$

and

$$[\tau_{rr,2}^{(n)}]_{r=R_2}=[-p_3^{(n)}]_{r=R_2}+[-f]_{r=R_2}, \tag{48}$$

where ρ_3 is the density of air.

Recall that the driving force is the ring force which acts at $z=z_0$ on the interior surface of the cylindrical shell structure. Since the driving force $f(z,t)$ is independent of θ, all the expressions used in Eqs. (35)–(48) must be independent

of θ. Therefore, the present problem is restricted to the case of $n=0$. Using Eqs. (35)–(48), one obtains a system of linear algebraic equations to be solved for the unknown coefficients of order $n=0$: $A_0^{(0)}$, $A_{1.1}^{(0)}$, $B_{1.1}^{(0)}$, $A_{2.1}^{(0)}$, $B_{2.1}^{(0)}$, $A_{3.1}^{(0)}$, $B_{3.1}^{(0)}$, $A_{1.2}^{(0)}$, $B_{1.2}^{(0)}$, $A_{2.2}^{(0)}$, $B_{2.2}^{(0)}$, $A_{3.2}^{(0)}$, $B_{3.2}^{(0)}$, and $B_0^{(0)}$ in terms of F_0 as shown in Appendix A. After the coefficient $A_0^{(0)}$ has been obtained by solving the algebraic equations in Appendix A, the pressure field at the axial location $z=z_0$ and the standoff distance $r=R_0+d_0$ in the water region, when the driving force is acting at $z=z_0$, is written as

$$p_b(\omega)=\frac{F_0}{2\pi}\int_{-\infty}^{\infty}T_b(k_z,\omega)dk_z, \tag{49}$$

where $T_b(k_z,\omega)$ is the transfer function at $z=z_0$ and $r=R_0+d_0$. Here, z_0 is considered zero since the location of the force application is arbitrary. Therefore, one has

$$T_b(k_z,\omega)=A_0^{(0)}H_0^{(2)}(\alpha_0[R_0+d_0]). \tag{50}$$

Remember that $\alpha_0=(k_0^2-k_z^2)^{1/2}$ for $k_0>k_z$, $\alpha_0=-i(k_z^2-k_0^2)^{1/2}$ for $k_0<k_z$, and the subscript b is referred to the baffle layer (coating).

B. One-layer shell model

The geometry used in this section is depicted in Fig. 1. In the present study, it is necessary to develop a single-layer model to obtain the pressure field at the standoff distance in the absence of the baffle layer. The analysis for the single-layer model follows the procedures used for the two-layer model in the previous section. The boundary conditions to be satisfied at the interface between the outer fluid and the outer surface of the cylindrical shell, i.e., $r=R_1$, are written as

$$[\tau_{rr,2}^{(n)}]_{r=R_1}=[-p_0^{(n)}]_{r=R_1}, \tag{51}$$

$$\left[\frac{\partial^2 u_{r,2}^{(n)}}{\partial t^2}\right]_{r=R_1}=-\frac{1}{\rho_0}\left[\frac{\partial p_0^{(n)}}{\partial r}\right]_{r=R_1}, \tag{52}$$

$$[\tau_{r\theta,2}^{(n)}]_{r=R_1}=0, \tag{53}$$

and

$$[\tau_{rz,2}^{(n)}]_{r=R_1}=0. \tag{54}$$

The boundary conditions to be satisfied at the interface between the cylindrical shell and the inner fluid, i.e., $r=R_2$, are written as

$$[\tau_{r\theta,2}^{(n)}]_{r=R_2}=0, \tag{55}$$

$$[\tau_{rz,2}^{(n)}]_{r=R_2}=0, \tag{56}$$

$$\left[\frac{\partial^2 u_{r,2}^{(n)}}{\partial t^2}\right]_{r=R_2}=-\frac{1}{\rho_3}\left[\frac{\partial p_3^{(n)}}{\partial r}\right]_{r=R_2}. \tag{57}$$

and

$$[\tau_{rr,2}^{(n)}]_{r=R_2}=[-p_3^{(n)}]_{r=R_2}+[-f]_{r=R_2}. \tag{58}$$

Using Eqs. (51)–(58), one obtains a system of linear algebraic equations to be solved for the unknown coefficients for the order $n=0$: $A_0^{(0)}$, $A_{1.2}^{(0)}$, $B_{1.2}^{(0)}$, $A_{2.2}^{(0)}$, $B_{2.2}^{(0)}$, $A_{3.2}^{(0)}$, $B_{3.2}^{(0)}$, and $B_0^{(0)}$ in terms of F_0 as shown in Appendix C. After the coefficient $A_0^{(0)}$ has been obtained by solving the algebraic equations in Appendix C, the pressure field at the axial location $z=z_0$ and the standoff distance $r=R_1+d_0$ in the water region, when the driving force is acting at $z=z_0$, is written as

$$p_w(\omega)=\frac{F_0}{2\pi}\int_{-\infty}^{\infty}T_w(k_z,\omega)dk_z, \tag{59}$$

where $T_w(k_z,\omega)$ is the transfer function at $r=R_1+d_0$ for $z_0=0$ in the absence of the baffle layer and is given by

$$T_w(k_z,\omega)=A_0^{(0)}H_0^{(2)}(\alpha_0[R_1+d_0]). \tag{60}$$

In summary, the noise reduction is written as

$$\mathrm{NR}=20\log_{10}\left|\frac{p_b(\omega)}{p_w(\omega)}\right| \ \mathrm{dB}, \tag{61}$$

where $p_b(\omega)$ and $p_w(\omega)$ are given by Eqs. (49) and (59), respectively (Fig. 3).

IV. NUMERICAL CALCULATIONS AND DISCUSSION

The numerical calculations for noise reductions have been made using Eq. (61). It is necessary to use the properties of the coating and cylindrical shell as well as their dimensions for evaluating the pressure fields $p_b(\omega)$ and $p_w(\omega)$. The major parameters required in the present study are the material density, the complex dilatational (longitudinal) and shear (transverse) wave speeds in the material, and the densities and acoustic wave speeds of the fluid media. The complex dilatational and shear wave speeds for the layer 1 (coating) are written as follows:

$$c_{d1}=c_{d01}(1+i\zeta_{d1})^{1/2} \tag{62}$$

and

$$c_{s1}=c_{s01}(1+i\zeta_{s1})^{1/2}, \tag{63}$$

where c_{d01} and c_{s01} are the dilatational and shear wave speeds in the coating, and ζ_{d1} and ζ_{s1} are the loss factors associated with dilatational and shear waves, respectively. The complex dilatational and shear wave speeds in the layer 2 (cylindrical shell) are written in a similar fashion. The baseline data used for the calculations of noise reductions are as follows:

Inner radius of cylindrical shell	$R_2=2.540$ m (100 in.),
Shell thickness	$h_2=0.0508$ m (2 in.),
Coating thickness	$h_1=0.0254$ m (1 in.),
Water density	$\rho_0=1000\ \mathrm{kg/m^3}$,
Sound speed in water	$c_0=1500$ m/s,
Air density	$\rho_3=1.215\ \mathrm{kg/m^3}$,
Sound speed in air	$c_3=340$ m/s,
Coating density	$\rho_1=600\ \mathrm{kg/m^3}$,
Shear wave speed in the coating	$c_{s01}=50$ m/s,
Dilatational wave speed in the coating	$c_{d01}=200$ m/s,

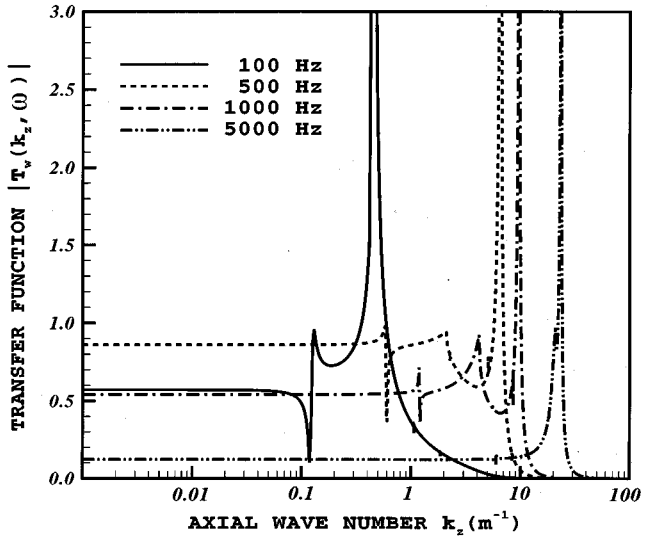


FIG. 4. Transfer functions in the absence of the baffle layer calculated for various frequencies.

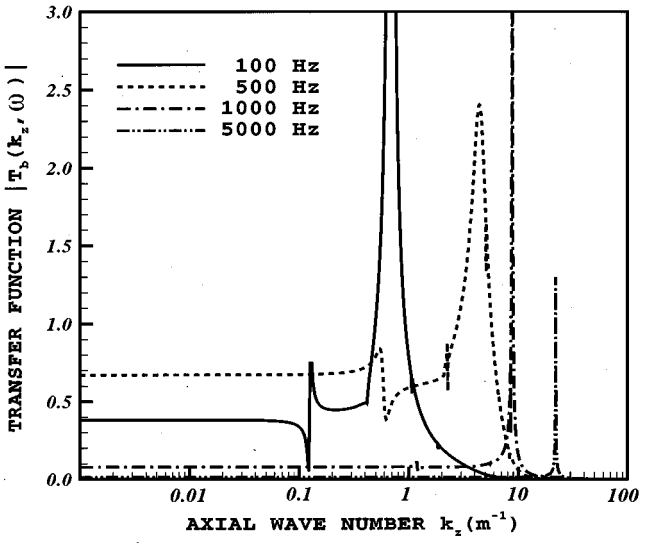


FIG. 5. Transfer functions in the presence of the baffle layer calculated for various frequencies.

Shear loss factor of the coating	$\zeta_{s1}=0.3$,
Dilatational loss factor of the coating	$\zeta_{d1}=0.03$,
Shell density	$\rho_2=7800\,\mathrm{kg/m^3}$,
Shear wave speed in the shell	$c_{s02}=3100\,\mathrm{m/s}$,
Dilatational wave speed in the shell	$c_{d02}=5800\,\mathrm{m/s}$,
Shear loss factor of the shell	$\zeta_{s2}=0.01$,
Dilatational loss factor of the shell	$\zeta_{d2}=0.001$,
Standoff distance	$d_0=0$.

In numerical calculations, as a parameter is varied, all other parameters are kept the same as listed in the baseline data. It is important to evaluate the integrands $T_b(k_z,\omega)$ and $T_w(k_z,\omega)$ to understand the wave-number frequency spectra in this study. Figure 4 presents the transfer functions $T_w(k_z,\omega)$ as a function of the axial wave number k_z for various frequencies. In Fig. 4, the solid, dashed, chain-dotted, and double chain-dotted lines denote the results calculated for 100, 500, 1000, and 5000 Hz, respectively. An observation of the solid line shows that the first cusp is the dilatational wave number in the shell which occurs at $k_d=0.108\,\mathrm{m}^{-1}$. The second cusp is expected to occur at the acoustic wave number in water $k_0=0.418\,\mathrm{m}^{-1}$, but is not shown clearly in the figure. This is because the acoustic wave number is so close to the resonant wave number and thus it is buried within the band of the resonant wave number. Resonant wave numbers here are the wave numbers that are obtainable from the dispersion relationship for the uncoated cylinder. Note that the resonant wave number for 100 Hz is $k_R=0.45\,\mathrm{m}^{-1}$. For 500 Hz, the first cusp occurs at the dilatational wave number $k_d=0.54\,\mathrm{m}^{-1}$. The second cusp is due to the acoustic wave number in water that occurs at $k_0=2.09\,\mathrm{m}^{-1}$. The resonant wave number for 500 Hz is given by $k_R=6.0\,\mathrm{m}^{-1}$. Similarly, one may deduce the results for 1 and 5 kHz from those obtained for 100 and 500 Hz, respectively. It is noticed that the resonant wave number shifts to the right as the frequency increases. Figure 5 presents the transfer functions $T_b(k_z,\omega)$ for various frequencies. The solid line in Fig. 5 shows that the first cusp occurs at $k_d=0.108\,\mathrm{m}^{-1}$ which appears at the same location as shown in Fig. 4. Note that the acoustic wave number in water is $k_0=0.418\,\mathrm{m}^{-1}$ shown as a discontinuity. However, the bandwidth of the resonant wave number at 100 Hz in Fig. 5 is wider than that shown in Fig. 4. Then, the integrated value of the solid line in Fig. 5 is greater than that in Fig. 4. This implies that the coating does not necessarily serve as the noise baffle which is designed for the structure-borne noise reduction. The coating lowers the impedance and increases the velocity when the acting force is the same with and without the coating. Specifically, the low impedance baffle decreases the shell impedance (radial direction) and raises the radial velocity sufficiently to make the pressure in the baffle layer higher than it is in the water without the baffle. How-

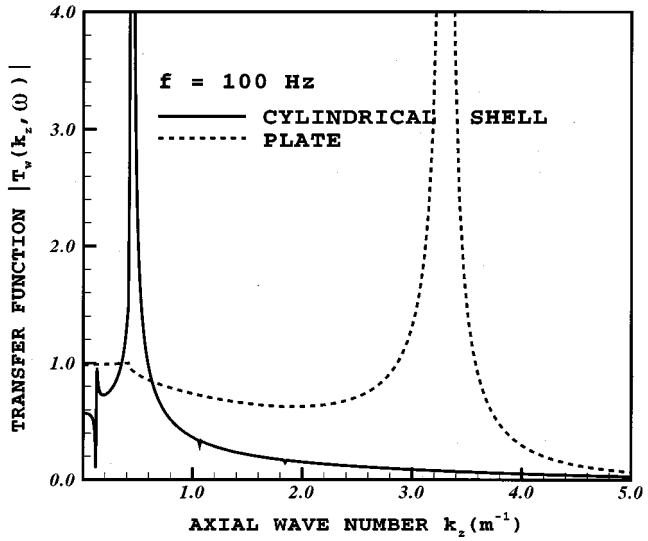


FIG. 6. Comparison of the transfer functions at 100 Hz calculated for a cylindrical shell and an elastic plate in the absence of the baffle layer.

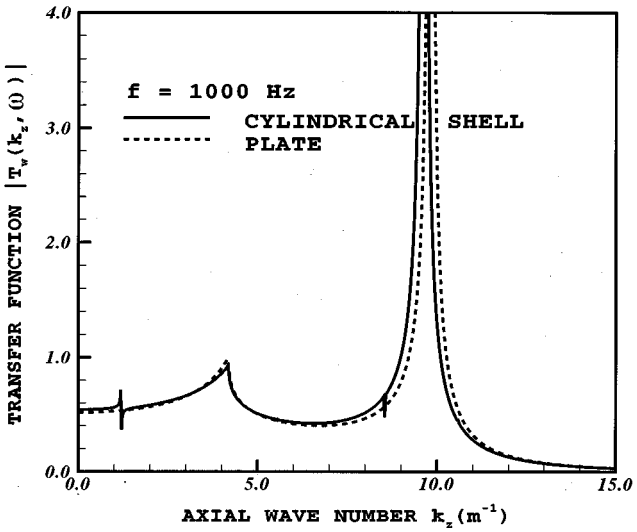


FIG. 7. Comparison of the transfer functions at 1000 Hz calculated for a cylindrical shell and an elastic plate in the absence of the baffle layer.

ever, for 5000 Hz, the area under the double chain-dotted line in Fig. 5 is much less than that in Fig. 4. In this case, the noise reduction is obtainable when the baffle layer is used.

Figure 6 presents a comparison between the transfer function $T_w(k_z,\omega)$ calculated for 100 Hz using the relevant equations for the cylindrical shell and that for the flat plate having the same thickness. As shown in Fig. 6, the difference between the locations of the resonant wave numbers computed at 100 Hz for the cylindrical shell and the flat plate is substantially large. This means that the effect of the cylindrical curvature is significant at this frequency. A similar comparison is shown in Fig. 7 for 1000 Hz. It is found in the figure that the difference between the resonant wave numbers computed at 1 kHz for the cylindrical shell and the flat plate is very small. Further, Fig. 8 shows that the resonant wave numbers computed at 5 kHz for the cylindrical shell and the flat plate are identical. It is found that the analysis of the cylindrical shell for a large value of *ka* (*k* and *a* being a wave

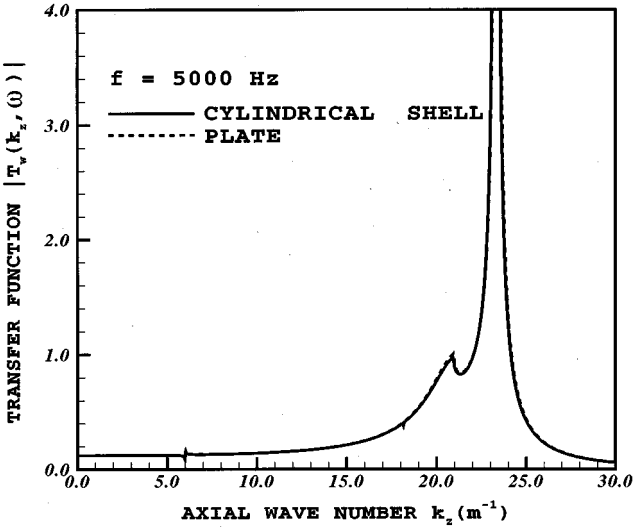


FIG. 8. Comparison of the transfer functions at 5000 Hz calculated for a cylindrical shell and an elastic plate in the absence of the baffle layer.

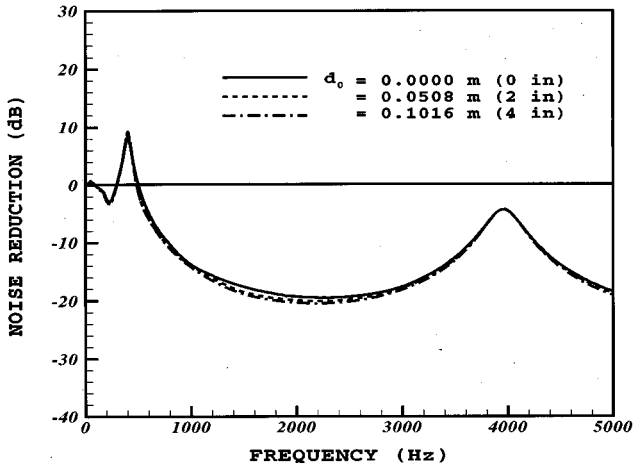


FIG. 9. Effect of the standoff distance on the noise reduction.

number and a shell radius, respectively) can be approximated by the analysis of the flat plate for the high frequency region.

Figure 9 presents the noise reductions calculated at $z_0=0$ using various standoff distances (d_0). As can be seen in the figure that the difference between three standoff distances is unnoticeable. It is reminded that the standoff distances used in Figs. 10–14 are zero as indicated in the baseline data.

Figure 10 shows the noise reduction curves as a function of frequency for different radii (R_2) including infinity that represents the flat plate. It is shown in Fig. 10 that the amounts of the noise reductions are approximately the same for the radii of $R_2=50$ and 100 in. as well as $R_2=\infty$ except for the low frequency region. This is because the value of *ka* is high enough to approximate the cylindrical shell to the flat plate. In the low frequency region, the baffle layer does not reduce the noise but enhances the noise contrary to the purpose of the baffle design. Remember that the areas underneath the transfer functions $T_b(k_z,\omega)$ calculated at 100 and 500 Hz shown in Fig. 5 are larger than the corresponding transfer functions $T_w(k_z,\omega)$ shown in Fig. 4. This provides the positive values in the noise reduction curves. The maximum amount of the noise reduction is obtained at the fre-

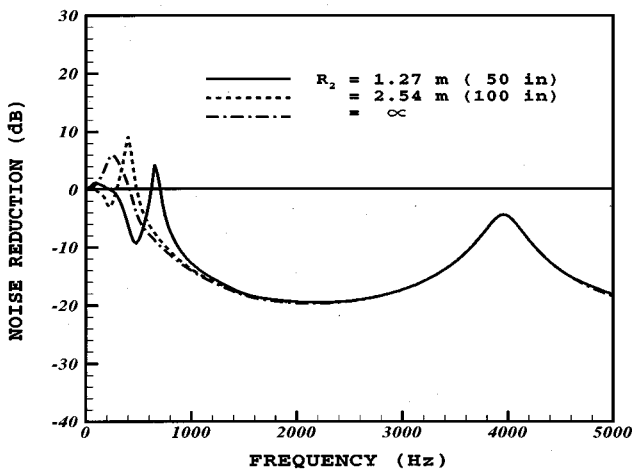


FIG. 10. Effect of the cylindrical shell radius on the noise reduction.

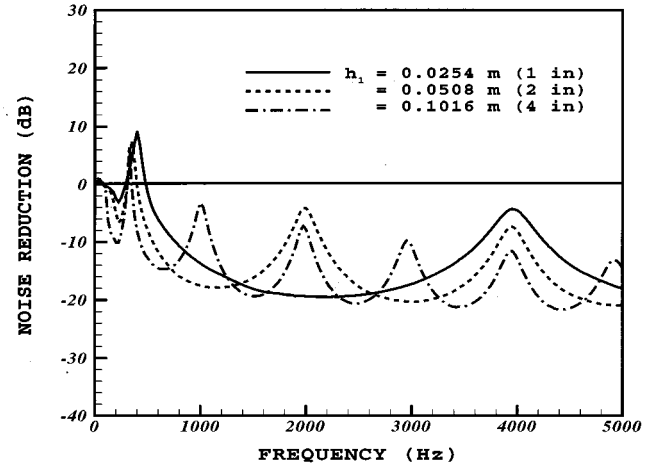


FIG. 11. Effect of the baffle layer thickness on the noise reduction.

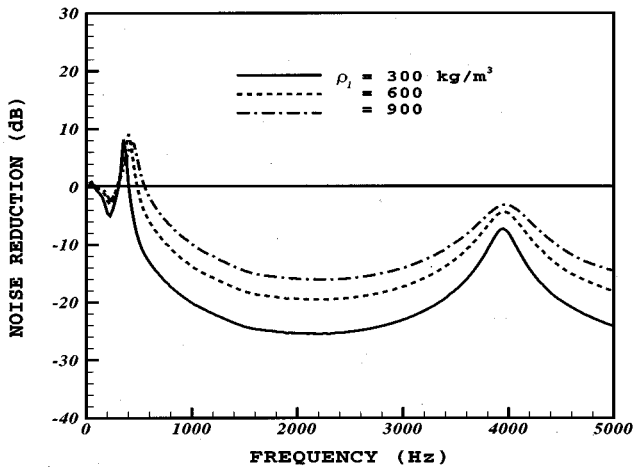


FIG. 12. Effect of the baffle layer density on the noise reduction.

quency where the baffle layer thickness is equal to a one-quarter wavelength. Here, it occurs at 2, 6, 10 kHz, etc. The minimum amount of the noise reduction occurs at the frequency where the baffle layer thickness is equal to a one-half wavelength. Here, it occurs at 4 kHz and its multiples. The sharp peak (positive noise reduction) occurs at the critical frequency (f_c) and is analogous to that of the spring-mass system in mechanical vibration. This critical frequency is written as

$$f_c = \frac{1}{2\pi}\sqrt{\frac{s}{m}}, \tag{64}$$

where s is the spring constant and m is the mass. In this case, $s=(B+4\mu/3)/h_1$, where B is the bulk modulus of the microvoided baffle, h_1 is the baffle layer thickness, and $m=\rho_2 h_2$ is the mass per unit area of the cylindrical shell. The substitution of the appropriate values ($B=2.4\times10^7$ N/m^2, $h_1=0.0254$ m, $\rho_2=7800$ kg/m^3, and $h_2=0.0508$ m) into Eq. (64) yields $f_c=247$ Hz, which falls in the vicinity of the peak shown on the left-hand side of the figure. Note that those peaks for different radii are located at different frequencies. Since this approximation is based on the two-dimensional geometry (flat plate), the location of $f_c=247$ Hz is most close to the noise reduction curve obtained for the case of $R_2=\infty$.

Figure 11 presents the effect of the baffle layer thickness (h_1) on the noise reduction. As mentioned earlier, the maximum reduction occurs at the frequency where the baffle layer thickness is equal to a one-quarter dilatational wavelength, i.e., $h_1=\lambda_{d1}/4$, where λ_{d1} is the wavelength in the baffle layer. The minimum reduction occurs at the frequency where the layer thickness is equal to a one-half dilatational wavelength, i.e., $h_1=\lambda_{d1}/2$. For $h_1=1$ in., as denoted by the solid line the maximum reduction occurs at 2, 6, 10 kHz, etc., whereas the minimum reduction occurs at 4, 8, 12 kHz, etc. For the case of $h_1=2$ in., as denoted by the dashed line the maximum reduction occurs at 1, 3, 5 kHz, etc., whereas the minimum reduction occurs at 2, 4, 6 kHz, etc. Similar results are shown in the figure for the case of $h_1=4$ in. The phenomena shown in the low frequency region are self-explanatory following the earlier discussion for Fig. 10. It is also noted that as the baffle layer thickness increases, the amount of the noise reduction increases depending on the frequency. The layer thickness also controls the frequency locations where the maximum and minimum noise reductions occur.

Figure 12 presents the effect of the density of the baffle layer (ρ_1) on the noise reduction. The maximum and minimum noise reductions occur at 2, 6, 10 kHz, etc. and 4, 8, 12 kHz, etc., respectively, regardless of the assigned values of the density. As shown in Fig. 12, the amount of the noise reduction increases as the density decreases. The phenomena shown in the low frequency region are similar to those discussed earlier. Figure 13 presents the effect of the dilatational wave speed (c_{d01}) on the noise reduction. The frequencies at which the maximum and minimum noise reductions occur are found to be 2, 6, 10 kHz, etc. and 4, 8, 12 kHz, etc., respectively, for the dilatational wave speed $c_{d01}=200$ m/s. Similar discussion is applicable to the cases of $c_{d01}=400$ m/s and $c_{d01}=600$ m/s for the maximum and minimum noise reductions. It is shown in the figure that more noise reduction is achieved as the value of c_{d01} decreases. In general, a light material provides a low dilata-

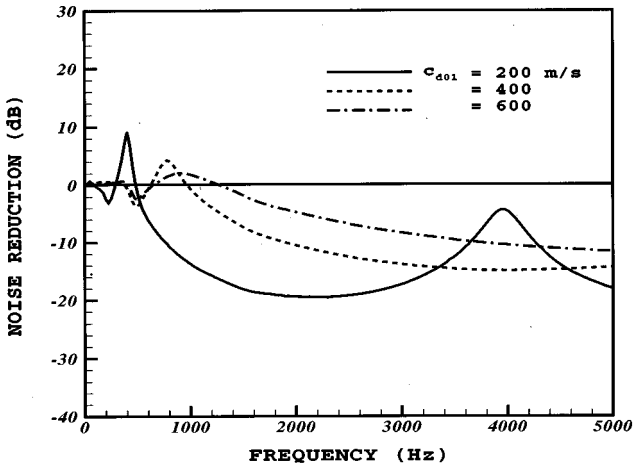


FIG. 13. Effect of the dilatational wave speed in the baffle on the noise reduction.

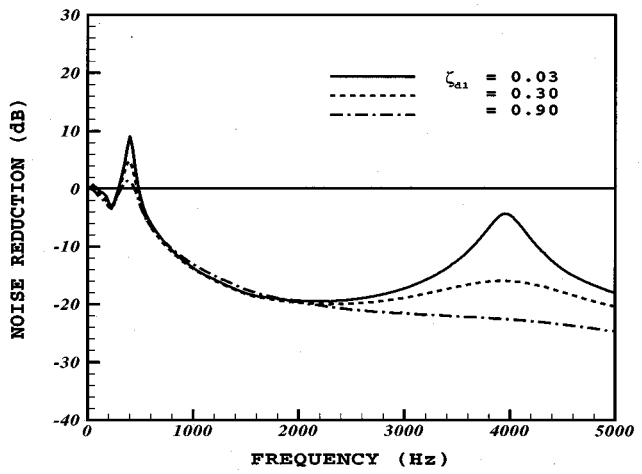


FIG. 14. Effect of the loss factor associated with the dilatational wave on the noise reduction.

tional wave speed. However, in this study, the variation of the dilatational wave speed due to the density change is not taken into account for noise reduction calculations. Figure 14 presents the effect of the loss factor (ζ_{d1}) associated with the dilatational wave in the baffle layer on the noise reduction. As can be seen in Fig. 14, the noise reduction levels below 2 kHz for various loss factors are approximately the same. However, beyond 2 kHz, the amount of the noise reduction increases (becomes larger negatively) as the loss factor increases. Note that the noise reduction is significantly affected by the loss factor near the frequency where the minimum noise reduction is obtained otherwise. Numerical calculations show that the shear wave speed (c_{s01}) and the loss factor (ζ_{s1}) have no significant effects on the noise reductions.

V. CONCLUDING REMARKS

A theoretical model has been developed for reducing the structure-borne noise generated by the cylindrical shell excited by an axially symmetric ring force. The theoretical model used here is a two-layer composite shell structure that is a vibrating cylindrical shell coated with a microvoided elastomer (baffle layer). Based on the limited study conducted in this work, the following conclusions are drawn:

1. The amount of noise reduction increases as (a) the density of the baffle layer decreases, (b) the dilatational wave speed in the baffle decreases, and (c) the loss factor associated with the dilatational wave increases;
2. The amount of noise reduction increases (depending on the frequency) as the baffle layer thickness increases;
3. The effects of the shear wave speed in the baffle layer and the associated loss factor on the noise reduction are insignificant.

The implication of item (1) is that a light material is more suitable for the reduction of the structure-borne noise. It is found that the analysis of the cylindrical shell for a large value of *ka* (*k* and *a* being a wave number and a shell radius, respectively) can be approximated by the analysis of the flat plate for the high frequency region.

ACKNOWLEDGMENTS

This work was supported by the Korea Science and Engineering Foundation. The authors are grateful to the reviewers of this paper for their valuable comments.

APPENDIX A: SYSTEM OF LINEAR ALGEBRAIC EQUATIONS (AXIALLY SYMMETRIC LOADING FOR A TWO-LAYER MODEL)

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & 0 & 0 & & & & & & & \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & 0 & 0 & & & & & & & \\ & 0 & 0 & a_{34} & a_{35} & a_{36} & a_{37} & & & & & & & \\ & a_{42} & a_{43} & a_{44} & a_{45} & 0 & 0 & & & & & & & \\ & a_{52} & a_{53} & a_{54} & a_{55} & 0 & 0 & a_{58} & a_{59} & a_{5,10} & a_{5,11} & 0 & 0 & \\ & 0 & 0 & a_{64} & a_{65} & a_{66} & a_{67} & 0 & 0 & a_{6,10} & a_{6,11} & a_{6,12} & a_{6,13} & \\ & a_{72} & a_{73} & a_{74} & a_{75} & 0 & 0 & a_{78} & a_{79} & a_{7,10} & a_{7,11} & 0 & 0 & \\ & a_{82} & a_{83} & a_{84} & a_{85} & 0 & 0 & a_{88} & a_{89} & a_{8,10} & a_{8,11} & 0 & 0 & \\ & 0 & 0 & a_{94} & a_{95} & a_{96} & a_{97} & 0 & 0 & a_{9,10} & a_{9,11} & a_{9,12} & a_{9,13} & \\ & a_{10,2} & a_{10,3} & a_{10,4} & a_{10,5} & 0 & 0 & a_{10,8} & a_{10,9} & a_{10,10} & a_{10,11} & 0 & 0 & \\ & & & & & & & 0 & 0 & a_{11,10} & a_{11,11} & a_{11,12} & a_{11,13} & \\ & & & & & & & a_{12,8} & a_{12,9} & a_{12,10} & a_{12,11} & 0 & 0 & \\ & & & & & & & a_{13,8} & a_{13,9} & a_{13,10} & a_{13,11} & 0 & 0 & a_{13,14} \\ & & & & & & & a_{14,8} & a_{14,9} & a_{14,10} & a_{14,11} & 0 & 0 & a_{14,14} \end{bmatrix} \begin{bmatrix} A_0^{(0)} \\ A_{1,1}^{(0)} \\ B_{1,1}^{(0)} \\ A_{2,1}^{(0)} \\ B_{2,1}^{(0)} \\ A_{3,1}^{(0)} \\ B_{3,1}^{(0)} \\ A_{1,2}^{(0)} \\ B_{1,2}^{(0)} \\ A_{2,2}^{(0)} \\ B_{2,2}^{(0)} \\ A_{3,2}^{(0)} \\ B_{3,2}^{(0)} \\ B_0^{(0)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ b_{14} \end{bmatrix}.$$

APPENDIX B: COEFFICIENTS SHOWN IN APPENDIX A

$$a_{11}=\begin{cases} H_0^{(2)}(\sqrt{k_0^2-k_z^2}R_0) & \text{for } k_0>k_z \\ H_0^{(2)}(-i\sqrt{k_z^2-k_0^2}R_0) & \text{for } k_0<k_z \end{cases},$$

$$a_{12}=\rho_1 c_{d1}^2\alpha_1^2 J_0''(\alpha_1R_0)+\rho_1(c_{d1}^2-2c_{s1}^2)\times\left\{-k_z^2J_0(\alpha_1R_0)+\frac{\alpha_1}{R_0}J_0'(\alpha_1R_0)\right\},$$

$$a_{13}=\rho_1 c_{d1}^2\alpha_1^2 Y_0''(\alpha_1R_0)+\rho_1(c_{d1}^2-2c_{s1}^2)\times\left\{-k_z^2Y_0(\alpha_1R_0)+\frac{\alpha_1}{R_0}Y_0'(\alpha_1R_0)\right\},$$

$$a_{14}=\rho_1 c_{d1}^2\{-ik_z\beta_1J_1'(\beta_1R_0)\}+\rho_1(c_{d1}^2-2c_{s1}^2)\{+ik_z\beta_1J_1'(\beta_1R_0)\},$$

$$a_{15}=\rho_1 c_{d1}^2\{-ik_z\beta_1Y_1'(\beta_1R_0)\}+\rho_1(c_{d1}^2-2c_{s1}^2)\{+ik_z\beta_1Y_1'(\beta_1R_0)\},$$

$$a_{21}=\begin{cases} -\sqrt{k_0^2-k_z^2}H_0^{(2)\prime}(\sqrt{k_0^2-k_z^2}R_0) & \text{for } k_0>k_z \\ +i\sqrt{k_z^2-k_0^2}H_0^{(2)\prime}(-i\sqrt{k_z^2-k_0^2}R_0) & \text{for } k_0<k_z \end{cases},$$

$$a_{22}=\rho_0\omega^2\alpha_1J_0'(\alpha_1R_0),\quad a_{23}=\rho_0\omega^2\alpha_1Y_0'(\alpha_1R_0),$$

$$a_{24}=-i\rho_0\omega^2k_zJ_1(\beta_1R_0),$$

$$a_{25}=-i\rho_0\omega^2k_zY_1(\beta_1R_0),$$

$$a_{34}=\rho_1 c_{s1}^2\left\{\frac{i}{R_0}k_zJ_1(\beta_1R_0)-ik_z\beta_1J_1'(\beta_1R_0)\right\},$$

$$a_{35}=\rho_1 c_{s1}^2\left\{\frac{i}{R_0}k_zY_1(\beta_1R_0)-ik_z\beta_1Y_1'(\beta_1R_0)\right\},$$

$$a_{36}=\rho_1 c_{s1}^2\left\{\frac{\beta_1}{R_0}J_0'(\beta_1R_0)-\beta_1^2J_0''(\beta_1R_0)\right\},$$

$$a_{37}=\rho_1 c_{s1}^2\left\{\frac{\beta_1}{R_0}Y_0'(\beta_1R_0)-\beta_1^2Y_0''(\beta_1R_0)\right\},$$

$$a_{42}=\rho_1 c_{s1}^2\{-2ik_z\alpha_1J_0'(R_0)\},$$

$$a_{43}=\rho_1 c_{s1}^2\{-2ik_z\alpha_1Y_0'(a_1R_0)\},$$

$$a_{44}=\rho_1 c_{s1}^2\left\{\left(\frac{1}{R_0^2}-k_z^2\right)J_1(\beta_1R_0)-\frac{1}{R_0}\beta_1J_1'(\beta_1R_0)-\beta_1^2J_1''(\beta_1R_0)\right\},$$

$$a_{45}=\rho_1 c_{s1}^2\left\{\left(\frac{1}{R_0^2}-k_z^2\right)Y_1(\beta_1R_0)-\frac{1}{R_0}\beta_1Y_1'(\beta_1R_0)-\beta_1^2Y_1''(\beta_1R_0)\right\},$$

$$a_{52}=\rho_1 c_{d1}^2\alpha_1^2 J_0''(\alpha_1R_1)+\rho_1(c_{d1}^2-2c_{s1}^2)\times\left\{-k_z^2J_0(\alpha_1R_1)+\frac{\alpha_1}{R_1}J_0'(\alpha_1R_1)\right\},$$

$$a_{53}=\rho_1 c_{d1}^2\alpha_1^2 Y_0''(\alpha_1R_1)+\rho_1(c_{d1}^2-2c_{s1}^2)\times\left\{-k_z^2Y_0(\alpha_1R_1)+\frac{\alpha_1}{R_1}Y_0'(\alpha_1R_1)\right\},$$

$$a_{54}=\rho_1 c_{d1}^2\{-ik_z\beta_1J_1'(\beta_1R_1)\}+\rho_1(c_{d1}^2-2c_{s1}^2)\times\{ik_z\beta_1J_1'(\beta_1R_1)\},$$

$$a_{55}=\rho_1 c_{d1}^2\{-ik_z\beta_1Y_1'(\beta_1R_1)\}+\rho_1(c_{d1}^2-2c_{s1}^2)\times\{ik_z\beta_1Y_1'(\beta_1R_1)\},$$

$$a_{58}=-\left\{\rho_2 c_{d2}^2\alpha_2^2 J_0''(\alpha_2R_1)+\rho_2(c_{d2}^2-2c_{s2}^2)\times\left[-k_z^2J_0(\alpha_2R_1)+\frac{\alpha_2}{R_1}J_0'(\alpha_2R_1)\right]\right\},$$

$$a_{59}=-\left\{\rho_2 c_{d2}^2\alpha_2^2 Y_0''(\alpha_2R_1)+\rho_2(c_{d2}^2-2c_{s2}^2)\times\left[-k_z^2Y_0(\alpha_2R_1)+\frac{\alpha_2}{R_1}Y_0'(\alpha_2R_1)\right]\right\},$$

$$a_{5,10}=-\{\rho_2 c_{d2}^2[-ik_z\beta_2J_1'(\beta_2R_1)]+\rho_2(c_{d2}^2-2c_{s2}^2)[ik_z\beta_2J_1'(\beta_2R_1)]\},$$

$$a_{5,11}=-\{\rho_2 c_{d2}^2[-ik_z\beta_2Y_1'(\beta_2R_1)]+\rho_2(c_{d2}^2-2c_{s2}^2)[ik_z\beta_2Y_1'(\beta_2R_1)]\},$$

$$a_{64}=\rho_1 c_{s1}^2\left\{\frac{ik_z}{R_1}J_1(\beta_1R_1)-ik_z\beta_1J_1'(\beta_1R_1)\right\},$$

$$a_{65}=\rho_1 c_{s1}^2\left\{\frac{ik_z}{R_1}Y_1(\beta_1R_1)-ik_z\beta_1Y_1'(\beta_1R_1)\right\},$$

$$a_{66}=\rho_1 c_{s1}^2\left\{\frac{\beta_1}{R_1}J_0'(\beta_1R_1)-\beta_1^2J_0''(\beta_1R_1)\right\},$$

$$a_{67}=\rho_1 c_{s1}^2\left\{\frac{\beta_1}{R_1}Y_0'(\beta_1R_1)-\beta_1^2Y_0''(\beta_1R_1)\right\},$$

$$a_{6,10}=-\left\{\rho_2 c_{s2}^2\left[\frac{ik_z}{R_1}J_1(\beta_2R_1)-ik_z\beta_2J_1'(\beta_2R_1)\right]\right\},$$

$$a_{6,11}=-\left\{\rho_2 c_{s2}^2\left[\frac{ik_z}{R_1}Y_1(\beta_2R_1)-ik_z\beta_2Y_1'(\beta_2R_1)\right]\right\},$$

$$a_{6,12}=-\left\{\rho_2 c_{s2}^2\left[\frac{\beta_2}{R_1}J_0'(\beta_2R_1)-\beta_2^2J_0''(\beta_2R_1)\right]\right\},$$

$$a_{6,13}=-\left\{\rho_2 c_{s2}^2\left[\frac{\beta_2}{R_1}Y_0'(\beta_2R_1)-\beta_2^2Y_0''(\beta_2R_1)\right]\right\},$$

$$a_{72}=\rho_1 c_{s1}^2\{-2ik_z\alpha_1J_0'(\alpha_1R_1)\},$$

$$a_{73}=\rho_1 c_{s1}^2\{-2ik_z\alpha_1Y_0'(\alpha_1R_1)\},$$

$$a_{74}=\rho_1 c_{s1}^2\left\{\left(\frac{1}{R_1^2}-k_z^2\right)J_1(\beta_1 R_1)-\frac{1}{R_1}\beta_1 J_1'(\beta_1 R_1)-\beta_1^2 J_1''(\beta_1 R_1)\right\},$$

$$a_{75}=\rho_1 c_{s1}^2\left\{\left(\frac{1}{R_1^2}-k_z^2\right)Y_1(\beta_1 R_1)-\frac{1}{R_1}\beta_1 Y_1'(\beta_1 R_1)-\beta_1^2 Y_1''(\beta_1 R_1)\right\},$$

$$a_{78}=-\{\rho_2 c_{s2}^2[-2ik_z\alpha_2 J_0'(\alpha_2 R_1)]\},$$

$$a_{79}=-\{\rho_2 c_{s2}^2[-2ik_z\alpha_2 Y_0'(a_2 R_1)]\},$$

$$a_{7,10}=-\left\{\rho_2 c_{s2}^2\left[\left(\frac{1}{R_1^2}-k_z^2\right)J_1(\beta_2 R_1)-\frac{1}{R_1}\beta_2 J_1'(\beta_2 R_1)-\beta_2^2 J_1''(\beta_2 R_1)\right]\right\},$$

$$a_{7,11}=-\left\{\rho_2 c_{s2}^2\left[\left(\frac{1}{R_1^2}-k_z^2\right)Y_1(\beta_2 R_1)-\frac{1}{R_1}\beta_2 Y_1'(\beta_2 R_1)-\beta_2^2 Y_1''(\beta_2 R_1)\right]\right\},$$

$$a_{82}=\alpha_1 J_0'(\alpha_1 R_1),\quad a_{83}=\alpha_1 Y_0'(\alpha_1 R_1),$$

$$a_{84}=-ik_z J_1(\beta_1 R_1),\quad a_{85}=-ik_z Y_1(\beta_1 R_1),$$

$$a_{88}=-\{\alpha_2 J_0'(\alpha_2 R_1)\},\quad a_{89}=-\{\alpha_2 Y_0'(\alpha_2 R_1)\},$$

$$a_{8,10}=-\{-ik_z J_1(\beta_2 R_1)\},\quad a_{8,11}=-\{-ik_z Y_1(\beta_2 R_1)\},$$

$$a_{94}=-ik_z J_1(\beta_1 R_1),\quad a_{95}=-ik_z Y_1(\beta_1 R_1),$$

$$a_{96}=-\beta_1 J_0'(\beta_1 R_1),\quad a_{97}=-\beta_1 Y_0'(\beta_1 R_1),$$

$$a_{9,10}=-\{-ik_z J_1(\beta_2 R_1)\},\quad a_{9,11}=-\{-ik_z Y_1(\beta_2 R_1)\},$$

$$a_{9,12}=-\{-\beta_2 J_0'(\beta_2 R_1)\},\quad a_{9,13}=-\{-\beta_2 Y_0'(\beta_2 R_1)\},$$

$$a_{10,2}=-ik_z J_0(\alpha_1 R_1),\quad a_{10,3}=-ik_z Y_0(\alpha_1 R_1),$$

$$a_{10,4}=-\left\{\frac{1}{R_1}J_1(\beta_1 R_1)+\beta_1 J_1'(\beta_1 R_1)\right\},$$

$$a_{10,5}=-\left\{\frac{1}{R_1}Y_1(\beta_1 R_1)+\beta_1 Y_1'(\beta_1 R_1)\right\},$$

$$a_{10,8}=-\{-ik_z J_0(\alpha_2 R_1)\},$$

$$a_{10,9}=-\{-ik_z Y_0(\alpha_2 R_1)\},$$

$$a_{10,10}=-\left\{-\left[\frac{1}{R_1}J_1(\beta_2 R_1)+\beta_2 J_1'(\beta_2 R_1)\right]\right\},$$

$$a_{10,11}=-\left\{-\left[\frac{1}{R_1}Y_1(\beta_2 R_1)+\beta_2 Y_1'(\beta_2 R_1)\right]\right\},$$

$$a_{11,10}=\rho_2 c_{s2}^2\left\{\frac{i}{R_2}k_z J_1(\beta_2 R_2)-ik_z\beta_2 J_1'(\beta_2 R_2)\right\},$$

$$a_{11,11}=\rho_2 c_{s2}^2\left\{\frac{i}{R_2}k_z Y(\beta_2 R_2)-ik_z\beta_2 Y_1'(\beta_2 R_2)\right\},$$

$$a_{11,12}=\rho_2 c_{s2}^2\left\{\frac{\beta_2}{R_2}J_0'(\beta_2 R_2)-\beta_2^2 J_0''(\beta_2 R_2)\right\},$$

$$a_{11,13}=\rho_2 c_{s2}^2\left\{\frac{\beta_2}{R_2}Y_0'(\beta_2 R_2)-\beta_2^2 Y_0''(\beta_2 R_2)\right\},$$

$$a_{12,8}=\rho_2 c_{s2}^2\{-2ik_z\alpha_2 J_0'(\alpha_2 R_2)\},$$

$$a_{12,9}=\rho_2 c_{s2}^2\{-2ik_z\alpha_2 Y_0'(\alpha_2 R_2)\},$$

$$a_{12,10}=\rho_2 c_{s2}^2\left\{\left(\frac{1}{R_2^2}-k_z^2\right)J_1(\beta_2 R_2)-\frac{1}{R_2}\beta_2 J_1'(\beta_2 R_2)-\beta_2^2 J_1''(\beta_2 R_2)\right\},$$

$$a_{12,11}=\rho_2 c_{s2}^2\left\{\left(\frac{1}{R_2^2}-k_z^2\right)Y_1(\beta_2 R_2)-\frac{1}{R_2}\beta_2 Y_1'(\beta_2 R_2)-\beta_2^2 Y_1''(\beta_2 R_2)\right\},$$

$$a_{13,8}=\rho_3\omega^2\alpha_2 J_0'(\alpha_2 R_2),$$

$$a_{13,9}=\rho_3\omega^2\alpha_2 Y_0'(\alpha_2 R_2),$$

$$a_{13,10}=-i\rho_3\omega^2 k_z J_1(\beta_2 R_2),$$

$$a_{13,11}=-i\rho_3\omega^2 k_z Y_1(\beta_2 R_2),$$

$$a_{13,14}=\begin{cases}-\sqrt{k_3^2-k_z^2}\,J_0'(\sqrt{k_3^2-k_z^2}R_2) & \text{for } k_3>k_z\\ -\sqrt{k_z^2-k_3^2}\,I_0'(\sqrt{k_z^2-k_3^2}R_2) & \text{for } k_3<k_z\end{cases},$$

$$a_{14,8}=\rho_2 c_{d2}^2\alpha_2^2 J_0''(\alpha_2 R_2)+\rho_2(c_{d2}^2-2c_{s2}^2)\times\left[-k_z^2 J_0(\alpha_2 R_2)+\frac{\alpha_2}{R_2}J_0'(\alpha_2 R_2)\right],$$

$$a_{14,9}=\rho_2 c_{d2}^2\alpha_2^2 Y_0''(\alpha_2 R_2)+\rho_2(c_{d2}^2-2c_{s2}^2)\times\left[-k_z^2 Y_0(\alpha_2 R_2)+\frac{\alpha_2}{R_2}Y_0'(\alpha_2 R_2)\right],$$

$$a_{14,10}=\rho_2 c_{d2}^2[-ik_z\beta_2 J_1'(\beta_2 R_2)]+\rho_2(c_{d2}^2-2c_{s2}^2)\times[ik_z\beta_2 J_1'(\beta_2 R_2)],$$

$$a_{14,11}=\rho_2 c_{d2}^2[-ik_z\beta_2 Y_1'(\beta_2 R_2)]+\rho_2(c_{d2}^2-2c_{s2}^2)\times[ik_z\beta_2 Y_1'(\beta_2 R_2)],$$

$$a_{14,14}=\begin{cases}J_0(\sqrt{k_3^2-k_z^2}R_2) & \text{for } k_3>k_z\\ I_0(\sqrt{k_z^2-k_3^2}R_2) & \text{for } k_3<k_z\end{cases},$$

$$b_{14}=-F_0.$$

APPENDIX C: SYSTEM OF LINEAR ALGEBRAIC EQUATIONS (AXIALLY SYMMETRIC LOADING FOR A ONE-LAYER MODEL)

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & 0 & 0 & \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & 0 & 0 & \\ & 0 & 0 & a_{34} & a_{35} & a_{36} & a_{37} & \\ & a_{42} & a_{43} & a_{44} & a_{45} & 0 & 0 & \\ & 0 & 0 & a_{54} & a_{55} & a_{56} & a_{57} & \\ & a_{62} & a_{63} & a_{64} & a_{65} & 0 & 0 & \\ & a_{72} & a_{73} & a_{74} & a_{75} & 0 & 0 & a_{78} \\ & a_{82} & a_{83} & a_{84} & a_{85} & 0 & 0 & a_{88} \end{bmatrix} \begin{bmatrix} A_0^{(0)} \\ A_{1,2}^{(0)} \\ B_{1,2}^{(0)} \\ A_{2,2}^{(0)} \\ B_{2,2}^{(0)} \\ A_{3,2}^{(0)} \\ B_{3,2}^{(0)} \\ B_0^{(0)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ b_8 \end{bmatrix}.$$

APPENDIX D: COEFFICIENTS SHOWN IN APPENDIX C

$$a_{11}=\begin{cases} H_0^{(2)}(\sqrt{k_0^2-k_z^2}R_1) & \text{for } k_0>k_z \\ H_0^{(2)}(-i\sqrt{k_z^2-k_0^2}R_1) & \text{for } k_0<k_z \end{cases},$$

$$a_{12}=\rho_2 c_{d2}^2\alpha_2^2 J_0''(\alpha_2R_1)+\rho_2(c_{d2}^2-2c_{s2}^2)\times\left\{-k_z^2J_0(\alpha_2R_1)+\frac{\alpha_2}{R_1}J_0'(\alpha_2R_1)\right\},$$

$$a_{13}=\rho_2 c_{d2}^2\alpha_2^2 Y_0''(\alpha_2R_1)+\rho_2(c_{d2}^2-2c_{s2}^2)\times\left\{-k_z^2Y_0(\alpha_2R_1)+\frac{\alpha_2}{R_1}Y_0'(\alpha_2R_1)\right\},$$

$$a_{14}=\rho_2 c_{d2}^2\{-ik_z\beta_2J_1'(\beta_2R_1)\}+\rho_2(c_{d2}^2-2c_{s2}^2)\times\{+ik_z\beta_2J_1'(\beta_2R_1)\},$$

$$a_{15}=\rho_2 c_{d2}^2\{-ik_z\beta_2Y_1'(\beta_2R_1)\}+\rho_2(c_{d2}^2-2c_{s2}^2)\times\{+ik_z\beta_2Y_1'(\beta_2R_1)\},$$

$$a_{21}=\begin{cases} -\sqrt{k_0^2-k_z^2}H_0^{(2)\prime}(\sqrt{k_0^2-k_z^2}R_1) & \text{for } k_0>k_z \\ +i\sqrt{k_z^2-k_0^2}H_0^{(2)\prime}(\sqrt{k_z^2-k_0^2}R_1) & \text{for } k_0<k_z \end{cases},$$

$$a_{22}=\rho_0\omega^2\alpha_2J_0'(\alpha_2R_1),\quad a_{23}=\rho_0\omega^2\alpha_2Y_0'(\alpha_2R_1),$$

$$a_{24}=-i\rho_0\omega^2k_zJ_1(\beta_2R_1),\quad a_{25}=-i\rho_0\omega^2k_zY_1(\beta_2R_1),$$

$$a_{34}=\rho_2c_{s2}^2\left\{\frac{i}{R_1}k_zJ_1(\beta_2R_1)-ik_z\beta_2J_1'(\beta_2R_1)\right\},$$

$$a_{35}=\rho_2c_{s2}^2\left\{\frac{i}{R_1}k_zY_1(\beta_2R_1)-ik_z\beta_2Y_1'(\beta_2R_1)\right\},$$

$$a_{36}=\rho_2c_{s2}^2\left\{\frac{\beta_2}{R_1}J_0'(\beta_2R_1)-\beta_2^2J_0''(\beta_2R_1)\right\},$$

$$a_{37}=\rho_2c_{s2}^2\left\{\frac{\beta_2}{R_1}Y_0'(\beta_2R_1)-\beta_2^2Y_0''(\beta_2R_1)\right\},$$

$$a_{42}=\rho_2c_{s2}^2\{-2ik_z\alpha_2J_0'(\alpha_2R_1)\},$$

$$a_{43}=\rho_2c_{s2}^2\{-2ik_z\alpha_2Y_0'(\alpha_2R_1)\},$$

$$a_{44}=\rho_2c_{s2}^2\left\{\left(\frac{1}{R_1^2}-k_z^2\right)J_1(\beta_2R_1)-\frac{1}{R_1}\beta_2J_1'(\beta_2R_1)-\beta_2^2J_1''(\beta_2R_1)\right\},$$

$$a_{45}=\rho_2c_{s2}^2\left\{\left(\frac{1}{R_1^2}-k_z^2\right)Y_1(\beta_2R_1)-\frac{1}{R_1}\beta_2Y_1'(\beta_2R_1)-\beta_2^2Y_1''(\beta_2R_1)\right\},$$

$$a_{54}=\rho_2c_{s2}^2\left\{\frac{i}{R_2}k_zJ_1(\beta_2R_2)-ik_z\beta_2J_1'(\beta_2R_2)\right\},$$

$$a_{55}=\rho_2c_{s2}^2\left\{\frac{i}{R_2}k_zY_1(\beta_2R_2)-ik_z\beta_2Y_1'(\beta_2R_2)\right\},$$

$$a_{56}=\rho_2c_{s2}^2\left\{\frac{\beta_2}{R_2}J_0'(\beta_2R_2)-\beta_2^2J_0''(\beta_2R_2)\right\},$$

$$a_{57}=\rho_2c_{s2}^2\left\{\frac{\beta_2}{R_2}Y_0'(\beta_2R_2)-\beta_2^2Y_0''(\beta_2R_2)\right\},$$

$$a_{62}=\rho_2c_{s2}^2\{-2ik_z\alpha_2J_0'(\alpha_2R_2)\},$$

$$a_{63}=\rho_2c_{s2}^2\{-2ik_z\alpha_2Y_0'(\alpha_2R_2)\},$$

$$a_{64}=\rho_2c_{s2}^2\left\{\left(\frac{1}{R_2^2}-k_z^2\right)J_1(\beta_2R_2)-\frac{1}{R_2}\beta_2J_1'(\beta_2R_2)-\beta_2^2J_1''(\beta_2R_2)\right\},$$

$$a_{65}=\rho_2c_{s2}^2\left\{\left(\frac{1}{R_2^2}-k_z^2\right)Y_1(\beta_2R_2)-\frac{1}{R_2}\beta_2Y_1'(\beta_2R_2)-\beta_2^2Y_1''(\beta_2R_2)\right\},$$

$$a_{72}=\rho_3\omega^2\alpha_2J_0'(\alpha_2R_2),\quad a_{73}=\rho_3\omega^2\alpha_2Y_0'(\alpha_2R_2),$$

$$a_{74}=-i\rho_3\omega^2k_zJ_1(\beta_2R_2),\quad a_{75}=-i\rho_3\omega^2k_zY_1(\beta_2R_2),$$

$$a_{78}=\begin{cases} -\sqrt{k_3^2-k_z^2}J_0'(\sqrt{k_3^2-k_z^2}R_2) & \text{for } k_3>k_z \\ -\sqrt{k_z^2-k_3^2}I_0'(\sqrt{k_z^2-k_3^2}R_2) & \text{for } k_3<k_z \end{cases},$$

$$a_{82}=\rho_2 c_{d2}^2 \alpha_2^2 J_0''(\alpha_2 R_2)+\rho_2(c_{d2}^2-2c_{s2}^2) \times\left[-k_z^2 J_0(\alpha_2 R_2)+\frac{\alpha_2}{R_2} J_0'(\alpha_2 R_2)\right],$$

$$a_{83}=\rho_2 c_{d2}^2 \alpha_2^2 Y_0''(\alpha_2 R_2)+\rho_2(c_{d2}^2-2c_{s2}^2) \times\left[-k_z^2 Y_0(\alpha_2 R_2)+\frac{\alpha_2}{R_2} Y_0'(\alpha_2 R_2)\right],$$

$$a_{84}=\rho_2 c_{d2}^2[-ik_z\beta_2 J_1'(\beta_2 R_2)]+\rho_2(c_{d2}^2-2c_{s2}^2) \times[ik_z\beta_2 J_1'(\beta_2 R_2)],$$

$$a_{85}=\rho_2 c_{d2}^2[-ik_z\beta_2 Y_1'(\beta_2 R_2)]+\rho_2(c_{d2}^2-2c_{s2}^2) \times[ik_z\beta_2 Y_1'(\beta_2 R_2)],$$

$$a_{88}=\begin{cases} J_0(\sqrt{k_3^2-k_z^2}R_2) & \text{for } k_3>k_z \\ I_0(\sqrt{k_z^2-k_3^2}R_2) & \text{for } k_3<k_z \end{cases},$$

$$b_8=-F_0.$$

[1] S. H. Ko, C. H. Sherman, and W. A. Strawderman, ''Baffling of flexural waves: noise reduction,'' Tech Memo, TD 12-80-74, Naval Underwater Systems Center, New London, Connecticut 06320, October, 1974.

[2] S. H. Ko and C. H. Sherman, ''Effectiveness of flexural wave baffles for improving performance of hull-mounted arrays,'' Tech Memo No. 771062, Naval Underwater Systems Center, New London, Connecticut 06320, April, 1977.

[3] S. H. Ko and C. H. Sherman, ''Flexural wave baffling,'' J. Acoust. Soc. Am. **66**, 566–570 (1979).

[4] G. Maidanik, ''Acoustic radiation from a driven coated infinite plate backed by a parallel infinite baffle,'' J. Acoust. Soc. Am. **42**, 32–35 (1967).

[5] S. H. Ko, ''Reduction of structure-borne noise using an air-voided elastomer,'' J. Acoust. Soc. Am. **101**, 3306–3312 (1997).

[6] H. Kolsky, *Stress Wave in Solids* (Dover, New York, 1963).

[7] T. R. Meeker and A. H. Meizler, ''Guided wave propagation in elongated cylinders and plates,'' *Physical Acoustics*, edited by W. P. Mason (Academic, New York, 1964).

[8] K. F. Graff, *Wave Motion in Elastic Solids* (Dover, New York, 1975).

[9] D. C. Gazis, ''Three-dimensional investigation of the propagation of waves in hollow circular cylinders. I. Analytical foundation,'' J. Acoust. Soc. Am. **31**, 568–573 (1957).

[10] S. H. Ko and B. E. Sandman, ''Three-dimensional analysis of acoustic scattering from a coated cylindrical shell,'' NUWC-NPT Technical Report No. 11,042, Naval Undersea Warfare Center Division, Newport, Rhode Island 02841, September, 1998.

Suppression of distortion product otoacoustic emissions and hearing threshold

Martin Pienkowski[a] and Hans Kunov
Institute of Biomaterials and Biomedical Engineering, University of Toronto, Rosebrugh Building, 4 Taddle Creek Road, Toronto, Ontario M5S 3G9, Canada

(Received 13 June 2000; accepted for publication 16 January 2001)

A distortion product otoacoustic emission (DPOAE) suppression tuning curve (STC) shows the minimum level of suppressor tone that is required to reduce DPOAE level by a fixed amount, as a function of suppressor frequency. Several years ago, Mills [J. Acoust. Soc. Am. **103**, 507–523 (1998)] derived, theoretically, an approximately linear relationship between the tip-to-tail suppressor level difference on a DPOAE STC, and the gain of the cochlear amplifier, defined as the maximum increase in the active over the passive basilar membrane (BM) response. In this paper, preliminary data from adult human subjects are presented that establish a correlation between this tip-to-tail DPOAE STC difference and the threshold of hearing, the latter measured at the frequency of the f_2 primary tone. Assuming that both suppression and the DPOAE are by-products of active, nonlinear BM dynamics, the above result suggests that threshold elevation in mild levels of hearing loss may be attributed, in part, to a reduction of cochlear amplifier gain, which is detectable with the suppression paradigm. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1354202]

PACS numbers: 43.64.Bt, 43.64.Jb, 43.64.Kc, 43.66.Sr, 43.66.Cb [BLM]

I. INTRODUCTION

The contemporary theory of hearing maintains that the mammalian cochlea is an ''active'' sensory transducer. Gold, its first proponent, asserted that an internal, metabolically powered motor counters cochlear fluid viscosity to boost the sound-induced response of the resonant basilar membrane (BM), and thus sharpen its frequency resolution (Gold and Pumphrey, 1948; Gold, 1948). Over the past two decades, much experimental and theoretical work has been presented in support of this hypothesis (reviewed by Dallos, 1988). The motor, or so-called ''cochlear amplifier'' (CA), is believed to reside in the outer hair cells (OHCs; Brownell, 1983), but the mechanisms of its operation remain largely unknown (reviewed by Yates, 1995; Holley, 1996), and there is some controversy about its capacity to impact BM mechanics (e.g., Allen and Fahey, 1992).

Nonetheless, in the dead animal, or when outer hair cells are selectively destroyed or their transduction process impaired in the live cochlea, the peak of the basilar membrane response to a pure tone is lower than in the normal animal by as much as 40 to 50 dB (e.g., Sellick *et al.*, 1982; Ruggero and Rich, 1991; reviewed by Ulfendahl, 1997). This difference is commonly attributed to the cochlear amplifier and adopted as a working definition of its ''gain'' (reviewed by Patuzzi, 1996; deBoer, 1996). As illustrated schematically in Fig. 1, the drop in basilar membrane response is confined to an ''effective width,'' where presumably, CA feedback is significant in relation to the BM impedance (Patuzzi, 1996; deBoer, 1996).

Cochlear trauma, or loss of OHC function, also linearizes the basilar membrane response (e.g., Sellick et al., 1982; Ruggero and Rich, 1991; reviewed by Ulfendahl, 1997), and eliminates certain mechanical phenomena, which in consequence, are thought to be by-products of nonlinear cochlear amplifier activity. One of these is two-tone suppression, or a decrease in the BM response to one, test tone, by the simultaneous sounding of a second, suppressor tone (Rhode, 1977). It appears that suppression in BM dynamics can be accounted for by the saturation of OHC feedback (Zwicker, 1986; Geisler *et al.*, 1990; Ruggero *et al.*, 1992a, 1992b; Cooper, 1996; Cai and Geisler, 1996; Geisler and Nuttall, 1997).

A second phenomenon generally attributed to cochlear amplifier nonlinearity are otoacoustic emissions (OAEs; Kemp, 1978, 1979; Zurek, 1981; Bialek and Wit, 1984; Robles *et al.*, 1989; reviewed by Probst *et al.*, 1991). OAEs are nonstimulus sounds detected with a microphone sealed inside the ear canal. For example, a distortion product otoacoustic emission (DPOAE) can be observed at frequency $2f_1-f_2$ (the cubic difference tone) after the cochlea is stimulated with a pair of primary tones, f_1 and f_2 ($f_2>f_1$), that are close in frequency. The level of the DPOAE is decreased, or rendered undetectable, by ototoxic stimuli or other experimental manipulations that selectively impair OHC function (reviewed by Probst and Harris, 1997). It is well-known that the level of the $2f_1-f_2$ DPOAE correlates (albeit weakly) with hearing threshold in humans (e.g., Kimberley *et al.*, 1997).

DPOAE level can also be reduced by the presentation of a third, suppressor tone in conjunction with the two primaries (Brown and Kemp, 1984). It has been common practice to measure the minimum level of suppressor tone required to reduce DPOAE (most often $2f_1-f_2$) level by some criterion amount, usually a few decibels. A plot of suppressor level as a function of suppressor frequency is called a DPOAE suppression tuning curve (STC). An example of a $2f_1-f_2$

[a]Author to whom correspondence should be addressed. Electronic mail: martin.pienkowski@utoronto.ca

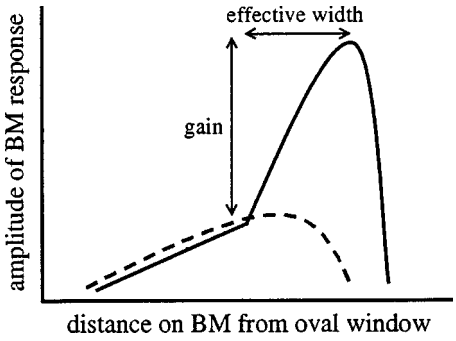


FIG. 1. Idealized basilar membrane responses to a pure tone in a normal mammalian cochlea (solid curve) and in a cochlea with dysfunctional outer hair cells (dashed curve). The difference is attributed to the cochlear amplifier (CA). CA ''gain'' is defined as the increase of peak response amplitude in the normal over the damaged cochlea. CA ''effective width'' is taken from the location on the basilar membrane (BM) where the response in the normal cochlea first exceeds that in the damaged one, to the peak of the normal response, beyond which viscous damping dominates BM dynamics.

DPOAE suppression tuning curve is given in Fig. 2. Typically, suppressor level is lowest when the suppressor frequency is just higher than that of the f_2 primary tone. The growth of level with frequency is rapid above f_2, more gradual for the first octave below f_2, and relatively small beyond the first octave. The difference in suppressor level between this tip and low-frequency tail or shoulder is on the order of a few tens of decibels. In these respects, the DPOAE STC is qualitatively similar to the pure-tone basilar membrane response function (Fig. 1 flipped upside down).

DPOAE STCs have been useful in identifying the place of resonance of the f_2 primary tone as the main source of DPOAE energy, and in exploring the contributions of other BM locations (e.g., Martin *et al.*, 1987, 1998a, 1999; Plinkert and Zenner, 1998). Their tuning sharpness, usually quantified by a Q_{10} ratio (tip frequency divided by the suppressor bandwidth at 10 dB above tip level) has been compared to that of basilar membrane, neural, and psychophysical tuning curves in various species, with mixed results (e.g., Koppl and Manley, 1993; Frank and Kossl, 1995; Abdala *et al.*, 1996; Taschenberger and Manley, 1998; Martin *et al.*, 1998b). Several papers have reported how DPOAE STCs are affected by the variation of primary tone frequencies, primary tone levels, and the criterion for amount of suppression (e.g., Brown and Kemp, 1984; Kummer *et al.*, 1995; Mills, 1998). Perhaps most telling was Mills' observation that increasing primary tone levels has the effect of raising the tip threshold and lowering the tip-to-tail difference on the DPOAE STC (Mills, 1998). The same trend is seen in pure-tone basilar membrane tuning curves with an increasing isoresponse criterion (Sellick *et al.*, 1982; Ulfendahl, 1997); at higher BM response velocities, OHC feedback plays a proportionately smaller role in the dynamics, because the viscous drag force grows linearly, while the feedback force grows only compressively. This parallelism between DPOAE STCs and BM tuning curves raises the possibility that the STC may be a noninvasive window on the gain of the cochlear amplifier. Recently, Martin and colleagues measured DPOAE STCs following the administration of diuretics such as furosemide (Martin *et al.*, 1998b), which reversibly inhibit OHC transduction (e.g., Evans and Klinke, 1982; Ruggero and Rich, 1991; Mills and Rubel, 1994). They reported increased DPOAE STC tip thresholds that correlated with furosemide-induced DPOAE level decreases.

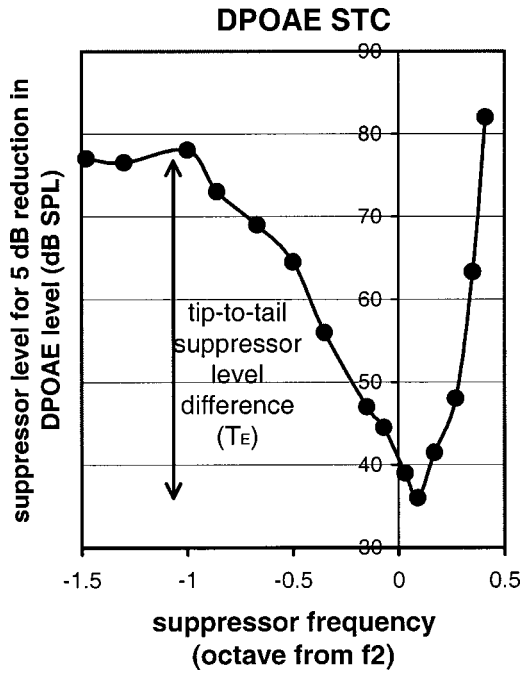


FIG. 2. Typical $2f_1-f_2$ distortion product otoacoustic emission suppression tuning curve (DPOAE STC) recorded from a human subject. Primary tone and DPOAE frequencies: $f_1=2584$, $f_2=3165$, $2f_1-f_2=2003$ Hz; primary tone levels: $L_1=60$, $L_2=40$ dB SPL; suppression criterion=5 dB. Level of the unsuppressed DPOAE was 9.8 dB SPL. The tip-to-tail suppressor level difference (abbreviated T_E, as introduced in the text below) is approximately 40 dB.

All the above-mentioned properties of DPOAE STCs can be accounted for by a model of the type developed by Mills on the basis of the amplifier-saturation suppression hypothesis (Mills, 1997, 1998; Geisler and Nuttall, 1997). Transcending the available data, Mills' model also predicts an approximately linear relationship between the tip-to-tail suppressor level difference on a DPOAE STC, and the gain of the cochlear amplifier (Mills, 1998). Simulations of our own nonlinear basilar membrane macromechanical model also predict a similar relationship (Pienkowski, 2000). However, it remains untested by direct experiment.

In this paper, we look into a possible relationship between the DPOAE STC tip-to-tail difference and the threshold of hearing in human subjects. Since a decrease in cochlear amplifier gain results in an increase in hearing threshold (e.g., Ryan and Dallos, 1975; Harrison, 1981; Liberman and Dodds, 1984; Patuzzi *et al.*, 1989), and since the DPOAE STC tip-to-tail difference is postulated to be proportional to CA gain, we hypothesize that there may be a correlation between a decrease in the DPOAE STC tip-to-tail difference and an increase in hearing threshold. Following Mills (1998), we use the abbreviation T_E (i.e., the ''threshold elevation'' in suppressor level) for the tip-to-tail difference on a DPOAE STC.

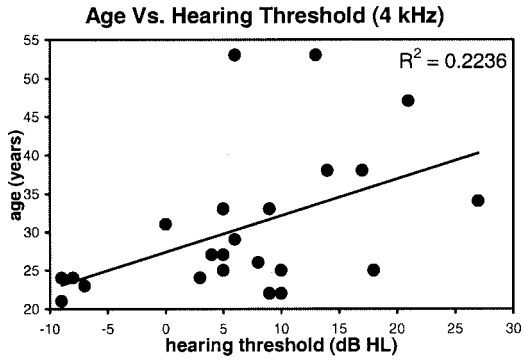


FIG. 3. Positive correlation between age and hearing threshold (4 kHz) in the 23 ears (from 18 subjects) tested in this study ($R^2=0.2236$).

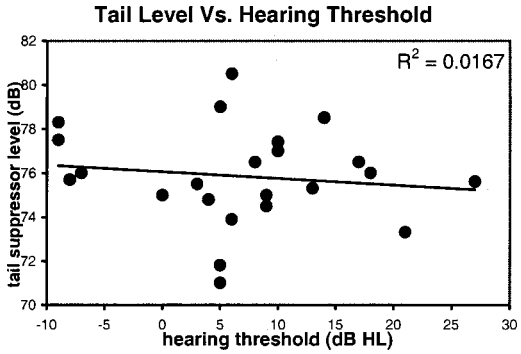


FIG. 4. No correlation between suppressor level in the tail region of the DPOAE STC and hearing threshold at 4 kHz.

II. METHODS

The subject pool consisted of 18 normally hearing individuals, aged between 21 and 53 (mean of 31), whose hearing thresholds ranged from −9 to 27 dB HL at 4 kHz, the main test frequency. Both the left and right ears were tested in five of the subjects, for a total of 23 ears. The levels and frequencies of the primary tones used to elicit DPOAEs were: $L_1=60$ dB SPL, $L_2=40$ dB SPL; $f_1=3338$ Hz, $f_2=4005$ Hz ($f_2/f_1=1.20$; $2f_1-f_2=2671$ Hz). As mentioned, low levels of primary tones, particularly L_2, yield DPOAE STCs with the largest tip-to-tail suppressor level differences, which more closely reflect the shapes of pure-tone basilar membrane threshold tuning curves. At lower levels of stimulation, the contribution of the cochlear amplifier to BM vibration (CA gain) is proportionately higher, and so, there is potential for a greater suppressive effect. This also explains why, in two-tone suppression, a given level of suppressor tone causes proportionately more suppression in low compared to high levels of test tone (e.g., Ruggero *et al.*, 1992a, 1992b). More insight into the choice of the atypically low level of L_2 primary tone can be found in Mills' paper (1998).

Full DPOAE suppression tuning curves were obtained from five of the subjects, by sweeping suppressor frequency from $1\frac{1}{2}$ octave below f_2 to $\frac{1}{2}$ octave above, and checking the suppressor level needed for a 5-dB reduction in DPOAE level. The tips of all five STCs occurred at suppressor frequencies within 0.1 octave of f_2 (four above f_2, one below), and the transitions from the steeper, low-frequency arms to the shallow tails were well defined at an average of about an octave below f_2. These observations appear consistent with previous findings, although the frequencies of DPOAE STC tails have not usually been reported (Brown and Kemp, 1984; Martin *et al.*, 1987; Kummer *et al.*, 1995; Abdala *et al.*, 1996; Mills, 1998). It was therefore decided that a single pair of suppressor frequencies (f_S) would be applied to measure the tip-to-tail difference (T_E) in all ears: $f_{S(\text{tip})}=4091$ Hz (0.03 octave above f_2) and $f_{S(\text{tail})}=2089$ Hz (0.94 octave below f_2). This greatly expedited the testing process at the cost of neglecting some intersubject variation in the precise suppressor frequencies at the STC tip and tail.

At both suppressor frequencies, $f_{S(\text{tip})}$ and $f_{S(\text{tail})}$, the level of the suppressor was successively incremented in 5-dB steps until the drop in DPOAE level (below its unsuppressed value) first exceeded the 5-dB criterion. An estimate of the precise suppressor level required for the criterion reduction was then obtained by linearly interpolating between the final two suppressor steps. These measurements were repeated a total of three times in each subject, and the results averaged to obtain the reported value of T_E. Repeatability was very good, to within a few decibels between different sessions.

DPOAEs were recorded using custom-built hardware driven by custom software. Outgoing and incoming signals were digitized at the rate of 22 050 Hz by the Turtle Beach ''Tahiti'' sound card onboard a PC, and converted to analog by a ''patient room unit,'' a custom system of converters, amplifiers, and attenuators developed in our lab for Poul Madsen Medical Devices, Ltd. Our acoustic probe housed two speakers (Knowles Electronics, model ED-8238) for generating primary and suppressor tone stimuli, and two microphones (Knowles Electronics, model FG-3329) for detecting otoacoustic emissions, wired in parallel to improve SNR. Microphone signals were processed online using a synchronous time-averaging algorithm developed and tested in our lab (Kulik, 1995). Synchronous time averaging has many advantages over traditional spectral averaging for DPOAE detection, most important of which is noise-floor reduction (Kulik, 1995). Briefly, the DPOAE recording is broken up into a number of segments and averaged prior to calculating the amplitude spectrum by FFT. To prevent the averaging from reducing the signal components of interest (primary tones, DPOAE), they must be correlated between all segments, while uncorrelated components (noise) are reduced. The primary tones and DPOAE are synchronized, or phase-locked, with the recorded segment by selecting stimulus frequencies according to a specific mathematical relationship with the sampling rate. To determine DPOAE level, we recorded 200 frames of 1024 points each (about 46 ms), such that each frame had exactly an integer number of cycles of both primary tones and the DPOAE. After time averaging and FFT, the noise floor was estimated from an average of six frequency bins (each with a resolution of about 20 Hz) around the DPOAE. At $f_{\text{DPOAE}}=2671$ Hz, the noise floor during our recordings (conducted in a soundproof booth) was typically between −20 and −22 dB SPL.

Probe speakers were calibrated before each test session by making sound-pressure measurements in an artificial ear with a 2-cm^3 coupler. The third, suppressor tone had to be combined electrically with one of the primary tones in a

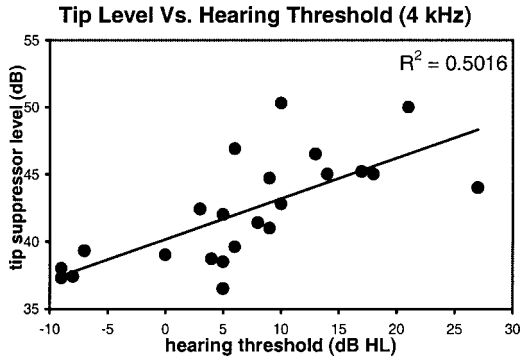


FIG. 5. Significant ($p<0.0001$), moderate ($R^2=0.50$), positive correlation between suppressor level in the tip region of the DPOAE STC and hearing threshold at 4 kHz.

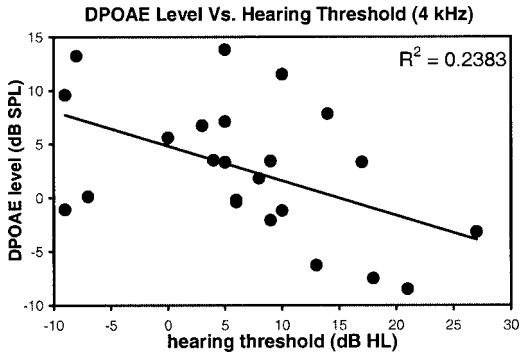


FIG. 7. Significant ($p<0.001$), weak ($R^2=0.24$), negative correlation between DPOAE level and hearing threshold at 4 kHz.

single speaker. To minimize distortion in the speakers while measuring T_E, the suppressor at frequency $f_{S(\text{tip})}$ was paired with f_1, and that at $f_{S(\text{tail})}$ with f_2.

Hearing thresholds at 4 kHz were measured in our soundproof booth with the Madsen Micro 5 clinical audiometer, by standard audiometry (method of limits). By measuring at 4 kHz, about the frequency of the f_2 primary, we follow the consensus that the place of resonance for f_2 on the basilar membrane is the dominant place of origin of the $2f_1-f_2$ DPOAE (Probst *et al.*, 1991). The level of 4 kHz was chosen because it yielded the highest SNR for DPOAE detection, and because noise-induced hearing loss typically develops at higher frequencies before lower ones, particularly at 4 kHz (e.g., Liberman and Kiang, 1978). Hearing thresholds in our subject pool ranged from −10 to 27 dB HL.

The testing procedures described above (measuring T_E and hearing threshold) were also attempted at $f_2=2$ and $f_2=3$ kHz in all subjects, and at $f_2=6$ kHz in a few of the subjects (the levels and frequency ratios of the primary tones, the suppression criterion, and the relative frequencies of the suppressors were the same as those used in the test at $f_2=4$ kHz). At these frequencies, sufficiently high-level DPOAEs could be recorded in only about half of the subjects. We required the unsuppressed DPOAE level to be at least 2 standard deviations plus the 5-dB suppression criterion above the mean noise floor. This requirement, as well as the low levels of primary tones used, are the probable causes for the lower than usual prevalence of DPOAEs at 2, 3, and 6 kHz. Therefore, results at these frequencies are based on a rather marginal number of data.

III. RESULTS

Twenty-three ears of 18 human subjects were tested. Figure 3 is a scatterplot of subjects' age versus threshold of hearing at 4 kHz. There is a weak correlation between the two variables (least-squares linear regression: $R^2=0.22$, $p<0.001$). Thus, although the hearing threshold of all subjects (−9 to 27 dB HL) is within ''normal'' range, it is unlikely that random variability alone (e.g., audiogram fine structure) can account for intersubject threshold differences.

Figures 4 and 5 show the main experiment result, scatterplots of suppressor level at the tail and tip on the DPOAE suppression tuning curve, respectively, versus hearing threshold. There is no correlation between threshold and the tail suppressor level, but there is a significant positive correlation ($R^2=0.50$, $p<0.0001$) between threshold and the tip suppressor level. Also, there is less variability in the tail data (mean=75.9, s.d.=2.2) than in the tip data (mean=42.2, s.d.=4.0). The data points of Fig. 5 are subtracted from Fig. 4 to give the tip-to-tail suppressor level difference (T_E), plotted against threshold in Fig. 6. The negative correlation between T_E (mean=33.8, s.d.=4.6) and threshold is significant ($R^2=0.45$, $p<0.0001$), and is the result of the increase in suppressor level at the tip of the DPOAE STC.

For comparison with Fig. 6, in Fig. 7 the unsuppressed

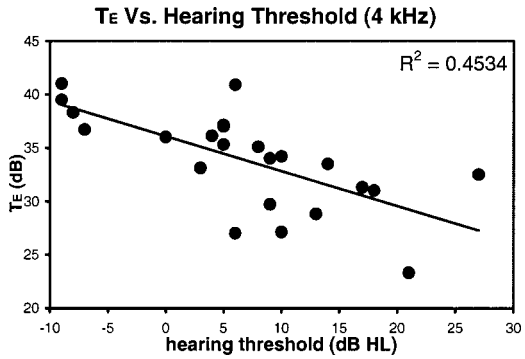


FIG. 6. Significant ($p<0.0001$), moderate ($R^2=0.45$), negative correlation between the tip-to-shoulder suppressor level difference (T_E;) on the DPOAE STC (Fig. 4 minus Fig. 5) and hearing threshold at 4 kHz.

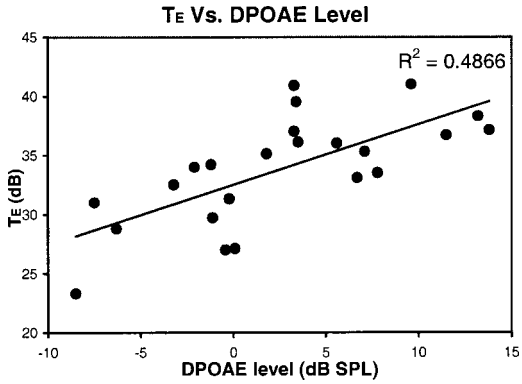


FIG. 8. Significant ($p<0.0001$), moderate ($R^2=0.49$), positive correlation between T_E and DPOAE level.

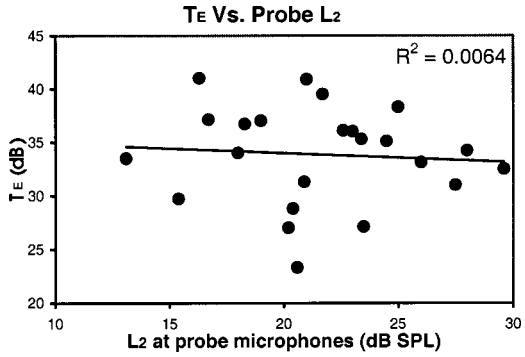


FIG. 9. No correlation between T_E and the level of the higher-frequency primary tone (L_2), measured at the acoustic probe microphones.

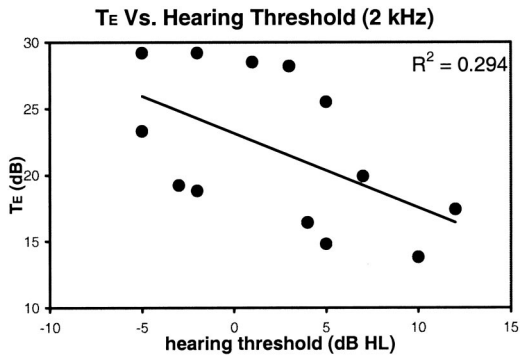


FIG. 11. Significant ($p<0.01$), low-moderate ($R^2=0.29$), negative correlation between T_E and hearing threshold at 2 kHz.

$2f_1-f_2$ DPOAE level is plotted against hearing threshold. As is commonly reported, there is a weak correlation between the two ($R^2=0.24$, $p<0.001$); R^2 here is comparable to that obtained from larger data sets (e.g., Kimberley *et al.*, 1997). A somewhat stronger correlation has been found between hearing threshold and the threshold primary tone level required to evoke detectable $2f_1-f_2$ DPOAEs (e.g., Nelson and Kimberley, 1992). In our subject pool, T_E correlated with hearing threshold more strongly than DPOAE level.

As one might expect, Fig. 8 shows that there is a significant correlation between T_E and DPOAE level ($R^2=0.49$, $p<0.0001$). An analogous correlation, between an elevation of tip suppressor level and a furosemide-induced decrease in DPOAE level, was recently found in the rabbit (Martin *et al.*, 1998b).

Finally, in Fig. 9 we show T_E as a function of the level of the higher-frequency primary tone (L_2). Calibration in an artificial ear (2-cm^3 coupler) does not take into account intersubject variability in ear canal length, volume, and shape, and in the depth of probe insertion. Since T_E has been shown to decrease with increasing primary tone level (Mills, 1998), it is important to check to what extent the reported correlation between T_E and threshold has been confounded by a systematic variation of T_E with in-ear L_2. The absence of correlation in Fig. 9 indicates that the relationship between T_E and hearing threshold was not confounded by our inability to keep primary tone level at the eardrum exactly constant across subjects. Note that L_2 was recorded by the microphones in the acoustic probe, not at the location of the eardrum.

Results similar to those presented above were obtained using different sets of primary tone (and suppressor tip and tail) frequencies, but were based on fewer ears tested. With $f_2=3$ kHz, DPOAEs sufficiently above the noise floor (2 standard deviations plus the 5-dB suppression criterion) could be detected in 14 of the 23 ears, and with $f_2=2$ kHz in 13 ears. Plots of T_E versus hearing threshold at 3 and 2 kHz are given in Figs. 10 and 11, respectively. A significant negative correlation ($p<0.01$) is seen in both cases ($R_2=0.40$ at 3 kHz; $R^2=0.29$ at 2 kHz). The main experiment results are summarized in Table I.

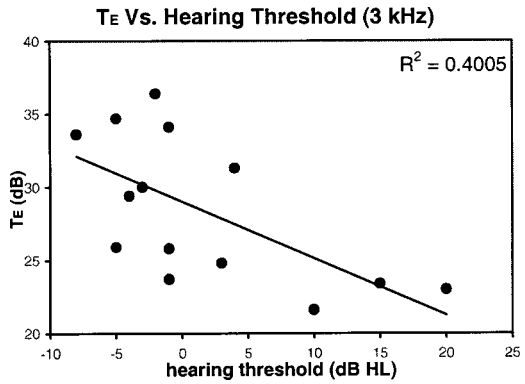


FIG. 10. Significant ($p<0.01$), moderate ($R^2=0.40$), negative correlation between T_E and hearing threshold at 3 kHz.

IV. DISCUSSION

The main result reported in this paper was a significant negative correlation between the tip-to-tail distance on a DPOAE suppression tuning curve (T_E) and the threshold of hearing in humans (Figs. 6, 10, 11; Table I). Notably, the basis for this relationship was primarily an increase in the tip suppressor level with hearing threshold; variation in suppressor level in the tail of the DPOAE STC did not correlate significantly with threshold (Figs. 4, 5; Table I). This pattern of decreased sensitivity in the tip and unchanged sensitivity in the tail regions is also seen in basilar membrane and auditory-nerve tuning curves following selective outer hair cell (i.e., cochlear amplifier) impairment (e.g., Liberman and Dodds, 1984; Ruggero and Rich, 1991). Likewise, the re-

TABLE I. Summary of the main experiment results. For the three sets of frequencies tested, the mean, standard deviation, and strength of correlation with hearing threshold is given for suppressor level in the tip and tail regions of the DPOAE STC, and for the tip-to-tail suppressor level difference (T_E).

		$f_2=4$ kHz ($n=23$)	$f_2=3$ kHz ($n=14$)	$f_2=2$ kHz ($n=13$)
T_E	Mean	33.8	28.4	21.9
	s.d.	4.6	5.0	5.7
	R^2	0.45	0.40	0.29
$L_{S(\text{tip})}$	Mean	42.2	44.4	44.3
	s.d.	4.0	4.3	2.9
	R^2	0.50	0.31	0.20
$L_{S(\text{tail})}$	Mean	75.9	72.8	66.2
	s.d.	2.2	2.4	3.3
	R^2	0.02	0.09	0.15

ported increase in the mean value of T_E with frequency (Table I) is also observed in basilar membrane tuning curves (Ulfendahl, 1997).

The above observations support, but do not prove, the theoretically postulated relationship between DPOAE STC tip-to-tail difference and cochlear amplifier gain (Mills, 1998; Pienkowski, 2000). Hearing threshold is, at best, an indirect measure of CA gain, and the experiment conducted constitutes only a feasibility study. A larger subject pool, including a population of patients diagnosed with noise-induced hearing loss (NIHL), is necessary to provide more conclusive (but still indirect) evidence. It is difficult to obtain a large pool of subjects with NIHL whose DPOAE levels are sufficiently high for the suppression protocol described here. (Papers often report the DPOAE levels of subjects with NIHL even though they lie within the range of noise-floor fluctuation—such low-level DPOAEs are obviously of no use in a suppression study.) A direct experiment to test the model predictions requires measuring basilar membrane responses and simultaneously recording DPOAEs from anesthetized animals. As the cochlear amplifier is reversibly inhibited by experimental manipulation, it would be possible to determine the rate of decrease of the T_E value with the fall of the BM response (i.e., lowered CA gain). Still, on the basis of model results (Mills, 1998; Pienkowski, 2000), and the results obtained in this study, we speculate that some of the threshold variability among ''normally hearing'' subjects can be attributed to differences in cochlear amplifier gain. The correlation of age and threshold (Fig. 3) upholds the plausibility of this assumption. We further hypothesize that threshold elevation in mild to moderate levels of hearing loss may be detectable with the suppression paradigm. Finally, the highest T_E values, generally obtained from subjects with the best thresholds, were about 40 dB. According to Mills' prediction (Mills, 1998), this puts the maximum cochlear amplifier gain in humans at about 45 dB, within the range of values measured in other mammals from basilar membrane vibration.

The strength of the observed correlation between T_E and hearing threshold was low to moderate: at $f_2=4$ kHz, $R^2=0.45$ ($n=23$ ears); $f_2=3$ kHz, $R^2=0.40$ ($n=14$); $f_2=2$ kHz, $R^2=0.29$ ($n=13$). Correlations of comparable strength have been reported between DPOAE level and hearing threshold, and between primary tone level at the threshold for DPOAE detection and hearing threshold (Harris, 1990; Nelson and Kimberley, 1992; Lonsbury-Martin and Martin, 1990; Kimberley *et al.*, 1997). A two-point (tip and tail) DPOAE suppression tuning ''curve'' requires about an order of magnitude greater measuring time than a single DPOAE recording, and an SNR that is higher by at least the criterion for suppression. Thus, in a screening application it appears inferior to the standard DF gram. Still, the DPOAE suppression measure might have a clinical use in a more comprehensive OAE-based diagnostic instrument.

V. CONCLUSIONS

A significant, moderate strength, negative linear correlation was observed between the tip-to-tail distance (T_E) on a DPOAE STC and hearing threshold in human subjects. This provides indirect support for theoretically derived relations between T_E and cochlear amplifier gain (Mills, 1998; Pienkowski, 2000). It also suggests that threshold elevation in mild levels of hearing loss may be detectable using the DPOAE suppression paradigm.

Abdala, C., Sininger, Y. S., Ekelid, M., and Zeng, F. G. (**1996**). ''Distortion product otoacoustic emission suppression tuning curves in human adults and neonates,'' Hear. Res. **98**, 38–53.

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.

Bialek, W. S., and Wit, H. P. (**1984**). ''Quantum limits to oscillator stability: Theory and experiments on acoustic emissions from the human ear,'' Phys. Lett. **104A**, 1973–1978.

Brown, A. M., and Kemp, D. T. (**1984**). ''Suppressibility of the $2f_1-f_2$ stimulated acoustic emissions in gerbil and man,'' Hear. Res. **13**, 29–37.

Brownell, W. E. (**1983**). ''Observations on a motile response in isolated outer hair cells,'' in *Mechanisms of Hearing*, edited by W. R. Webster and L. M. Aitkin (Monash University Press, Clayton, Australia), pp. 5–10.

Cai, Y., and Geisler, C. D. (**1996**). ''Suppression in auditory-nerve fibers of cats using low-side suppressors. III. Model results,'' Hear. Res. **96**, 126–140.

Cooper, N. P. (**1996**). ''Two-tone suppression in cochlear mechanics,'' J. Acoust. Soc. Am. **99**, 3087–3098.

Dallos, P. (**1988**). ''Cochlear neurobiology: Some key experiments and concepts of the past two decades,'' in *Auditory Function: Neurobiological Bases of Hearing*, edited by G. W. Edelman, W. E. Gall, and M. Cowan (Wiley, New York), pp. 153–188.

deBoer, E. (**1996**) ''Modeling the cochlea,'' in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer, New York), pp. 258–317.

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.

Frank, G., and Kossl, M. (**1995**). ''The shape of $2f_1-f_2$ suppression tuning curves reflects basilar membrane specializations in the moustached bat,'' Hear. Res. **83**, 151–160.

Geisler, C. D., Yates, G. K., Patuzzi, R. B., and Johnstone, B. M. (**1990**). ''Saturation of outer hair cell receptor currents causes two-tone suppression,'' Hear. Res. **44**, 241–256.

Geisler, C. D., and Nuttall, A. L. (**1997**). ''Two-tone suppression of basilar membrane vibrations in the base of the guinea pig cochlea using ''low-side'' suppressors,'' J. Acoust. Soc. Am. **102**, 430–440.

Gold, T., and Pumphrey, R. J. (**1948**). ''Hearing. I. The cochlea as a frequency analyzer,'' Proc. R. Soc. London, Ser. B **135**, 462–491.

Gold, T. (**1948**). ''Hearing. II. The physical basis of the action of the cochlea,'' Proc. R. Soc. London, Ser. B **135**, 492–498.

Harris, F. P. (**1990**). ''Distortion-product otoacoustic emissions in humans with high frequency sensorineural hearing loss,'' J. Speech Hear. Res. **33**, 594–600.

Harrison, R. V. (**1981**). ''Rate-versus-intensity functions and related AP responses in normal and pathological guinea pig and human cochleas,'' J. Acoust. Soc. Am. **70**, 1036–1044.

Holley, M. C. (**1996**). ''Outer hair cell motility,'' in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer, New York), pp. 386–434.

Kemp, D. T. (**1978**). ''Stimulated acoustic emissions from within the human auditory system,'' J. Acoust. Soc. Am. **64**, 1386–1391.

Kemp, D. T. (**1979**). ''Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea,'' Arch. Otorhinolaryngol. **224**, 37–45.

Kimberley, B. P., Brown, D. K., and Allen, J. B. (**1997**). ''Distortion product otoacoustic emissions and sensorineural hearing loss,'' in *Otoacoustic Emissions: Clinical Applications*, edited by M. S. Robinette and T. J. Glattke (Thieme, New York), pp. 181–204.

Koppl, C., and Manley, G. A. (**1993**). ''Distortion product otoacoustic emissions in the bobtail lizard. II. Suppression tuning characteristics,'' J. Acoust. Soc. Am. **93**, 2834–2844.

Kulik, R. (**1995**). ''A synchronous averaging instrument for the measurement of distortion product otoacoustic emissions,'' M.A.Sc. thesis, Department of Electrical and Computer Engineering, University of Toronto.

Kummer, P., Janssen, T., and Arnold, W. (**1995**). ''Suppression tuning char-

acteristics of the $2f_1-f_2$ distortion product otoacoustic emission in humans,'' J. Acoust. Soc. Am. **98**, 197–210.
Liberman, M. C., and Dodds, L. W. (**1984**). ''Single-neuron labeling and chronic cochlear pathology. III. Stereocilia damage and alternations of threshold tuning curves,'' Hear. Res. **16**, 55–74.
Liberman, M. C., and Kiang, N.Y. (**1978**). ''Acoustic trauma in cats. Cochlear pathology and auditory-nerve activity,'' Acta. Oto-Laryngol. Suppl. **358**, 1–68.
Lonsbury-Martin, B. L., and Martin, G. K. (**1990**). ''The clinical utility of distortion-product otoacoustic emissions.'' Ear Hear. **11**, 144–154.
Martin, G. K., Probst, R., Scheinin, S. A., Coats, A. C., and Lonsbury-Martin, B. L. (**1987**). ''Acoustic distortion products in rabbits. II. Sites of origin revealed by suppression and pure-tone exposures,'' Hear. Res. **28**, 191–208.
Martin, G. K., Jassir, D., Stagner, B. B., Whitehead, M. L., and Lonsbury-Martin, B. L. (**1998a**). ''Locus of generation for the $2f_1-f_2$ vs $2f_2-f_1$ distortion-product otoacoustic emissions in normal-hearing humans revealed by suppression tuning, onset latencies, and amplitude correlations,'' J. Acoust. Soc. Am. **103**, 1957–1971.
Martin, G. K., Jassir, D., Stagner, B. B., and Lonsbury-Martin, B. K. (**1998b**). ''Effects of loop diuretics on the suppression tuning of distortion-product otoacoustic emissions in rabbits,'' J. Acoust. Soc. Am. **104**, 972–983.
Martin, G. K., Stagner, B. B., Jassir, D., Telischi, F. F., and Lonsbury-Martin, B. L. (**1999**). ''Suppression and enhancement of distortion-product otoacoustic emissions by interference tones above f_2. I. Basic findings in rabbits,'' Hear. Res. **136**, 105–123.
Mills, D. M., and Rubel, E. W. (**1994**). ''Variation of distortion product otoacoustic emissions with furosemide injection.'' Hear. Res. **77**, 183–199.
Mills, D. M. (**1997**). ''Interpretation of distortion product otoacoustic emission measurements. I. Two stimulus tones,'' J. Acoust. Soc. Am. **102**, 413–429.
Mills, D. M. (**1998**). ''Interpretation of distortion product otoacoustic emission measurements. II. Estimating tuning characteristics using three stimulus tones,'' J. Acoust. Soc. Am. **103**, 507–523.
Nelson, D. A., and Kimberley, B. P. (**1992**). ''Distortion-product emissions and auditory sensitivity in human ears with normal hearing and cochlear hearing loss,'' J. Speech Hear. Res. **35**, 1142–1159.
Patuzzi, R. B., Yates, G. K., and Johnstone, B. M. (**1989**). ''Outer hair cell receptor current and sensorineural hearing loss,'' Hear. Res. **42**, 47–72.
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.
Pienkowski, M. (**2000**). ''Evidence for a relationship between the suppression of distortion product otoacoustic emissions and hearing threshold,'' M.Sc. thesis, Department of Physiology, University of Toronto.
Plinkert, P. K., and Zenner, H.-P. (**1998**). ''Suppression of distortion product otoacoustic emissions (DPOAE) near $2f_1-f_2$ removes DP-gram fine structure—Evidence for a secondary generator,'' J. Acoust. Soc. Am. **103**, 1527–1531.
Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). ''A review of otoacoustic emissions,'' J. Acoust. Soc. Am. **89**, 2027–2067.
Probst, R., and Harris, F. P. (**1997**). ''Otoacoustic emissions,'' Adv. Oto-Rhino-Laryngol. **53**, 182—204.
Rhode, W. S. (**1977**). ''Some observations on two-tone interaction measured with the Mossbauer effect,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London).
Robles, L., Ruggero, M. A., and Rich, N. C. (1**989**). ''Nonlinear interaction in the mechanical response of the cochlea to two-tone stimuli,'' in *Cochlear Mechanisms. Structure, Function and Models*, edited by J. D. Wilson and D. T. Kemp (Plenum, New York), pp. 369–375.
Ruggero, M. A., and Rich, N. C. (**1991**). ''Furosemide alters organ of Corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane,'' J. Neurosci. **11**, 1057–1067.
Ruggero, M. A., Robles, L., and Rich, N. C. (**1992a**). ''Two-tone suppression in the basilar membrane of the cochlea: Mechanical basis of auditory-nerve rate suppression,'' J. Neurophysiol. **68**, 1087–1099.
Ruggero, M. A., Robles, L., Rich, N. C., and Recio, A. (**1992b**). ''Basilar membrane responses to two-tone and broadband stimuli,'' Philos. Trans. R. Soc. Lond. Biol. Sci. **336**, 307–314.
Ryan, A., and Dallos, P. (**1975**). ''Effect of absence of cochlear outer hair cells on behavioral auditory threshold,'' Nature (London) **253**, 44–46.
Sellick, P. M., Patuzzi, R. B., and Johnstone, B. M. (**1982**). ''Measurement of basilar membrane motion in the guinea pig using the Mossbauer technique,'' J. Acoust. Soc. Am. **72**, 131–141.
Taschenberger, G., and Manley, G. A. (**1998**). ''General characteristics and suppression tuning properties of the distortion-product otoacoustic emission $2f_1-f_2$ in the barn owl,'' Hear. Res. **123**, 183–200.
Ulfendahl, M. (**1997**). ''Mechanical responses of the mammalian cochlea,'' Prog. Neurobiol. (Oxford) **53**, 331–380.
Yates, G. K. (**1995**). ''Cochlear structure and function,'' in *Hearing*, edited by B. C. J. Moore (Academic, San Diego), pp. 41–74.
Zurek, P. (**1981**). ''Spontaneous narrowband acoustic signals emitted by human ears,'' J. Acoust. Soc. Am. **69**, 514–522.
Zwicker, E. (**1986**). ''Suppression and $(2f_1-f_2)$ difference tones in a nonlinear cochlear preprocessing model with active feedback,'' J. Acoust. Soc. Am. **80**, 163–176.

DPOAE group delays versus electrophysiological measures of cochlear delay in normal human ears

Ruurd Schoonhoven,[a] Vera F. Prijs, and Sandra Schneider
Leiden University Medical Center, Department of ENT/Audiology, P.O. Box 9600, 2300 RC Leiden, The Netherlands

(Received 15 July 2000; revised 15 January 2001; accepted 18 January 2001)

Group delays of $2f_1-f_2$ distortion product otoacoustic emissions (DPOAEs) were determined using both f_1- and f_2-sweep paradigms in 24 normal-hearing subjects. These DPOAE group delays were studied in comparison with cochlear delays estimated from derived band VIIIth nerve compound action potentials (CAPs) and auditory brainstem responses (ABRs) in the same subjects. The center frequencies of the derived bands in the electrophysiological experiment were matched with the f_2-frequencies in the DPOAE recording to ensure that DPOAEs and derived CAPs and ABRs were generated at the same places along the cochlear partition, thus allowing for a direct comparison. The degree to which DPOAE group delays are larger in the f_2- than in the f_1-sweep paradigm is consistent with a theoretical analysis of the so-called wave-fixed model. Both DPOAE group delays are highly correlated with CAP- and ABR-derived measures of cochlear delay. The principal result of this study is that ''roundtrip'' DPOAE group delay in the f_1-sweep paradigm is exactly twice as large as the neural estimate of the ''forward'' cochlear delay. The interpretation of this notion in the context of cochlear wave propagation properties and DPOAE-generating mechanisms is discussed. [DOI: 10.1121/1.1354987]

PACS numbers: 43.64.Jb, 43.64.Kc, 43.64.Pg, 43.64.Ri [BLM]

I. INTRODUCTION

Distortion product otoacoustic emissions (DPOAEs) can be measured in the ear canal when the ear is stimulated with two pure tones, the primaries, with frequencies f_1 and f_2 ($>f_1$). Generation of the distortion is generally thought to be concentrated in a region near the f_2 characteristic place where the overlap of the excitation patterns of f_1 and f_2 is maximal. DPOAEs provide a simple, noninvasive tool to study various properties of cochlear filter processes. With the phase-gradient method the group delay of the DPOAE can be determined (Kimberley *et al.*, 1993; O'Mahoney and Kemp, 1995). This measure of cochlear delay can be compared with cochlear mechanical group delays (Ruggero, 1994; Ruggero *et al.*, 1997), with neural group delays (Anderson *et al.*, 1971; Allen, 1983; Gummer and Johnstone, 1984; Joris and Yin, 1992), or PSTH latencies in single VIIIth nerve fibers (Kiang *et al.*, 1965; Versnel *et al.*, 1992a), with latencies of VIIIth nerve compound action potentials (CAPs) (Elberling, 1974; Eggermont, 1979) or with latency measures derived, e.g., from the auditory brainstem response (ABR) (Don and Eggermont, 1978; Neely *et al.*, 1988; Donaldson and Ruth, 1993; Serbetcioglu and Parker, 1999). To study the DPOAE's phase-frequency gradient, variation of the DPOAE frequency can be achieved either by varying f_1 (the f_1-sweep paradigm, with group delay D_1) or by varying f_2 (the f_2-sweep paradigm, with group delay D_2). Like other measures of cochlear delay DPOAE group delays decrease with increasing frequency of the stimulating primaries as related to the tonotopic organization of the cochlea. In early considerations of DPOAE group delays (e.g., Kimberley *et al.*, 1993) the ''roundtrip'' concept has been proposed, reflecting the idea that the measured delay consists of a ''forward'' delay of the primaries to the DP generation site and of a ''backward'' delay of the DP component traveling back to the ear canal.

Various experimental observations and theoretical considerations have complicated this elementary interpretation of DPOAE group delays. First, DPOAE group delays observed in the f_2-sweep paradigm are larger than delays found in the f_1-sweep paradigm; for the $2f_1-f_2$ component D_2 is reported to be typically 20%–40% larger than D_1 (O'Mahoney and Kemp, 1995; Moulin and Kemp, 1996a, b; Whitehead *et al.*, 1996; Bowman *et al.*, 1997). Qualitatively similar conclusions hold for the higher-order lower sideband components $3f_1-2f_2$ and $4f_1-3f_2$ (Moulin and Kemp, 1996b; Schneider *et al.*, 1999), but the group delay of the upper sideband component $2f_2-f_1$ is independent of the sweep paradigm (Moulin and Kemp, 1996a, b; Whitehead *et al.*, 1996; Schneider *et al.*, 1999). To explain the difference in group delays between f_1- and f_2-sweeps it was first noted that, when sweeping f_1, the place where distortion is generated, thought to be near the f_2-characteristic place, remains unchanged; in contrast, when sweeping f_2 the generation site may either remain unchanged (the place-fixed model) or move along with the envelope of the f_2-wave (the wave-fixed model) (Moulin and Kemp, 1996b; Prijs *et al.*, 2000). However, the notion has now arisen that the DPOAE recorded in the ear canal is composed of two components (Brown *et al.*, 1996; Shera and Guinan, 1999; Talmadge *et al.*, 1999). One is the f_{dp}-wave propagated backward from the generation site near the f_2-characteristic place; this component is thought to obey the wave-fixed model (Knight and Kemp, 2000; Prijs *et al.*, 2000). The other component results

[a]Electronic mail: r.schoonhoven@lumc.nl

from reflections, at the f_{dp} characteristic place, of the f_{dp} wave traveling apically from its generation site; this mechanism is thought to obey the place-fixed model. Which contribution dominates the response in the ear canal depends on the level of the primaries (Fahey and Allen, 1997) and in particular on their frequency ratio (Knight and Kemp, 2000). DPOAEs measured close to the optimal primary ratio of $f_2/f_1 \approx 1.2$ seem to be dominated by the wave-fixed component from the f_2-generation site (Prijs *et al.*, 2000; Knight and Kemp, 2000).

Because of the various uncertainties regarding the interpretation of DPOAE group delays we have designed an experiment to compare DPOAE group delays with electrophysiological measures of cochlear delay, which have been extensively documented and are much better understood. A comprehensive set of data was collected in a group of normal-hearing human subjects. The data comprise f_1- and f_2-sweep group delays of the $2f_1-f_2$ DPOAE component, and both VIIIth nerve CAP and ABR latencies in a half-octave derived band paradigm. In order to allow for a direct comparison of various delay measures, the center frequencies of the derived bands were chosen equal to the f_2-frequencies in the DPOAE experiment. This way the presumed generation regions of otoacoustic and electrophysiological responses were matched. For the same reason stimulus levels of pure tones (DPOAE) and clicks (CAP and ABR) were chosen at comparable hearing levels. The analysis first aims at comparing how DPOAE group delays and CAP and ABR latencies vary as a function of the generating place along the cochlear partition. Next, a direct statistical comparison of DPOAE group delays versus electrophysiological latencies at corresponding generator places is made. The results of the study are discussed in terms of the relation between f_1- and f_2-sweep DPOAE, and the forward and retrograde delays involved in the DPOAE group delay as measured in the ear canal.

II. METHODS

A. Subjects

Recordings were made in the left ears of 24 human subjects (12 male, 12 female) in the age range 18–35 years. Pure tone audiograms were made to assess normal hearing, using as a criterion that thresholds were 10 dB HL or better at all octave frequencies from 125 to 8000 Hz, with at most one point at 15 dB HL accepted. Prior to recording otoscopy was performed to remove earwax and to check for normal aspect of the eardrum.

B. Otoacoustic emissions

DPOAEs were recorded with an Etymotic ER10C probe with 40-dB microphone preamplifier, connected to a PC provided with an Ariel DSP board and running CUBDIS© software. The software was adapted to improve the rejection of artifacts and to allow for automatic recording of series of f_1- and f_2-swept phase profiles of the $2f_1-f_2$ DPOAE component. During the recording subjects were seated comfortably in a sound isolated room, with PC and investigator outside. First a DP-gram was made at primary levels of $L_1=65$ dB SPL and $L_2=55$ dB, SPL and with f_2 frequencies at half-octave steps from 0.7 to 8 kHz. Associated f_1 frequencies were chosen at f_2/f_1 ratios that we previously found to produce maximal DPOAE amplitude in a large group of normal ears. The ratio f_2/f_1 varied from 1.19 for $f_2=8$ kHz to 1.32 for $f_2=0.7$ kHz. Then at each of the above-mentioned fixed f_2's an f_1-sweep paradigm was applied. Recordings were made for seven values of f_1, centered around that used in the above DP-gram. While the primaries were presented continuously, repeated trials were collected of 160 averaged sweeps of 20.48-ms duration. A maximum of 16 trials were recorded, and averaged after rejection of artifacts; criterion for rejection was a noise level (averaged over eight frequency bins surrounding the DP frequency) above 0 dB SPL. However, a smaller number of averaged trials was considered sufficient if the absolute noise level reached a level of -25 dB SPL, or if the DP component reached a signal-to-noise ratio of at least 5 dB. In the present experiments (recordings in normal ears and close to the optimal f_2/f_1 ratio) the latter condition was fulfilled for the vast majority of data points. Frequency steps were chosen so that at least three recordings were made within a 2π range of DPOAE phase shift. DPOAE phase was unwrapped and the group delay D_1 of the DPOAE was determined from the slope of the linear regression line fitted to the DPOAE's phase-frequency relation. Before computing the fit, visual inspection of the seven data points was used to identify any anomalous points, e.g., due to improper phase unwrapping. Such points appeared to be rare (overall less than a few percent) and were discarded from the fit. Fits based on less than five remaining points were considered unreliable and not included in the further analysis. Similarly an f_2-sweep paradigm was applied at each of the f_1's used in the DP-gram to determine D_2. Choice of frequency step size, recording procedure, and further data analysis were analogous to the f_1-sweep paradigm.

C. Derived band CAPs and ABRs

Electrophysiological recordings were made using a standard clinical auditory evoked potential system. Subjects were lying comfortably in a sound isolated and electrically shielded room, with equipment and investigator outside. Auditory nerve CAPs were recorded with a noninvasive electrocochleographic procedure (Ruth *et al.*, 1988). The active recording electrode was a 2-mm foam wick connected to a 0.1-mm-diam teflon-coated silver wire that was guided through a 1-mm-diam silastic tube. The wick was soaked in conductive gel and placed on the eardrum. Typical impedances were 20–50 kΩ. The reference electrode was placed on the ipsilateral mastoid, and the ground electrode on the contralateral mastoid. Sound was presented through an ER3 insert ear phone, the foam plug of which was carefully inserted along the tube guiding the electrode wire, thus sealing the outer ear canal. Stimuli were 0.1-ms rectangular clicks with alternating polarity, presented at a repetition rate of 7.9/s. Click level was 70 dB nHL; its baseline-to-peak pressure corresponded with the rms pressure of a 98.4 dB SPL tone (cf. Burkard, 1984), as measured in a dedicated 2-cc cavity that accomodated both the insert phone and, at the

virtual position of the eardrum, a $\frac{1}{2}$-in. recording microphone (B&K 4134 connected to B&K 2610 measuring amplifier). Responses to 2048 clicks were averaged after preamplification (PAR113, 10000×), filtering (5–3000 Hz), and rejection of artefacts. ABRs were recorded in a parallel recording channel, with active electrode on the forehead, and reference and ground electrodes shared with those for the CAP. Here the filter passband was 30–3000 Hz. The contralateral ear was always masked with broadband noise.

After recording the unmasked responses, noise was electronically mixed with the click stimulus in the evoked potential system. An external noise input was used, fed with wideband noise led through a high-pass filter (Krohn Hite 3988, two cascaded sections of 48 dB/oct each, Butterworth setting). Wideband noise was set at a level of 96.5 dB SPL (fixed for all subjects) that was found to just completely mask CAP and ABR evoked by the wideband click. Cutoff settings of the high-pass noise were then varied in half-octave steps. When subtracting responses associated with two subsequent noise cutoff frequencies f_i and f_{i+1} ($f_{i+1}=\sqrt{2}f_i$), the center frequency f_c of the resulting narrow-band derived response is identified with the maximum in the difference of the two noise power spectra, which equals 0.92 f_i as we derived from Prijs and Eggermont (1981); f_c is slightly lower than the lower of the two cutoff frequencies due to the finite filter slopes. In order to achieve that the resulting derived band center frequencies in the electrophysiological recordings coincided with the f_2 frequencies in the DPOAE experiments, noise cutoff frequencies were therefore chosen at values $f_i=f_2/0.92$ of 12.3, 8.72, 6.17, 4.36, 3.07, 2.18, 1.54, 1.09, 0.77 and 0.55 kHz. After recording, derived CAP N_1 latencies were corrected for the acoustic delay of the click of 1.0 ms (as measured in the aforementioned cavity) and a synaptic delay of 0.8 ms (Eggermont, 1979) to estimate the underlying mechanical delay, D_{CAP}. Derived ABR peak V latencies were corrected by the mean male/female $I-V$ peak interval in the unmasked response, and then also by acoustic and synaptic delay, providing the estimate D_{ABR}.

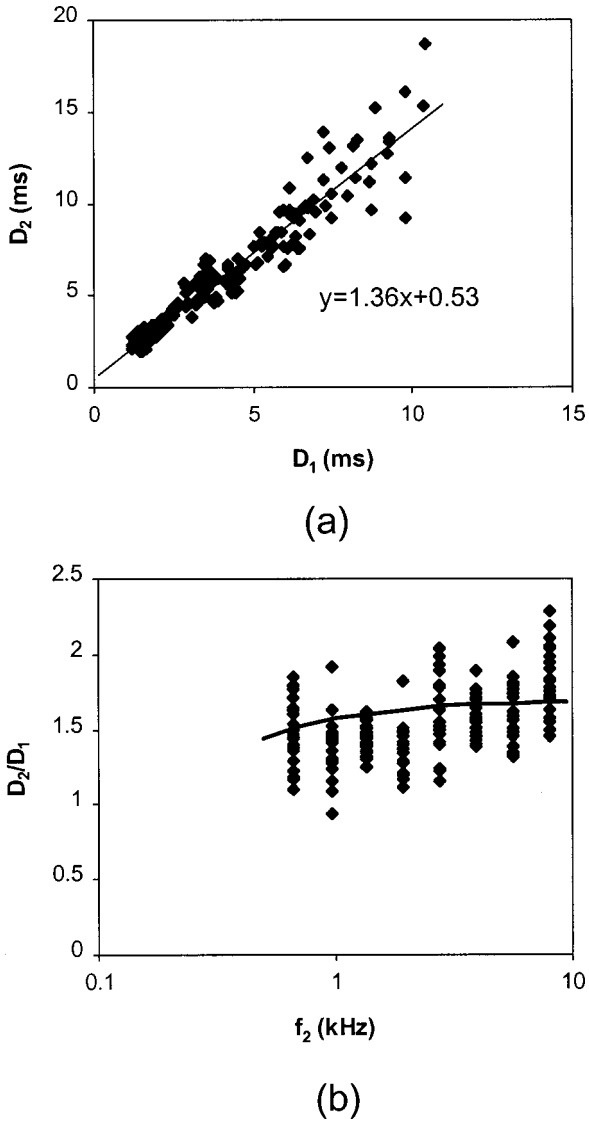


FIG. 1. Comparison of DPOAE group delays in f_1- vs f_2-sweep paradigm, D_1 and D_2. (a) Scatter plot with regression line and (b) D_2/D_1 ratio versus f_2 compared with the theoretical prediction on the basis of the wave-fixed model [Eq. (1)].

III. RESULTS

DPOAE group delays could reliably be measured in virtually all subjects for f_2 frequencies from 8 down to 0.7 kHz. Errors in the group delay (i.e., in the slope of the fitted regression lines) were typically 3%–5% for all f_2's and for both sweep paradigms. No significant differences across gender were found for f_1- and f_2-sweep group delays, D_1 and D_2, for any of the f_2-frequencies. Therefore, male and female data were pooled. Figure 1 presents a comparison of D_1 and D_2. The two group delay measures were highly correlated ($r=0.96$). D_2 was significantly larger than D_1, with a slope of the regression line of 1.36 ± 0.03 [Fig. 1(a)], and a D_2/D_1 ratio of 1.55 ± 0.23. Figure 1(b) presents D_2/D_1 for the individual data points in comparison with the theoretical prediction in the wave-fixed model (Schneider *et al.*, 2000; Talmadge *et al.*, 2000; Prijs *et al.*, 2000; Tubis *et al.*, 2000; Shera *et al.*, 2000):

$$\frac{D_2}{D_1}=2\frac{f_1}{f_2}. \tag{1}$$

Derived band CAPs and ABRs could reliably be measured in basically all subjects, but the small signal-to-noise ratio limited the yield of data particularly in the low center frequency bands. On average, derived CAPs could be identified in seven and derived ABRs in six out of nine frequency bands with center frequencies of 11.3 to 0.7 kHz. For illustration, high-pass noise masked and derived CAPs for one of the subjects are shown in Figs. 2(a) and (b), respectively; derived band ABR waveforms were similar to those extensively shown in the literature (e.g., Don and Eggermont, 1978; Schoonhoven, 1992). No significant male/female differences in derived CAP latencies were found in any derived band (*t*-test, $p>0.4$); latency differences between derived bands of, e.g., 8 vs 2 and 8 vs 1.4 kHz were slightly larger in males than in females, but the difference did not reach sta-

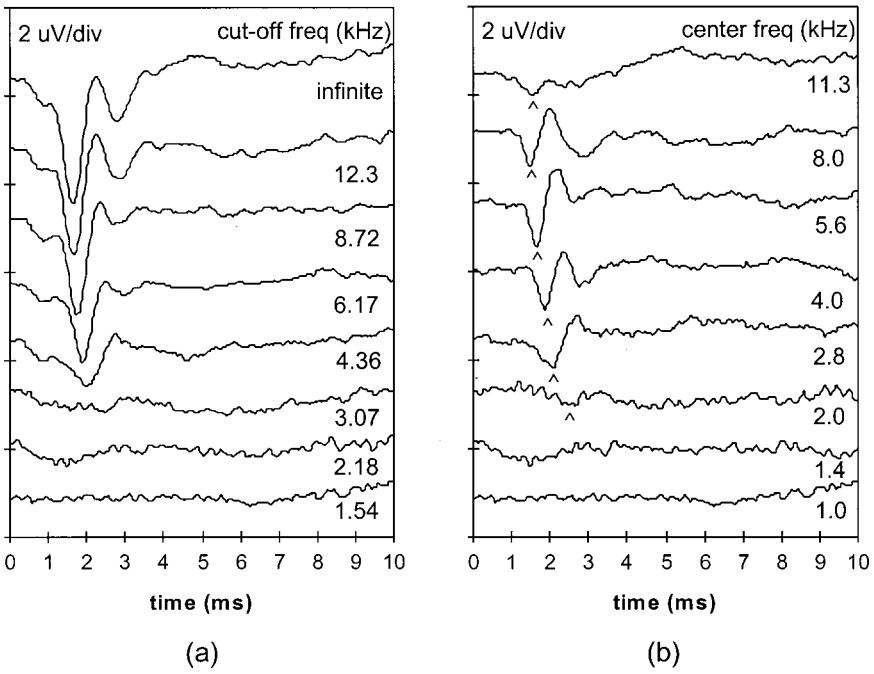


FIG. 2. Representative example of (a) high-pass noise masked CAPs and (b) derived band CAPs recorded from the tympanic membrane, at cutoff and center frequencies as indicated. The data were corrected for a 1-ms acoustic delay. Arrowheads in (b) mark the derived band N_1 responses identified in this subject.

tistical significance ($p=0.24$ and 0.07, respectively) (cf. Don *et al.*, 1993). Therefore all data were pooled across gender. The estimate of cochlear delay, D_{CAP}, was obtained by correcting derived band N_1 latencies with acoustic delay and synaptic delay. Derived ABR peak V latencies demonstrated a significant male/female difference in the 4- and 8-kHz frequency bands (*t*-test, $p<0.05$), but not in the other. The estimate of cochlear delay D_{ABR} was obtained by subtracting from all peak V latencies the mean I-V interval (males 4.14 ms, females 3.96 ms), and the acoustic and synaptic delays. The resulting values of D_{ABR} showed no gender differences and were therefore pooled.

Figure 3 presents a comprehensive survey of DPOAE group delays plotted against f_2, and delays estimated from the derived CAP and ABR data plotted against f_c. Regression equations [obtained from a linear fit of log(delay) to log(frequency)] are given in each panel. All average delay measures demonstrate the expected monotonic decrease with frequency; across subject variability appears to be larger in D_2 (for which nonmonotonic relations were seen in several individual subjects) than in D_1, and larger in D_{ABR} than in D_{CAP}. The exponents in the regression equations indicate that the frequency dependence of all four measures is highly similar. The constants in the equations for D_1 and D_2 obviously have a similar ratio as observed in Fig. 1. Comparison of regression parameters of D_{CAP} and D_{ABR} shows that the two quantities are basically equal. Indeed, statistical comparison revealed no significant difference between these two delay measures (paired *t*-test, $p<0.05$; $r=0.95$; $D_{\mathrm{ABR}}=0.98\,D_{\mathrm{CAP}}+0.04$). Comparing regression parameters of DPOAE group delays with those of D_{CAP} and D_{ABR} [Figs. 3(a) and (b) versus (c) and (d)] suggests that D_1 equals about twice D_{CAP} and D_{ABR}, while D_2 is almost three times as large.

A more direct comparison of otoacoustic and electrophysiological measures of delay is presented in Figs. 4 and 5 for CAP and ABR, respectively. The top part of these figures gives scatter plots of DPOAE versus electrophysiological delays together with the associated regression equations. Correlation coefficients between DPOAE and CAP delays in Fig. 4 were 0.92 for D_1 and 0.90 for D_2; for the comparison with ABR delays correlations were found of 0.92 and 0.89, respectively. The regression lines for D_1 are close to the lines representing the function $y=2x$, included in Figs. 4(a) and 5(a) for illustration. Therefore, in another way than in Fig. 3, the data again suggest that D_1 is essentially equal to $2*D_{\mathrm{CAP}}$ and to $2*D_{\mathrm{ABR}}$. A paired *t*-test confirmed that there is no statistically significant difference between D_1 and $2*D_{\mathrm{CAP}}$ nor between D_1 and $2*D_{\mathrm{ABR}}$. Obviously, the difference with D_{CAP} and D_{ABR} was highly significant for D_2. Finally, the bottom parts of Figs. 4 and 5 represent D_1 and D_2 as a function of f_2, plotted together with $2*D_{\mathrm{CAP}}$ and $2*D_{\mathrm{ABR}}$ as a function of f_c, respectively (cf. Fig. 3), also illustrating the similarity between D_1 and $2*D_{\mathrm{CAP}}$ and $2*D_{\mathrm{ABR}}$.

IV. DISCUSSION

The aim of the present project was to study DPOAE group delays in comparison with electrophysiological measures of cochlear delay in normal human ears. The experiment was designed to allow for the closest comparison possible. Two main critical aspects of our experimental approach are, first, the derived-band paradigm and the way narrow-band center frequency was matched to f_2 in the DPOAE experiment and, second, the selection of the level of the (transient) click versus the levels of the (continuous) primaries.

The choice for a click stimulus in combination with the derived band paradigm (Teas *et al.*, 1962), rather than, e.g., the use of tone bursts (cf. Neely *et al.*, 1988), was made for two reasons. First, a click warrants a high degree of neural

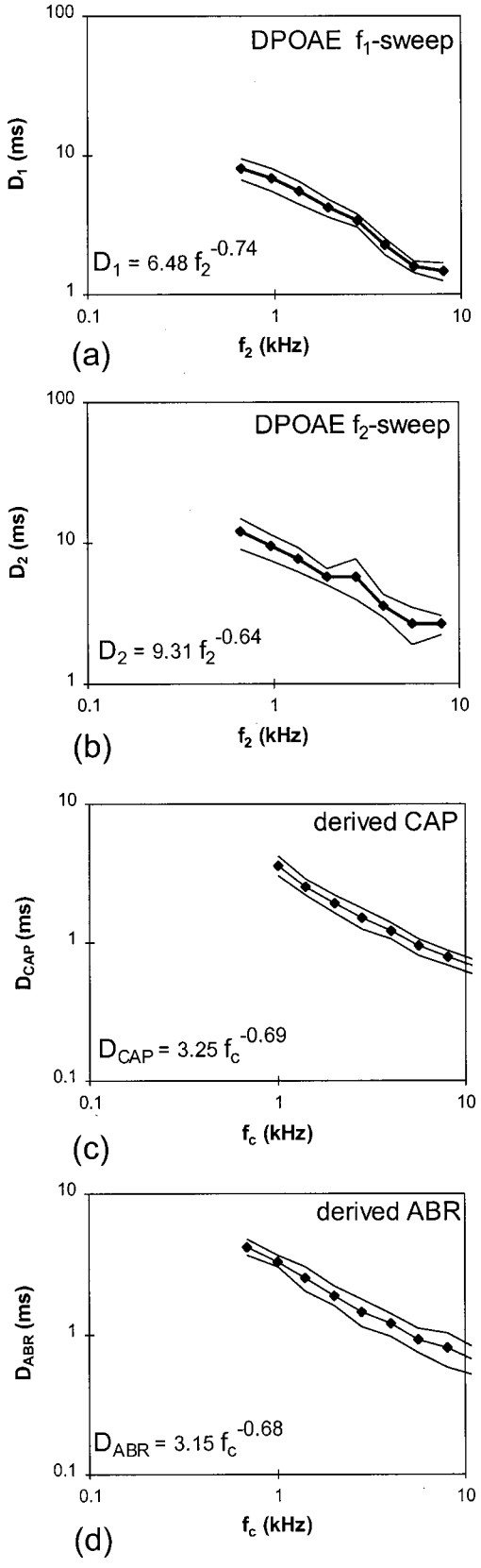


FIG. 3. (a) and (b) DPOAE group delays in the f_1- and f_2-sweep paradigms, D_1 and D_2 respectively, vs f_2. (c) and (d) Cochlear delays estimated from derived band CAPs and ABRs, D_{CAP} and D_{ABR}, respectively, vs center frequency f_c. Data represent mean±sd of 24 subjects, males and females pooled. Note the different scaling of the coordinate in (a) and (b) vs (c) and (d).

synchronization of the response and therefore a precise definition of its latency, which is not generally the case for tone bursts where rise time is always a compromise between the required synchrony and frequency specificity. Second, a solid theoretical framework is available relating click response latencies in the time domain to group delays determined in the frequency domain (see later in this work). The derived band method has found widespread application in studies based on VIIIth nerve CAP (Elberling, 1974; Eggermont, 1976, 1979; Prijs and Eggermont, 1981; Charlet de Sauvage *et al.*, 1983) and ABR (Parker and Thornton, 1978a, b, c; Don and Eggermont, 1978; Thornton and Farrell, 1991; Schoonhoven, 1992; Don *et al.*, 1993, 1996, 1997; Donaldson and Ruth, 1993; Serbetcioglu and Parker, 1999). The most solid validation of the derived band technique was given by Evans and Elberling (1982) who investigated the paradigm at the level of single auditory nerve fibers in the cat. These authors found that fibers responding to a click in the presence of a given high-pass noise masker but not in the presence of a high-pass masker with a one-octave lower cutoff indeed span a range of characteristic frequencies (CFs) of about one octave. These conclusions hold for CFs larger than 1–2 kHz in cat, and the authors extrapolated that in humans the procedure is valid for frequencies of 0.5–1 kHz and above. The center frequency of the derived band is defined as the mean characteristic frequency of fibers masked by the high-pass noise at one but not at the successive filter cutoff setting. When using one-octave filter steps, the center frequency roughly equals the lower of the two cutoff frequencies as can be concluded from the single-fiber data by Evans and Elberling (1982), and from studies where a second narrow-band masking noise was introduced to mask the derived band response itself (Parker and Thornton, 1978c; Stapells and So, 1999). By extrapolation the center frequency of half-octave bands will be slightly below the lower of the two successive cutoff frequencies. Using a more theoretical approach, the cochlear response pattern to the click under a given masking condition can be constructed by combining the power spectrum of the masker and the function that describes the degree of masking as a function of noise level. Subtracting such response patterns for successive high-pass filter settings produces the cochlear activation pattern underlying the derived band response. Since masking, at least in humans, increases about linearly with noise level (Eggermont and Odenthal, 1974), this approach predicts that the activation pattern underlying the derived band response follows exactly the same profile as the difference between successive high-pass noise power spectra (Prijs and Eggermont, 1981), as we assumed in the present study. From the various considerations we are confident that our derived band responses and DPOAEs originate from properly matched cochlear frequency regions, and estimate the uncertainty in this match to be not greater than about $\frac{1}{8}$ octave. Within this uncertainty the main conclusions drawn from the experimental data remain unchanged.

A second essential aspect of the experimental design concerns matching the intensities of the click in the electrophysiological experiment and the primaries in the DPOAE measurement, respectively. The issue is important because

both derived CAP and ABR latencies and DPOAE group delays demonstrate a notable level dependence (Eggermont and Don, 1980; Prijs and Eggermont, 1981; Bowman *et al.*, 1997; Wable *et al.*, 1997). Because of the opposite spectro-temporal properties of clicks and tones an appropriate match of their levels is not self-evident, and in fact several authors have come to different conclusions based on different considerations (e.g., Versnel *et al.*, 1992a; Recio *et al.*, 1998, and others). We have aimed at using a click of a similar (normal) hearing level as the primaries. From a psychoacoustic perspective, the total loudness of the click results from integration over about 20 critical bands, and a half-octave band covers almost two critical bands (Zwicker and Fastl, 1990). Thus the equivalent loudness of the click per half octave band is about 10 dB below the overall level of 70 dB HL. Together with a conductive hearing loss due to the tympanic membrane electrode, which we found to be in the order of 5 dB, the equivalent click level per derived band then closely matches the level of f_2, taking that differences between SPL and HL scales are small in the frequency region of interest. We are aware that this reasoning, although at least a partial justification of our approach, may not necessarily be conclusive. Given the intensity dependence of the various delay measures it is obvious that the results from the present experiments may not directly be extrapolated to other stimulus levels; literature data on intensity dependence, however, may help extrapolating the present results (e.g., Eggermont and Don, 1980; Bowman *et al.*, 1997).

The general properties of the present DPOAE group delays are in agreement with other studies in humans (Kimberley *et al.*, 1993; O'Mahoney and Kemp, 1995; Moulin and Kemp, 1996a, b; Wable *et al.*, 1996; Bowman *et al.*, 1997) and laboratory animals (Whitehead *et al.*, 1996; Schneider *et al.*, 1999) in that D_1 and D_2 decrease with f_2, and that D_2 is larger than D_1 to the amount shown in Fig. 1. The theoretical relation for the D_2/D_1 ratio [Eq. (1); Prijs *et al.*, 2000; Tubis *et al.*, 2000; Shera *et al.*, 2000] is obeyed by the experimental data in humans [Fig. 1(b)] to a similar extent as earlier shown for guinea pig DPOAE group delays (Prijs *et al.*, 2000). In view of the theoretical considerations in the latter paper this supports the notion that human DPOAE group delays as measured here near the f_2/f_1 ratio of 1.2 are dominated by the contribution from the f_2-characteristic place where the distortion product is generated and obey the wave-fixed model (Brown *et al.*, 1996; Talmadge *et al.*, 1999; Shera and Guinan, 1999; Prijs *et al.*, 2000; Knight and Kemp, 2000).

The general properties of the derived CAP and ABR results correspond to results from other studies. After Elberling (1974) had shown that derived VIIIth nerve responses can be recorded from the ear canal, the present study is the first to show that this can easily be done with a completely noninvasive procedure. Reliability of CAP (and ABR) responses was mostly judged by considering the consistency of responses in subsequent derived bands; errors in latency assignment are usually small (~0.1 ms) for responses with sufficient signal-to-noise ratio (SNR) [top traces in Fig. 2(b)]; for responses with lower SNR the uncertainty may be somewhat larger. The derived CAP amplitudes in the present experiment (Fig. 2) are a factor of 2 to 3 smaller than half-octave derived band responses in transtympanic recording (Eggermont, 1976), in agreement with other studies comparing extratympanic and transtympanic electrocochleography procedures (for an overview see, e.g., Schoonhoven *et al.*, 1995). For that reason, the yield of derived responses in the low-frequency bands is limited by the SNR. Also because of the sometimes unfavorable SNR, derived ABRs could not be detected in all frequency bands; in several other studies cited above better SNRs were obtained by using one-octave rather than half-octave bands, by applying a large number of averages and/or an objective method to determine the number of averages required to reach a predetermined noise level, or by using separate rarefaction and condensation clicks. Conclusions that I-V intervals are independent of frequency band, and that the purely cochlear delay can be obtained by correcting ABR peak V latency for the male/female I-V interval have been drawn before (e.g., Don and Eggermont, 1978; Don *et al.*, 1993). After correction with acoustic and synaptic delays of derived CAP and ABR latencies, D_{CAP} and D_{ABR} are similar to results of other studies including the numerical values of the fits given in Fig. 3 (cf. Eggermont, 1979). The exact numerical results depend somewhat on the corrections that were applied to CAPs and ABRs for acoustic and synaptic delay. A slightly (~0.1 ms) smaller acoustic delay for the ER3 tubes has been suggested (Whitehead *et al.*, 1996), which might, however, be cancelled by a slightly larger synaptic delay as sometimes assumed (e.g., 1.0 ms; Neely *et al.*, 1988; cf. Ruggero *et al.*, 1987).

The principal results of the present study concern the direct comparisons between DPOAE group delays and derived CAP/ABR latencies (Figs. 4 and 5). Since both measures were determined in the same subjects, a more rigorous analysis was feasible than in comparisons of DPOAE group delays with literature data on electrophysiological measures of cochlear delay (e.g., Kimberley *et al.*, 1993). The central conclusion from the analysis given in Figs. 4 and 5 is that D_1 is indistinguishable from both $2{*}D_{\mathrm{CAP}}$ and $2{*}D_{\mathrm{ABR}}$. The various foregoing considerations indicate that the overall uncertainties in all delay measures are typically in the order of 0.1 to 0.2 ms. Within that uncertainty range the difference between D_1 and $2{*}D_{\mathrm{CAP}}$ and $2{*}D_{\mathrm{ABR}}$ may in fact become statistically significant, but the conclusion that D_1 is very close to both $2{*}D_{\mathrm{CAP}}$ and $2{*}D_{\mathrm{ABR}}$ and that D_2 is greatly larger than either of those remains unchanged.

Which interpretation can be given to this conclusion regarding the understanding of DPOAE group delays? As argued above, the group delays measured with the present experimental parameters refer to the DPOAE component that is generated near the f_2-characteristic place X_2, and not, e.g., to a reflection component from the f_{dp}-characteristic place X_{dp}. Thus, DPOAE group delays reflect the the combined forward group delay of the primaries to X_2 and backward group delay of the DP component to the ear canal (Prijs *et al.*, 2000). The situation for the f_2-sweep paradigm is complicated by the fact that the generation site shifts along with the changing primary f_2. However, in the f_1-sweep, the generator is fixed at X_2, and simply reads out the phase changes of f_1 at that place. Therefore the forward delay com-

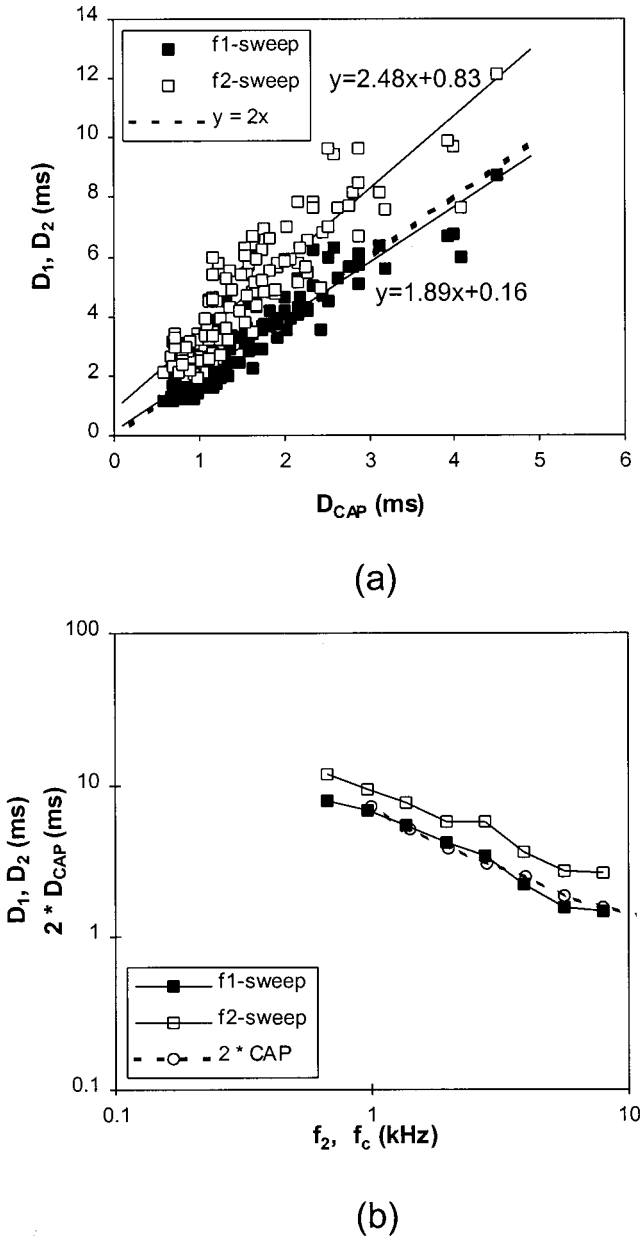


FIG. 4. DPOAE group delays versus delays estimated from derived band CAPs. (a) Scatter plot of D_1 and D_2 vs D_{CAP}, with regression lines, and compared with the relation $y=2x$. (b) D_1, D_2 and D_{CAP} vs f_2 and f_c, respectively.

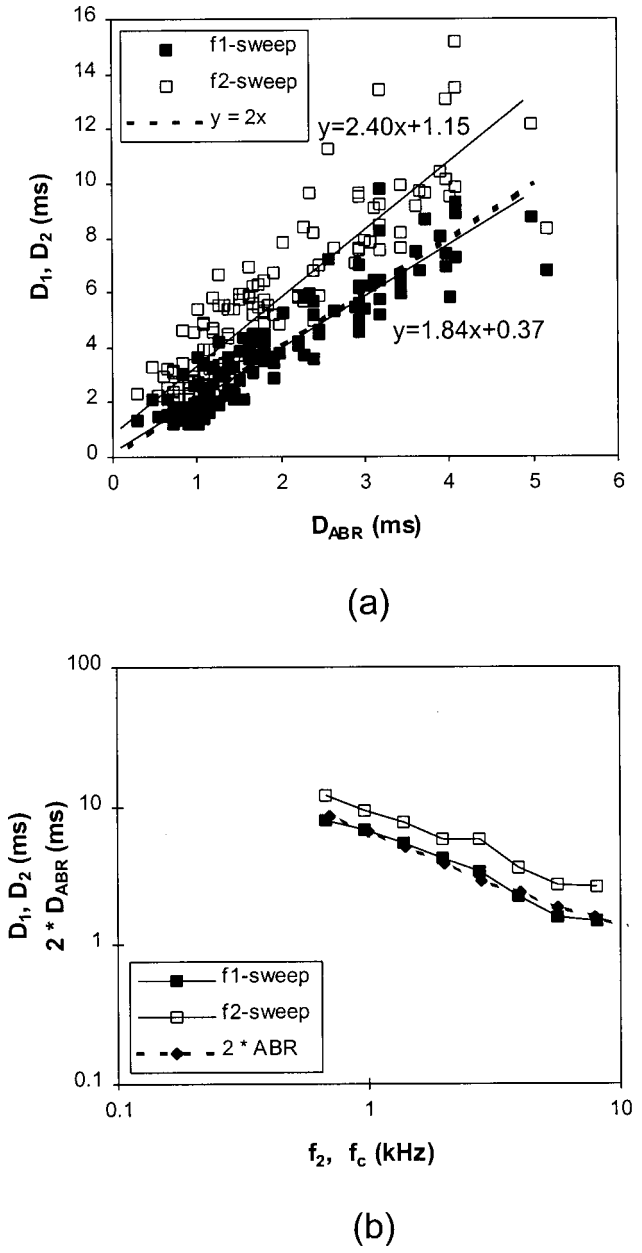


FIG. 5. DPOAE group delays versus delays estimated from derived band ABRs. (a) Scatter plot of D_1 and D_2 vs D_{ABR}, with regression lines, and compared with the relation $y=2x$. (b) D_1, D_2 and D_{ABR} vs f_2 and f_c, respectively.

ponent in D_1 can be interpreted as the group delay of the f_1-wave at X_2, which we will denote by $d(X_2;f_1)$. If we assume that the wave pattern set up along the cochlear partition does not depend on where in the cochlear fluids a given pressure variation is applied, then the wave pattern for the DP component traveling from X_2 to the cochlear base is equivalent with the wave pattern resulting from the external application of a pure tone with that same frequency. In that approach forward and backward group delays of the same frequency component are equal, and the backward group delay of f_{dp}, forming the second part of D_1, equals $d(X_2;f_{\mathrm{dp}})$.

Latencies of click responses in the time domain can be related to group delays in the frequency domain by noting that in a linear system the time delay of the center of gravity of the impulse response is identical to the weighted-average group delay (i.e., the phase-gradient averaged over frequency, weighted with the system's power gain). Although the cochlea is certainly not linear, this relation was still shown to hold for tone and click responses of single cochlear nerve fibers at low click levels, where PST histograms are only minimally affected by neural refractoriness (Goldstein *et al.*, 1971). Ruggero and co-workers more recently showed the same relation to hold at the level of the basilar membrane, at each stimulus level, and also demonstrated that amplitude and phase functions derived from click and tone responses are essentially identical and vary similarly as a function of intensity (Ruggero, 1994; Recio *et al.*, 1998; though the latter authors came to a different match of click and tone levels than we used in the present article). Since center frequencies of derived band CAP and ABR were lined

up with f_2, D_{CAP} and D_{ABR} may therefore be identified with the weighted average group delay at X_2, i.e., the weighted average of $d(X_2;f)$. At low stimulus levels this weighted average is essentially equal to the group delay at f_2, $d(X_2;f_2)$, because there the power gain of the system is dominated by the peak around the characteristic frequency. At intermediate stimulus levels as used in this study the low-frequency response, with its smaller phase gradient, becomes increasingly important in the weighting function. This makes that the weighted average group delay underlying our click responses will be somewhat smaller than the group delay at CF as such (Ruggero, 1994; Ruggero *et al.*, 1997; M. A. Ruggero, personal communication). In addition, at the levels used neural refractoriness shows a distinct influence on click PSTHs (e.g., Schoonhoven *et al.*, 1994), also making that their peak latency tends to shorten relative to the center of gravity of the underlying BM response. Finally, since the CAP is a convolution of the underlying neural firing pattern with the so-called unit contribution (Goldstein and Kiang, 1958), and noting that the unit contribution has a biphasic waveform and thus acts as a differentiator (Kiang *et al.*, 1976; Prijs, 1986; Versnel *et al.*, 1992b), the CAP latency may not correspond with the PSTH peak latency, but with a slightly smaller latency associated with the PSTH's onset slope. These are three reasons why D_{CAP} and D_{ABR} will eventually be smaller than $d(X_2;f_2)$. The foregoing results and reasoning can be summarized by

$$D_1 = 2 * D_{CAP} \tag{2}$$

with

$$D_1 = d(X_2;f_1) + d(X_2;f_{dp}) \tag{3}$$

and

$$D_{CAP} < d(X_2;f_2), \tag{4}$$

and of course similarly for D_{ABR}. Given that $f_{dp} < f_1 < f_2$, this result is consistent with the known phase behavior of traveling waves in the cochlea showing that (e.g., Ruggero *et al.*, 1997)

$$d(X_2;f_{dp}) < d(X_2;f_1) < d(X_2;f_2). \tag{5}$$

A solid quantitative comparison of Eqs. (2)–(5) cannot easily be given, particularly because the difference between D_{CAP} (and D_{ABR}) and $d(X_2;f_2)$ in Eq. (4) and the difference between the terms in Eq. (5) are not quantitatively known. Yet, Eqs. (2)–(5) are consistent with the notion that the f_1-sweep group delay, with experimental parameters as applied, is simply composed of a forward delay of the primaries to the generating place X_2 and a backward delay of the DP frequency to the ear canal. Experiments, e.g., combining DPOAE measurements with recording of single auditory nerve fiber period histograms and group delays based thereupon, may further clarify the relation between DPOAE group delays and forward and retrograde waves in the cochlea.

V. CONCLUSIONS

(1) The relation between f_1-sweep and f_2-sweep group delays of the $2f_1-f_2$ DPOAE in normal human ears is consistent with a theoretical elaboration of the wave-fixed model.
(2) Derived band CAP and ABR latencies, corrected for synaptic delay and for the ABR I-V interval, provide identical measures of cochlear delay.
(3) At least at the stimulus levels used, DPOAE group delays in the f_1-sweep paradigm amount to twice the associated, electrophysiologically determined cochlear delay.
(4) The findings are theoretically consistent with the notion that DPOAE group delay is a combination of the forward delay of the primaries to the f_2 (generation) place and a backward delay of the DP frequency from that place to the ear canal.

ACKNOWLEDGMENTS

This study was supported by the Netherlands Organization for Scientific Research (NWO) and by the Heinsius Houbolt Fund.

Allen, J. B. (**1983**). ''Magnitude and phase-frequency response to single tones in the auditory nerve,'' J. Acoust. Soc. Am. **73**, 2071–2091.

Anderson, D. J., Rose, J. E., Hind, J. E., and Brugge, J. F. (**1971**). ''Temporal position of discharges of single auditory nerve fibers within the cycle of a sine-wave stimulus: frequency and intensity effects,'' J. Acoust. Soc. Am. **49**, 1131–1139.

Bowman, D. M., Brown, D. K., Eggermont, J. J., and Kimberley, B. P. (**1997**). ''The effect of sound intensity on f_1-sweep and f_2-sweep distortion product otoacoustic emissions phase delay estimates in human adults,'' J. Acoust. Soc. Am. **101**, 1550–1559.

Brown, A. M., Harris, F. P., and Beveridge, H. A. (**1996**). ''Two sources of acoustic distortion products from the human cochlea,'' J. Acoust. Soc. Am. **100**, 3260–3267.

Burkard, R. (**1984**). ''Sound pressure level measurement and spectral analysis of brief acoustic transients,'' Electroencephalogr. Clin. Neurophysiol. **57**, 83–91.

Charlet de Sauvage, R., Cazals, Y., Erre, J. P., and Aran, J. M. (**1983**). ''Acoustically derived auditory nerve action potential evoked by electrical stimulation: An estimation of the waveform of single unit contribution,'' J. Acoust. Soc. Am. **73**, 616–627.

Don, M., and Eggermont, J. J. (**1978**). ''Analysis of the click-evoked brainstem potentials in man using high-pass noise masking,'' J. Acoust. Soc. Am. **63**, 1084–1092.

Don, M., Ponton, C. W., Eggermont, J. J., and Masuda, A. (**1993**). ''Gender differences in cochlear response time: An explanation for gender amplitude differences in the unmasked auditory brainstem response,'' J. Acoust. Soc. Am. **94**, 2135–2148.

Don, M., Vermiglio, A. J., Ponton, C. W., Eggermont, J. J., and Masuda, A. (**1996**). ''Variable effects of click polarity on auditory brainstem response latencies: Analyses of narrow-band ABRs suggest possible explanations,'' J. Acoust. Soc. Am. **100**, 458–466.

Don, M., Masuda, A., Nelson, R. A., and Brackmann, D. E. (**1997**). ''Successful detection of small acoustic tumors using the stacked derived-band auditory brainstem response amplitude,'' Am. J. Otology **18**, 608–621.

Donaldson, G. S., and Ruth, R. A. (**1993**). ''Derived band auditory brainstem response estimates of traveling wave velocity in humans. I. Normal hearing subjects,'' J. Acoust. Soc. Am. **93**, 940–951.

Eggermont, J. J. (**1976**). ''Analysis of compound action potential responses to tone bursts in the human and guinea pig cochlea,'' J. Acoust. Soc. Am. **60**, 1132–1139.

Eggermont, J. J. (**1979**). ''Narrow-band AP latencies in normal and recruiting human ears,'' J. Acoust. Soc. Am. **65**, 463–470.

Eggermont, J. J., and Odenthal, D. W. (**1974**). ''Electrophysiological investigation of the human cochlea. Recruitment, masking and adaptation,'' Audiology **13**, 1–22.
Eggermont, J. J., and Don, M. (**1980**). ''Analysis of the click-evoked brainstem potentials in humans using high-pass noise masking. II. Effect of click intensity,'' J. Acoust. Soc. Am. **68**, 1671–1675.
Elberling, C. (**1974**). ''Action potentials recorded along the cochlear partition in man,'' Scand. Audiol. **3**, 13–19.
Evans, E. F., and Elberling, C. (**1982**). ''Location-specific components of the gross cochlear action potential. An assessment of the validity of the high-pass masking technique by cochlear nerve fibre recording in the cat,'' Audiology **21**, 204–227.
Fahey, P. F., and Allen, J. B. (**1997**). ''Measurement of distortion product phase in the ear canal of the cat,'' J. Acoust. Soc. Am. **102**, 2880–2891.
Goldstein, M. H., and Kiang, N. Y. S. (**1958**). ''Synchrony of neural activity in electric responses evoked by transient acoustic stimuli,'' J. Acoust. Soc. Am. **30**, 107–114.
Goldstein, J. L., Baer, T., and Kiang, N. Y. S. (**1971**). ''A theoretical treatment of latency, group delay and tuning characteristics for auditory nerve responses to clicks and tones,'' in *Physiology of the Auditory System*, edited by M. B. Sachs (National Educational Consultants, Baltimore), pp. 133–141.
Gummer, A. W., and Johnstone, B. M. (**1984**). ''Group delay measurement from spiral ganglion cells in the basal turn of the guinea pig cochlea,'' J. Acoust. Soc. Am. **76**, 1388–1400.
Joris, P. X., and Yin, T. C. T. (**1992**). ''Responses to amplitude-modulated tones in the auditory nerve of the cat,'' J. Acoust. Soc. Am. **91**, 215–232.
Kiang, N. Y. S., Watanabe, T., Thomas, E. C., and Clark, L. F. (**1965**). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve*, MIT Research Monograph 35 (MIT, Cambridge, MA).
Kiang, N. Y. S., Moxon, E. C., and Kahn, A. R. (**1976**). ''The relationship of gross potentials recorded from the cochlea to single unit activity in the auditory nerve,'' in *Electrochleography*, edited by R. J. Ruben, C. Elberling, and G. Salomon (University Park, Baltimore), pp. 95–115.
Kimberley, B. P., Brown, D. K., and Eggermont, J. J., (**1993**). ''Measuring human cochlear traveling wave delay using distortion product otoacoustic emission phase responses,'' J. Acoust. Soc. Am. **94**, 1343–1350.
Knight, R. D., and Kemp, D. T. (**2000**). ''Indications of different distortion product otoacoustic emission mechanisms from a detailed f_1, f_2 study,'' J. Acoust. Soc. Am. **107**, 457–473.
Moulin, A., and Kemp, D. T. (**1996a**). ''Multicomponent acoustic distortion product otoacoustic emission phase in humans. I. General characteristics,'' J. Acoust. Soc. Am. **100**, 1617–1639.
Moulin, A., and Kemp, D. T. (**1996b**). ''Multicomponent acoustic distortion product otoacoustic emission phase in humans. II. Implications for distortion product otoacoustic emissions generation,'' J. Acoust. Soc. Am. **100**, 1640–1662.
Neely, S. T., Norton, S. J., Gorga, M. P., and Jesteadt, W. (**1988**). ''Latency of auditory-brainstem responses and otoacoustic emissions using tone-burst stimuli,'' J. Acoust. Soc. Am. **83**, 652–656.
O'Mahoney, C. F., and Kemp, D. T. (**1995**). ''Distortion product otoacoustic emission delay measurement in human ears,'' J. Acoust. Soc. Am. **97**, 3721–3735.
Parker, D. J., and Thornton, A. R. D. (**1978a**). ''The validity of the derived cochlear nerve and brainstem evoked responses of the human auditory system,'' Scand. Audiol. **7**, 45–52.
Parker, D. J., and Thornton, A. R. D. (**1978b**). ''Frequency specific components of the cochlear nerve and brainstem evoked responses of the human auditory system,'' Scand. Audiol. **7**, 53–60.
Parker, D. J., and Thornton, A. R. D. (**1978c**). ''Derived cochlear nerve and brainstem evoked responses of the human auditory system. The effect of masking in the derived band,'' Scand. Audiol. **7**, 73–80.
Prijs, V. F. (**1986**). ''Single-unit response at the round window of the guinea pig,'' Hear. Res. **21**, 127–133.
Prijs, V. F., and Eggermont, J. J. (**1981**). ''Narrow-band analysis of compound action potentials for several stimulus conditions in the guinea pig,'' Hear. Res. **4**, 23–41.
Prijs, V. F., Schneider, S., and Schoonhoven, R. (**2000**). ''Group delays of distortion product otoacoustic emissions: Relating delays measured with f_1- and f_2-sweep paradigms,'' J. Acoust. Soc. Am. **107**, 3298–3307.
Recio, A., Rich, N. C., Narayan, S. S., and Ruggero, M. A. (**1998**). ''Basilar-membrane responses to clicks at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **103**, 1972–1989.
Ruggero, M. A., and Rich, N. C. (**1987**). ''Timing of spike initiation in cochlear afferents: depending on site of innervation,'' J. Neurophysiol. **58**, 379–403.
Ruggero, M. A. (**1994**). ''Cochlear delays and traveling waves: Comments on 'Experimental look at cochlear mechanics,'' ' Audiology **33**, 133–142.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.
Ruth, R. A., Lambert, P. R., and Ferraro, J. A. (**1988**). ''Electrocochleography: methods and clinical applications,'' Am. J. Otology **9**, 1–11.
Schoonhoven, R. (**1992**). ''Dependence of auditory brainstem response on click polarity and high-frequency sensorineural hearing loss,'' Audiology **31**, 72–86.
Schoonhoven, R., Keijzer, J., Versnel, H., and Prijs, V. F. (**1994**). ''A dual filter model describing single-fiber responses to clicks in the normal and noise-damaged cochlea,'' J. Acoust. Soc. Am. **95**, 2104–2121.
Schoonhoven, R., Fabius, M. A. W., and Grote, J. J. (**1995**). ''Input/output curves to tone bursts and clicks in extratympanic and transtympanic electrocochleography,'' Ear Hear. **16**, 619–630.
Schneider, S., Prijs, V. F., and Schoonhoven, R. (**1999**). ''Group delays of distortion product otoacoustic emissions in the guinea pig,'' J. Acoust. Soc. Am. **105**, 2722–2730.
Schneider, S., Prijs, V. F., Schoonhoven, R., and van Hengel, P. W. J. (**2000**). ''f_1- versus f_2-sweep group delays of distortion product otoacoustic emissions in the guinea pig; experimental results and theoretical predictions,'' in *Recent Developments in Auditory Mechanics*, edited by H. Wada, T. Takasaka, K. Ikeda, K Ohyama, and T. Koike (World Scientific, Singapore, 2000), pp. 360–366.
Serbetcioglu, M. B., and Parker, D. J. (**1999**). ''Measures of cochlear travelling wave delay in humans. I. Comparison of three techniques in subjects with normal hearing,'' Acta Oto-Laryngol. **119**, 537–543.
Shera, C. A., and Guinan, J. J. (**1999**). ''Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs,'' J. Acoust. Soc. Am. **105**, 782–798.
Shera, C. A., Talmadge, C. L., and Tubis, A. (**2000**). ''Interrelations among distortion-product phase-gradient delays: Their connection to scaling symmetry and its breaking,'' J. Acoust. Soc. Am. **108**, 2933–2948.
Stapells, D. R., and So, M. R. (**1999**). ''High-pass noise masked/derived-band auditory brainstem responses: Cochlear contributions determined by narrowband maskers,'' in *Abstracts of XVI Biennial Symposium, International Evoked Response Audiometry Study Group, Tromsø, Norway*, edited by E. Laukli.
Talmadge, C. L., Long, G. R., 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., Tong, C., Long, G. R., and Dhar, S. (**2000**). ''Temporal aspects of otoacoustic emissions,'' in *Recent Developments in Auditory Mechanics*, edited by H. Wada, T. Takasaka, K. Ikeda, K. Ohyama, and T. Koike (World Scientific, Singapore, 2000), pp. 353–359.
Teas, D. C., Eldredge, D. H., and Davis, H. (**1962**). ''Cochlear responses to acoustic transients: An interpretation of whole-nerve action potentials,'' J. Acoust. Soc. Am. **34**, 1438–1459.
Thornton, A. R. D., and Farrell, G. (**1991**). ''Apparent travelling wave velocity changes in cases of endolymphatic hydrops,'' Scand. Audiol. **20**, 13–18.
Tubis, A., Talmadge, C. L., Tong, C., and Dhar, S. (**2000**). ''On the relationships between the fixed-f_1, fixed-f_2 and fixed-ratio phase derivatives of the 2 f_1-f_2 distortion product otoacoustic emission,'' J. Acoust. Soc. Am. **108**, 1772–1785.
Versnel, H., Schoonhoven, R., and Prijs, V. F. (**1992a**). ''Single-fibre and whole-nerve responses to clicks as a function of sound intensity in the guinea pig,'' Hear. Res. **59**, 138–156.
Versnel, H., Prijs, V. F., and Schoonhoven, R. (**1992b**). ''Round-window recorded potential of single-fibre discharge (unit response) in normal and noise damaged cochleas,'' Hear. Res. **59**, 157–170.

Wable, J., Collet, L., and Chéry-Croze, S. (**1996**). ''Phase delay measurements of distortion product otoacoustic emissions at $2f_1-f_2$ and $2f_2-f_1$ in human ears,'' J. Acoust. Soc. Am. **100**, 2228–2235.

Wable, J., Collet, L., Chéry-Croze, S., and Berger-Vachon, C. (**1997**). ''$2f_1-f_2$ distortion product otoacoustic emission latency: changes with frequency and level of primaries,'' Audiology **36**, 72–82.

Whitehead, M. L., Stagner, B. B., Martin, G. K., and Lonsbury-Martin, B. L. (**1996**). ''Visualization of the onset of distortion-product otoacoustic emissions, and measurement of their latency,'' J. Acoust. Soc. Am. **100**, 1663–1679.

Zwicker, E., and Fastl, H. (**1990**). *Psychoacoustics. Facts and Models* (Springer, Berlin).

Wave and place fixed DPOAE maps of the human ear

Richard D. Knight and David T. Kemp
Auditory Biophysics Group, Institute of Laryngology and Otology, University College London, 330/332 Gray's Inn Road, London WC1X 8EE, United Kingdom

(Received 3 July 2000; revised 27 December 2000; accepted 3 January 2001)

Human intermodulation distortion product otoacoustic emissions (DPOAE) can be a mixture of low and high latency components. They have different level, phase, and suppression characteristics, which indicate that emissions arise both from the frequency region of the primary tones directly and indirectly via the DP frequency place. Which component dominates the measured DPOAE in the ear canal depends on the stimulus parameters, especially the frequency ratio, f_2/f_1. Interference between the two emissions adds complexity to measurements of DPOAE. The behavior and even existence of whichever emission route is lower in level often cannot directly be deduced from the raw DPOAE data because the other emission covers it. It is therefore not known whether both emissions are present for all stimulus parameters or whether the trends seen in each emission when they are the dominant emission route continue under stimulus conditions when they are not dominant. In this study, the two DPOAE components are separated by a post-processing method. Previously, maps of raw DPOAE data against f_2/f_1 and DP frequency have been obtained. To separate the components, sets of data consisting of f_2/f_1 sweeps were transformed by an inverse Fourier transform into the time domain. The low and high latency components appeared as two distinct peaks because of their different phase gradients. These peaks were separated by windowing in the time domain and two frequency domain maps were reconstructed, representing the low and high latency DPOAEs. It was found that the low latency component of the $2f_1-f_2$ DP was only emitted strongly with f_2/f_1 between approximately 1.1 and 1.3. The removal of the high latency component revealed the low ratio edge of this region, at which the level falls sharply. However, the low latency emission has been traced at reduced amplitude over a wide range of stimulus parameters. Although previously only observed at small frequency ratios, the high latency component was found to be present widely in the lower sideband, its level reducing slowly at larger f_2/f_1. Its phase behavior changes in the lower sideband, being approximately constant with DP frequency at small ratios of f_2/f_1, but deviating from this at wider ratios. These results support the hypothesis that a DPOAE component which propagates to and is re-emitted from the DP frequency place (place fixed emission) is present across a wide parameter range. However, for all but the close primary condition the lower sideband DPOAE is dominated by direct emission from the region of f_2 and f_1 wave interaction (wave fixed emission). A simple transmission line model is presented to illustrate how the observed DPOAE maps can arise on the basis of this hypothesis. [DOI: 10.1121/1.1354197]

PACS numbers: 43.64.Jb, 43.64.Kc, 43.64.Ri [BLM]

I. INTRODUCTION

Distortion product otoacoustic emission (DPOAE) are of great research and clinical interest. Recent research interest has focused attention on the two possible paths by which lower sideband DPOAE [of the form $(a+1)f_1-af_2$, $a=1,2,3$, etc.] may travel to the ear canal once generated within the region of f_2 excitation.

A. PATH 1: The DP travels directly to the base from the generation site

Kemp and Knight (1999) have suggested that the mechanism of DP generation may, for certain f_2/f_1, preferentially direct energy basally. Kemp (1986) described $2f_1-f_2$ DP production with $f_2/f_1>1.1$ as ''wave fixed'' because its phase characteristics required the generator to move smoothly with the f_2 traveling wave in a frequency sweep.

Shera and Guinan (1999) labeled the same phenomenon ''distortion emission'' to emphasize that only nonlinearity (not a reflection) is required for an ear canal DP to be produced. Path 1 results in ''low latency'' DPOAEs.

B. PATH 2: The DP travels initially toward the apex, until it is reflected back toward the base

This mode of DPOAE, seen in the $2f_1-f_2$ DPOAE for small f_2/f_1, was separated by Kemp and Brown (1983) who suppressed activity at the DP place by adding an extra stimulus tone close to the DP frequency. It was described as ''place fixed'' by Kemp (1986), because the phase characteristics required reflectors to be fixed in position on the basilar membrane. TEOAEs have similar phase characteristics and have been said to require an impedance discontinuity (Kemp, 1978) or impedance irregularities along the basilar membrane to cause the reflection (Kemp, 1986; Lonsbury-Martin *et al.*, 1988; Zweig and Shera, 1995). Shera and Guinan (1999) described this mode of DPOAE as ''reflection DPs'' to emphasize that the ''turnaround'' mechanism is a reflection. Path 2 results in ''high latency'' DPOAEs.

Upper sideband DPOAE differ from this as the initial generation region is thought to be around the DP frequency place and the DP will propagate poorly beyond this place. Upper sideband DPOAE are always emitted predominantly by a place fixed mechanism from around the DP frequency place (e.g., Martin *et al.*, 1987, 1998; Kemp, 1998), although a small wave fixed emission is theoretically possible.

Martin *et al.* (1999, 2000) and Fahey *et al.* (2000) have shown that there may also be alternative basal sites for DP generation. These require more complex generation mechanisms, for example, involving the generation of intermodulation distortion from harmonic distortion products and difference tones derived from the original stimulus tones.

The region of ''reflection'' in lower sideband DPOAE probably lies in the DP frequency place region, as suggested by suppression studies using a third stimulus tone (e.g., Kummer *et al.*, 1995; Gaskill and Brown, 1996; Heitmann *et al.*, 1998), observation of DPOAE fine structure in frequency limited hearing loss and relationships between the fine structure of different distortion products (Mauermann *et al.*, 1999a, b) and comparison of experimental data with mathematical models (Talmadge *et al.*, 1999). If outer hair cell activity adequately compensates for the transmission losses in propagating back to the base and in reflection back a second time to the DP place, then the distortion product also could reflect repeatedly between the DP characteristic place and the base of the cochlea resulting in a feedback system causing interference effects, fine structure and occasionally resulting in spontaneous otoacoustic emissions (Kemp, 1980; Zweig and Shera, 1995; Talmadge *et al.*, 2000).

Whichever DP emission path results in the higher DP amplitude in the ear canal will dominate the phase behavior of the measured DPOAE. This depends on various aspects of the stimulus parameters, particularly the ratio of f_2/f_1. Previous studies (Kemp, 1986; Knight and Kemp, 1999) have shown that there appears to be a modal change in the $2f_1-f_2$ DPOAE phase behavior. In a frequency sweep with f_2/f_1 held constant, if f_2/f_1 is less than about 1.1 the $2f_1-f_2$ DPOAE has a steep phase gradient (high latency), indicating that the DP is dominated by a place fixed emission. But if the ratio is larger a shallow phase gradient is found (low latency), indicating that the emission is wave fixed.

Knight and Kemp (2000a) investigated the relationship between DP frequency and the ratio of f_2/f_1 in detail and showed that the place fixed behavior of the $2f_1-f_2$ DP when f_2/f_1 is small is a continuation of the patterns that are seen in the $2f_2-f_1$ DP for all f_2/f_1. However, the ear canal $2f_1-f_2$ DP was dominated by a wave fixed emission mode when f_2/f_1 exceeded approximately 1.1. A similar relationship was also seen between the $3f_1-2f_2$ and $3f_2-2f_1$ DPs.

The DP source, in the region of the f_2 frequency place, has been assumed to have a constant directivity for all f_2/f_1 (e.g., Allen and Fahey, 1992). However, more recently it has been suggested that this directivity changes as the frequency ratio changes (Knight and Kemp, 2000a).

Kemp and Knight (1999) have suggested a mechanism by which this may occur. Essentially this involves analyzing the modeled phase of f_1 and f_2 in the presumed generating region around f_2 and calculating the phase gradient of the $2f_1-f_2$ distortion product across this region as it is generated. This gradient changes from positive to negative as the ratio between f_1 and f_2 changes, resulting in constructive summation of the $2f_1-f_2$ distortion product occurring either toward the base or the apex. Therefore the phase gradients in the region of DP generation may to a large extent control the relative amplitudes of the wave and place fixed emission mechanisms.

As the DPOAE behavior which is observed is dominated by whichever mode is stronger, it is not known whether the two DP paths both continue to be significant for all f_2/f_1 ratios or whether DP production tends to be bi-modal with only one mode usually present and just a small transition region at which both emissions can occur. If both modes were always present and could be separated then the phase behavior of the ''residual'' or ''minority'' emissions would be of great interest. The phase characteristic of such a ''residual emission'' could be seen in relation to the traveling wave excitation patterns and the systematic trends that have previously only been seen in restricted stimulus parameter conditions could be pursued over a wider range of frequency ratios. Observation of the ''place fixed'' phase behavior will be of particular interest. The crossover between the lower and upper sidebands, the lower sideband f_2/f_1 ratio at which the phase pattern which continues from the upper sideband changes, and the phase behavior in the lower sideband at moderately wide ratios would provide further clues regarding the locations of the initial DP generating region relative to the f_2 frequency place and the reflecting region relative to the f_{DP} place.

However, separating the wave and place fixed emissions is not a straightforward task. One way is to add a third stimulus tone to suppress activity at the DP place, leaving only the DP emitted directly from the f_2 region. Comparison of DPOAE measurements obtained with and without the third tone can yield the DP place contribution by vector subtraction (Dreisbach and Siegel, 1999). However, it can be difficult to ensure that the DP place is fully suppressed without the f_2 region, and hence DP generation, also being affected. This risk is greatest with small stimulus frequency ratios (Siegel *et al.*, 2000).

An alternative method is pursued here, in which DPOAE is evoked by two stimulus tones and the wave and place fixed DPOAE are separated by a phase gradient dependent post-processing method. Shera and Zweig (1993) and Zweig and Shera (1995) have used Fourier transforms to investigate the spectral oscillations in SFOAE. Stover *et al.* (1996), Fahey and Allen (1997), and Kalluri and Shera (2000) used Fourier transforms on DPOAE frequency sweeps to observe the phase gradient-derived latency. It has been shown by Knight and Kemp (2000a) that an inverse fourier transform of a DP sweep with constant f_2/f_1 results in a quasi-time domain representation with two distinct peaks, the relative sizes of the peaks depending on the ratio of f_2/f_1 at which the sweep is performed. In the present study, these low and high latency parts are windowed in the pseudo-time domain

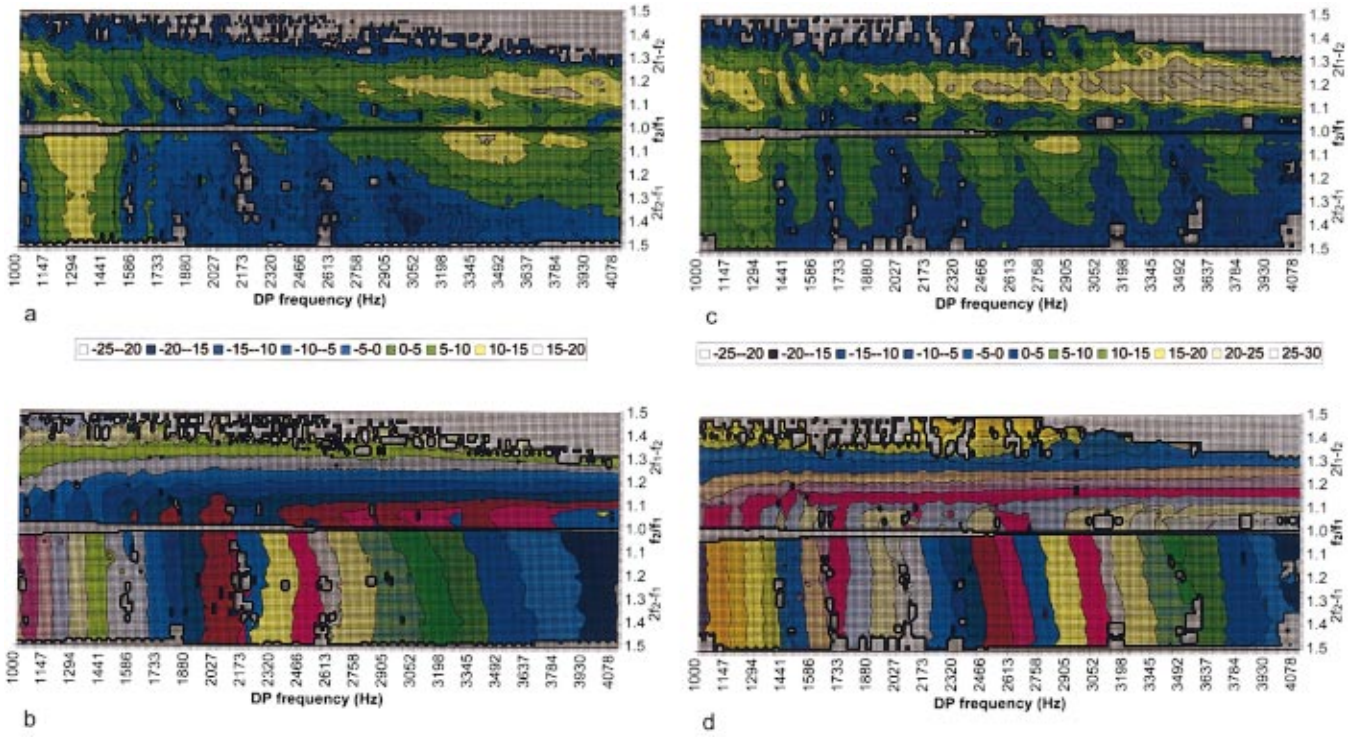


FIG. 1. $2f_1-f_2$ and $2f_2-f_1$ DPOAE data for subjects RDK in parts (a) and (b) and RN in parts (c) and (d) before the separation process [previously presented in J. Acoust. Soc. Am. **107**, 457–473 (2000)]. For each figure presented in this format, level (dB SPL) is shown in parts (a) and (c) in 5 dB steps and phase is shown in parts (b) and (d) in 180° steps. Phase becomes more positive toward the top and left of the chart.

and then transformed back to the frequency domain separately.

This method has the advantage of avoiding the unpredictable nonlinear effects of a third tone. A detailed and wide frequency sweep is required in order for the Fourier transformations to be adequately performed, from which definite separation of emissions with different phase behaviors can be achieved. This technique has been used to observe the ''wave'' and ''place'' fixed DPOAEs separately, allowing behavior which is usually masked to be uncovered.

There is an underlying difference in the third tone and the Fourier transform methods. The third tone separates the emission modes on the basis of place on the basilar membrane and it turns out that these emissions are different modes. The Fourier transform method separates the emission modes on the basis of their different phase gradients and it seems that they are from different places. The methods are comparable if the place fixed emission is from the region of the DP frequency place and the wave fixed emission is from elsewhere. Kalluri and Shera (2000) have shown an example in which similar results were obtained by the two methods, and Knight and Kemp (2000a) showed that the later peak in the time domain can be selectively suppressed by a third tone close to the DP frequency, supporting the theory that the two methods are comparable.

II. MEASUREMENTS AND DATA ANALYSIS

Data presented in this study have been obtained from an additional analysis of data presented by Knight and Kemp (2000a, reproduced in Fig. 1). The DPOAE measurements were obtained using the left ears of two subjects: a male (RDK) aged 29 and a female (RN) aged 24, both of whom had normal auditory thresholds and middle ear function.

DPOAE measurements were obtained with $L_1=L_2=70$ dB SPL with DP frequencies ranging from 1 kHz to 4.1 kHz in 12 Hz steps. Stimulus frequency ratios from 1.01 to 1.5 were employed. The DPOAE phase is corrected for the phase changes induced by the probe loudspeakers and microphone (the system detects the stimulus phase via the microphone and corrects automatically) but no phase correction is applied for acoustic delays along the ear canal.

From these data, the two emission components (wave and place fixed) have distinctly different latencies and therefore can be separated by the following post-processing method. A 512 point frequency array was set up, ranging from 0 to 3100 Hz, thereby preserving the original 3100 Hz range. An exponential frequency spacing was adopted in order to linearize the underlying curve in the phase versus frequency relationship of the place fixed DPOAE and therefore produce clearer peaks in the time domain (Zweig and Shera, 1995). The exponential frequency points were calculated as follows:

$$f_i = 1000 \times 4.1^{((i-1)/511)} - 1000,$$

where f_i = the frequency of the ith array point.

The measured data points were frequency shifted down 1000 Hz and inserted into this array with linear interpolation between data points where required. The level and phase data were converted to complex number format for the inverse Fourier transformation.

No frequency domain windowing was applied before the inverse Fourier transformation. Although there is a risk of

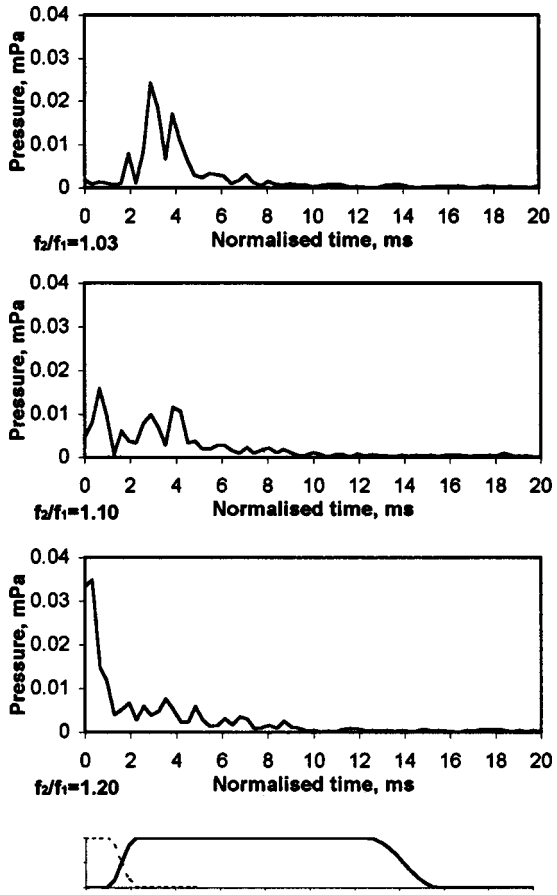


FIG. 2. Magnitude of inverse Fourier transforms of DPOAE constant frequency ratio sweeps selected from data in Fig. 1 (subject RDK, a and b). The time windowing employed to separate the early and late components is indicated at the bottom of the figure. In each case the total area under the curve corresponds to a DPOAE level of between 15 and 20 dB SPL.

artifacts resulting from this, the artifacts would be easily distinguishable from the signal in this case as the artefacts would be spread throughout the time window whereas the true signal would form peaks early in the time window. The time domain windowing which was subsequently employed would therefore largely remove any such artefacts. The advantage of this approach was that the data were not therefore weighted in favor of data from the middle of the frequency range and the final separated data were easy to compare directly with the original unprocessed data.

As the frequency array had an exponential characteristic, the meaning of the time scale of the inverse Fourier transform was not straightforward. The time scale in the time domain which is indicated in Fig. 2 was calculated as follows:

$$\text{Time increment per point (ms)} = 1000/\text{frequency range (Hz)}.$$

This gives a time increment of approximately 0.32 ms per point. This calculated time interval most accurately reflects true time for data from the middle of the frequency range (2–3 kHz). It is an underestimate for the lower frequencies and an overestimate for the higher frequencies. The definition of this time scale is not relevant for the subsequent windowing and processing of the data. The transformation into the ''time'' domain produced separate amplitude peaks for the low and high latency components (Fig. 2).

The peaks were separated by time windowing, the windowing being designed to separate the wave and place fixed DPOAE at the point at which the magnitude function had a dip between the wave and place fixed peaks. The chosen crossover point for the ''early'' and ''late'' time windows was the fifth data point, which is about 1.6 ms. A flat topped window with a raised cosine edge was adopted. The cosine curve ''edge'' extended from the third to the seventh data points (approx. 1.00–2.25 ms). The ''late'' window was rolled off more gradually between the 40th and 50th data points (12.9–16.1 ms), comfortably after the DPOAE peaks had died down to the noise floor.

Each windowed part of the time domain response was returned to the frequency domain by a forward Fourier transform. The data were returned to level and phase format and converted back to the original 256 point linear frequency array. Separate area charts were then constructed for the wave and place fixed DPOAE.

III. RESULTS

A. Low latency component

The low latency portion of the $2f_1-f_2$ DPOAE had a high level when f_2/f_1 was between 1.1 and 1.3 [Figs. 3 and 4, parts (a) and (b)]. Outside this range the level fell rapidly. The DPOAE is likely to be restricted to a limited ratio range because the phases of the DP sources in the generating region of the basilar membrane can only positively sum in the basal direction with a limited range of ratios (Kemp and Knight, 1999). The phase contours were approximately horizontal, indicating that phase was dependent on the frequency ratio, but almost independent of DP frequency. There was good vertical agreement between successive constant f_2/f_1 sweeps, with phase being steadily more positive at larger stimulus frequency ratios.

The low latency portion of the $2f_2-f_1$ DPOAE (Figs. 3 and 4, parts a and b) was of very small amplitude. In one subject (RDK, Fig. 3) it was mostly below -20 dB SPL, but the other subject (RN, Fig. 4) had a clear low latency peak in the inverse Fourier transform (not shown) which was above -20 dB SPL in the frequency domain [the lower half of Fig. 4(a)] with most combinations of f_1 and f_2. In the frequency domain the associated phase, where measurable, had a tendency to gradually be more positive when f_2/f_1 was smaller, but the phase was less dependent on the DP frequency.

The absence after processing of the fine structure which had been seen in the unprocessed data of Fig. 1, even when the wave fixed emission appeared to dominate, suggests that the fine structure was caused by interference between the low and high latency DPOAE components.

B. High latency

The phase of the high latency part was predominantly dictated by the emission frequency, in contrast to the low latency DPOAE [Figs. 3 and 4, parts (c) and (d)]. The level of the high latency component was less sensitive to f_2/f_1 changes as DP frequency was held constant (vertical move-

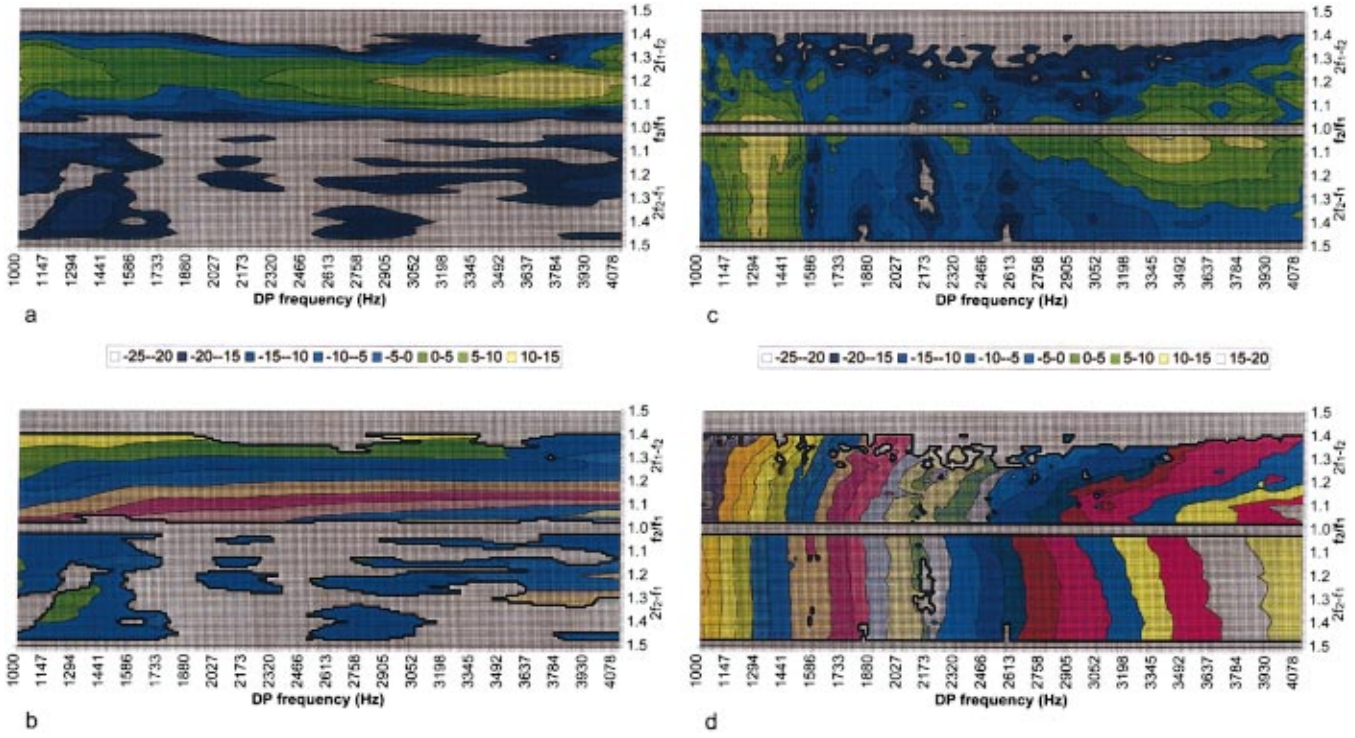


FIG. 3. Low (a and b) and high (c and d) latency windowed DPOAE for subject RDK, extracted from data in Fig. 1(a and b).

ment on the figures). Factors which affect the level of the place fixed DP and are sensitive to f_2/f_1 include the amount of apical propagation from the site of DP generation and the amount of suppression of the apically traveling DP by the proximity of f_1 and f_2.

There are interesting secondary relationships with the frequency ratio visible in the data. At higher frequencies (above about 2 kHz) the upper sideband DPOAE phase from both subjects shows a slight, accelerating negative drift toward smaller ratios (indicated by the phase contours bending to the left). The phase of the lower sideband high latency DPOAE was, at small frequency ratios, a continuation of the upper sideband phase. However, above $f_2/f_1 \sim 1.07$ the phase was steadily more positive at wider frequency ratios

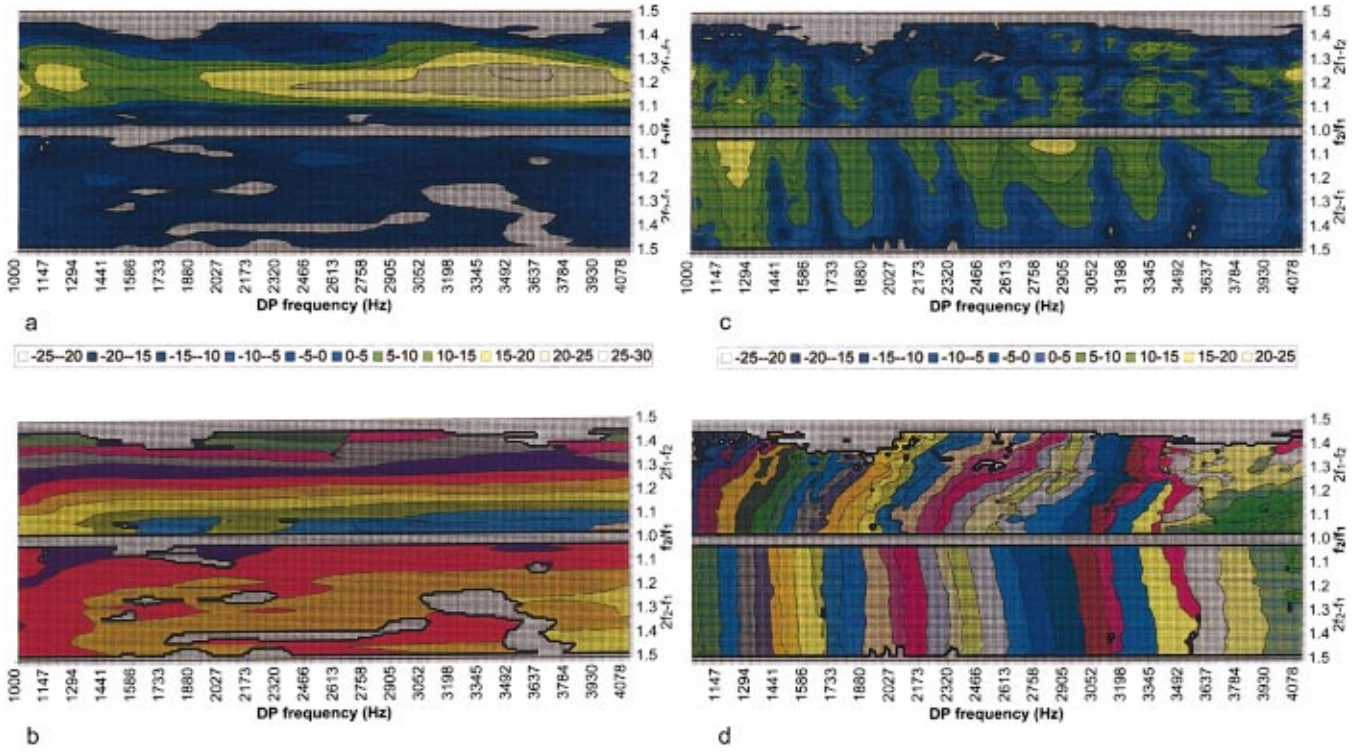


FIG. 4. Low (a and b) and high (c and d) latency windowed DPOAE for subject RN, extracted from data in Fig. 1(c and d).

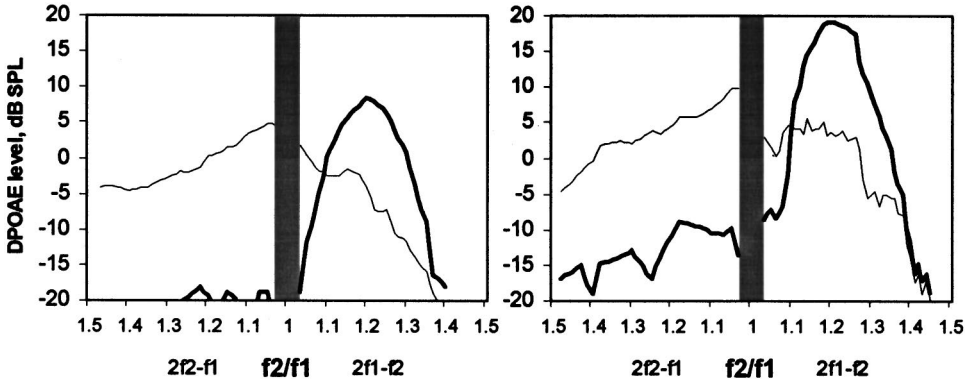


FIG. 5. Wave fixed (thick line) and place fixed (thin line) DPOAE level for subject RDK (left) and RN (right) plotted against frequency ratio, calculated as a power average from the data in Figs. 3 and 4. The shaded region where f_2/f_1 is small indicates where data could not be obtained because of limitations on the smallest frequency spacing at which the DPOAE could be resolved from the stimuli. Random noise floor estimates are approximately −30 dB SPL for the wave fixed DPOAE and −25 dB SPL for the place fixed DPOAE (see Table I).

(contours leaning to the right). The retention of the pronounced fine structure in the high latency DPOAE indicates that this is not caused by interactions with a low latency emission. Similar patterns were observed from $3f_1-2f_2$ and $3f_2-2f_1$ DPOAE data (not shown), except that the DPs were only present when f_2/f_1 was small.

C. Wave and place fixed level, averaged across DP frequency

In order to see clearly the preferred frequency ratio for the wave and place fixed DPOAE separately, a power average of the DPOAE level from Figs. 3 and 4 was calculated across DP frequency and plotted against frequency ratio (Fig. 5). The results show clearly that the wave fixed emission favors a restricted ratio range from $f_2/f_1=1.1-1.35$ in the lower sideband only, whereas the place fixed emission is present in both sidebands much more widely with the strongest emission when f_2/f_1 is small. The residual upper sideband wave fixed emission can be seen from subject RN, typically at −10 to −15 dB SPL.

In the lower sideband when f_2/f_1 is larger, the rapid fall seen in the wave fixed level as f_2/f_1 increases beyond 1.3 and the more gradual fall in the place fixed level with increasing f_2/f_1 means that the levels of the two components become similar. Tubis *et al.* (2000) have suggested that at wide ratios, the lower sideband place fixed DPOAE level may exceed the wave fixed level (as is also the case when $f_2/f_1>1.1$). The present data suggest that this is indeed possible although it does not quite occur in our data, the difference may be due to different stimulus conditions being adopted.

IV. DISCUSSION

In general, the low latency derived DPOAE phase is dependent on the ratio f_2/f_1 but changes only slowly with DP frequency (so phase contours run left to right across the figures). The high latency derived DPOAE phase changes far more rapidly with DP frequency (Figs. 3 and 4). This difference between the low and high latency DPOAE phase is not surprising and is precisely what would be expected due to the windowing employed in the time domain, because windowing at low latency excludes components with a steep phase gradient. Of far greater interest is the way in which the phase contours behave with respect to the frequency ratio (vertical in the figures) when the sweeps are recombined. The agreement of the phase of adjacent sweeps confirms that the patterns seen in the data do not arise from merely an artifact. It is significant that a substantial amplitude of DP above the noise level was obtained for both modes under most stimulus conditions.

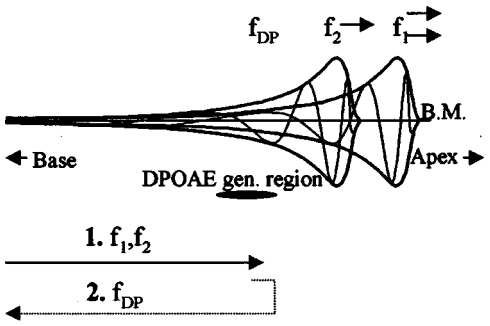
The levels of the wave and place fixed emissions have very different relationships to f_2/f_1, as shown clearly in Fig. 5. The wave fixed emission is emitted strongly only in a restricted range of the lower sideband and a residual emission is present more widely. This limited range of f_2/f_1 for which the wave fixed DPOAE is strong may result from the phase gradient across the DPOAE generating region only being favorable for basal propagation under these conditions (Kemp and Knight, 1999).

The place fixed emission is emitted most strongly when f_2/f_1 is close to 1 in either sideband but the ratio is less critical. As f_2/f_1 increases the level of the place fixed emission falls faster in the lower sideband than the upper sideband, this may be because the phase gradient switches from favoring apical propagation to favoring basal propagation as the DPOAE generating region leaves the DP frequency place and moves with the f_2 envelope.

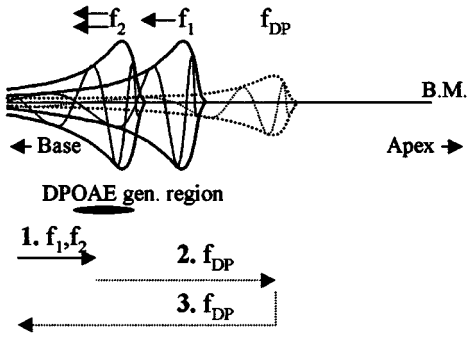
The analysis method has successfully uncovered the full extent of the two modes of emission. The basic features seen in the unprocessed data are preserved and the continuity between the upper and lower sidebands remains. However, there is some evidence of small artefacts being generated at wide ratios and high frequency, where some DPOAE is seen after windowing when none was present beforehand. Such ''blurring'' of the data is a result of applying short windows in the time domain. The possibility of random noise arising from the whole sweep is discussed later with respect to the significance of the low latency $2f_2-f_1$ emission data.

A. Possible general explanation for the high latency DPOAE phase behavior (place fixed mechanism)

The following arguments consider the case of a constant f_{DP} sweep. It is helpful to consider a constant DP frequency sweep when considering place fixed DPOAE phase, as the presumed reflection site does not move, leaving only the changing phases of the inward and outward journeys to cause a DPOAE phase gradient. Such a sweep corresponds to vertical slices of the results in Figs. 1, 3, and 4. The descriptions below consider only the case of a place fixed emission.



a) Place fixed upper sideband DP (e.g. $2f_2$-f_1)



b) Place fixed lower sideband DP (e.g. $2f_1$-f_2)

FIG. 6. Illustrations of the path taken by place fixed DPOAE, considering a constant DP frequency sweep in which the peaks of the f_1 and f_2 traveling waves expand away from the DP frequency place as the frequency ratio increases. (a) Upper sideband DPOAE, and (b) lower sideband DPOAE.

1. Upper sideband DP (e.g., 2 f_2-f_1)

DP generation and emission are from the region of f_{DP} [Fig. 6(a)].

(1) In a constant DP frequency sweep, as the ratio f_2/f_1 is increased, the f_1 and f_2 places move apically away from the f_{DP} place, so the phase lag of both $f_1(\phi_1)$ and $f_2(\phi_2)$ at the DP generating site must decrease. The f_1 frequency place moves away at twice the rate of f_2. However, the phase of the $2f_2-f_1$ DP as it is generated changes twice as much in response to changes in ϕ_2 as to changes in ϕ_1. The phase gradient of ϕ_1 is more gradual than ϕ_2 at the DP generating site. Therefore there are opposing factors influencing the DP phase which means that the DP phase at its generation site will not change much with changing f_2/f_1.
(2) The phase change resulting from the return journey is unchanged with changing ratio. Therefore the return journey doesn't contribute to the phase gradient.

As the upper sideband experimental data has an almost flat phase gradient when DP frequency is fixed (a vertical section through Figs. 3 and 4, parts c and d), the opposing factors described appear to be roughly equal (and opposite) in magnitude.

The argument above suggests that any offset from flat phase (vertical phase contours on the charts) is associated with the difference in the gradients of ϕ_1 and ϕ_2 at the DP generating site.

2. Lower sideband DP (e.g., 2 f_1-f_2)

DP generation is in the region of f_2, but the emission travels via the f_{DP} region [Fig. 6(b)].

(1) In a constant DP frequency sweep, the phase lag of $f_1(\phi_1)$ at the DP generating site decreases with increasing ratio, whereas the phase of $f_2(\phi_2)$ at the DP generating site remains largely unchanged because the site moves with the f_2 place. Therefore the phase lag of the DP as it is generated decreases by $2\phi_1$.
(2) As the ratio increases and the DP generating site moves basally away from the DP place, the phase lag of the journey onward to the DP place increases. However, of necessity the phase gradient of f_{DP} at the f_2 place is less than the f_1 phase gradient at the f_2 place. Also, in point (1) above, the DP phase change due to the change in ϕ_1 was multiplied by 2. Therefore the change in DP phase due to (1) will be greater than the opposing change in (2).
(3) The phase of the return journey is not affected by the stimulus frequency ratio. Therefore (3) does not contribute to the DPOAE phase gradient.

The factors involved here (1, 2, and 3) do not cancel, so there is a net reduction in DP phase lag as the frequency ratio increases with DP frequency constant, contrary to the upper sideband. This is consistent with the lower sideband experimental data when the frequency ratio exceeded about 1.1.

However, it has been shown that the lower sideband high latency phase does not follow this pattern at the smallest frequency ratios, indicating a divergence from the argument above. The phase is continuous between the lower sideband and the upper sideband, suggesting continuity of mechanism. For the upper sideband the DPOAE generating region is associated with the DP frequency place, which is basal to the f_2 place. So for the lower sideband DPOAE at close ratios, where f_2 and f_{DP} become close, the generating site may also be more associated with the f_{DP} site than with the f_2 site. Alternatively, the lower sideband DP generating region may simply be changing shape when f_2/f_1 is small. Changes in the phase gradients within the DP generating region (discussed by Kemp and Knight, 1999) may be significant. This behavior would be consistent with a tendency for DPs to always be generated somewhat basal to the f_2 peak.

B. Possible explanations for the low latency 2 f_2-f_1 emission

As a secondary observation, a small low latency upper sideband DPOAE has been observed in one of the two subjects. The low latency emission, which was seen in the $2f_2-f_1$ DP from subject RN, is of interest as it has not previously been reported in humans. The level of the emission was typically -20 to -5 dB SPL, whereas the high latency emission was typically 0 to 15 dB SPL. The low latency upper sideband emission has a slight and consistent depen-

TABLE I. Mean noise level estimates, dB, SPL, obtained from the normalized time domain responses of the $2f_2-f_1$ DP with $f_2/f_1=1.03$ and 1.1. The ''short window'' was employed to isolate the low latency emission, the ''long window'' was used for the high latency emission.

	RDK	RN
Short window	−30.4	−29.3
Long window	−26.2	−24.6

dence of phase on the frequency ratio but little dependence on DP frequency. This is consistent with a low latency wave fixed emission, so the emission may be from the DP frequency region of the cochlea.

Where is the noise floor?

As the data have been considerably processed, it is difficult to directly infer a noise floor for individual data points. In order to obtain an estimate of where the noise floor lies after the DPOAE separation processing, ''pseudo-time'' windows were drawn from the data at a latency well after the DPOAE had died away. These samples of data were then subjected to a forward Fourier transform to provide an estimate of the noise level in the frequency domain. The average across the frequency range of the noise level estimate was obtained for the $2f_2-f_1$DP with $f_2/f_1=1.03$ and 1.1 (Table I).

In each example, the noise estimate occasionally exceeded −20 dB SPL (exceptionally up to −13 dB SPL), so small regions of data points exceeding −20 dB SPL can be explained by random noise (e.g., the ''low latency'' $2f_2-f_1$ data of RDK). However, the ''low latency'' $2f_2-f_1$ data of RN consistently exceeds this level by 5–15 dB and so this cannot be explained by random noise.

C. How typical are the results?

Regarding the general features of the separated DPOAE, both subjects follow the same patterns, despite having different underlying latencies and DP levels. The general patterns can be accounted for on the basis of the description given in Sec. IV A. In both subjects the wave fixed emission was only strong when f_2/f_1 was between 1.1 and 1.3. The peak occurred at around $f_2/f_1=1.2-1.25$, which is consistent with other data (Harris *et al.*, 1989; Knight and Kemp, 1999; Moulin, 2000). It is likely, therefore, that the general features are typical of normal-hearing ears. The more localized features are not correlated between the two subjects and it seems that these will be different for each subject.

The balance between the levels of the wave and place fixed emissions with $L_1=L_2=70$ dB SPL may be different with other stimulus levels but the overall pattern of behavior is similar (Knight and Kemp, 1999). Use of stimuli with $L_1>L_2$ can be expected to favor the lower sideband (Kemp, 1987).

The possibility has arisen of a ''wave fixed'' emission being present in the $2f_2-f_1$ data, approximately 20 dB SPL lower in level than the place fixed emission. It is not known whether this will emerge as a ''normal feature'' of normally hearing ears as it has only been successfully identified as being above noise in 1 of the 2 subjects.

V. FURTHER ASSESSMENT OF THE TWO COMPONENT DP PHASE HYPOTHESIS

We have demonstrated the separation of wave fixed and place fixed DPOAE in humans using a frequency domain editing method. The distinctive features that have been observed in the phase behavior of DPOAEs appear consistent with the qualitative explanation offered, which was based on the general characteristics of the cochlear traveling wave and the anticipated effect of nonlinearity and reflectors.

In this section we demonstrate the appropriateness of our general explanation by deriving the main experimentally observed DPOAE phase patterns from a basic computational model. We constructed a simple transmission line model of traveling wave interactions and used this to generate maps of backward-traveling DP amplitude and phase maps to compare with the experimental DPOAE data. A region of DP generation was introduced, associated with nonlinearity of a localized region of gain leading up to the traveling wave peaks. The intensity and phase of the summated DPOAE propagating back to the input of the model (middle ear) was computed as a function of primary stimulus frequency and frequency ratio. We then introduced a series of partial reflectors into the transmission line and computed the phase and intensity of DPOAE that these reflectors caused to be propagated back to the input. We reasoned that if the model reproduced the main features seen in the experimental results in a manner not critically dependant on the choice of parameters, it would support the general explanation offered and its underlying assumptions.

Signal propagation within our frequency domain transmission line model is based on a propagation phase curve similar to that found in basilar membrane observations. It has an exponential acceleration of phase accumulation with distance and a logarithmic spacing of characteristic frequencies. Transmission characteristics, which include frequency dependent gain and damping functions, are designed to produce traveling wave envelopes with growth with distance up to a peak, followed by a rapid decline. A nonlinear gain function is introduced which achieved compressive growth with input intensity at the traveling wave peak but linear growth further toward the start of the transmission line and DP generation. Details of the model are given in Appendices A and B.

In our model the DP can propagate in both directions, the DP generated at each model element being vectorally added to the propagating DP traveling wave as it passes. This differs from the approach used by Sun *et al.* (1994) who summed the DP sources to the point of maximum overlap between the stimulus tones and then radiated the DP as a point source from there. Transmission of the DP from the DP generating region, propagated directly back to the start of the transmission line, was taken to represent ''wave fixed'' DPOAE generation. These data were generated over a range of stimulus frequencies, comparable to that used to obtain the experimental data.

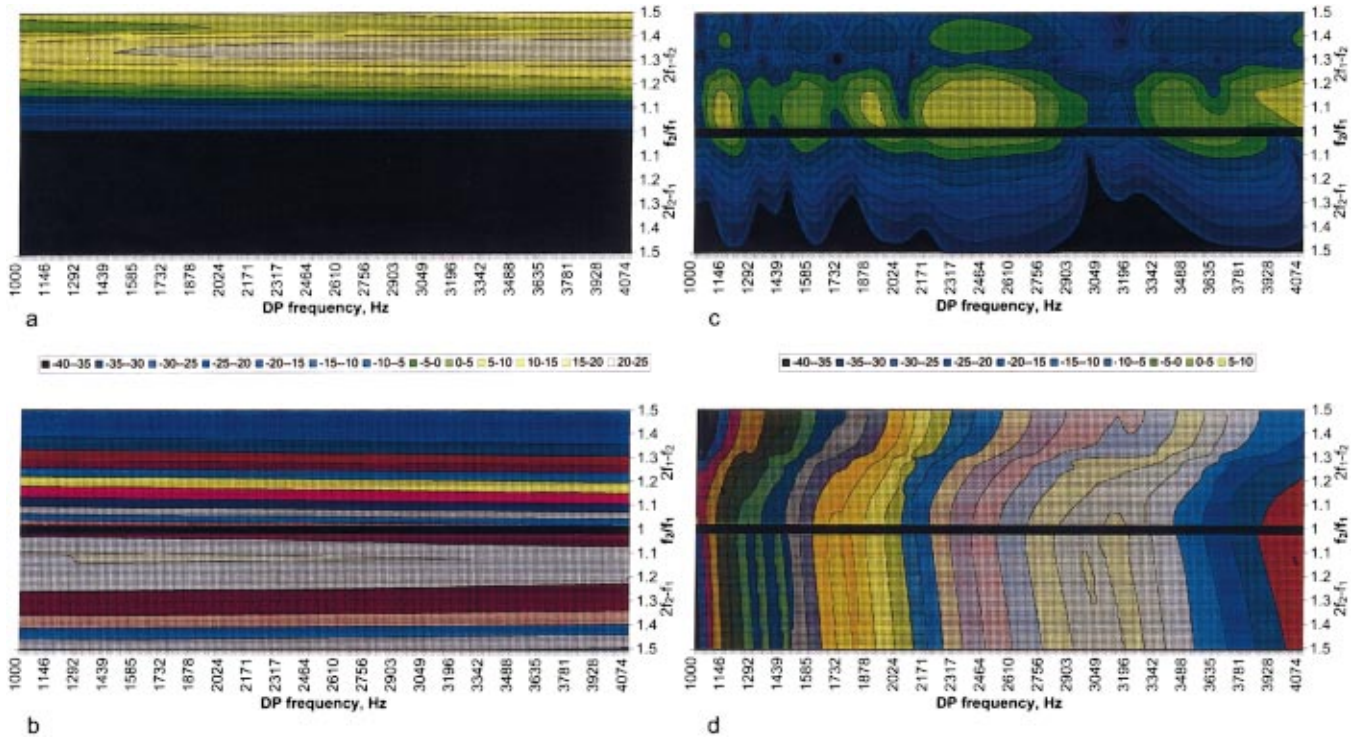


FIG. 7. Wave (a and b) and place (c and d) fixed DP from a transmission line model. The overall level and phase patterns are similar to those seen in measured DPOAE.

The phase and amplitude of this direct component is presented in Fig. 7 (a and b) in a similar format to the DPOAE data of Figs. 3 and 4 (a and b). The ''directly emitted'' data reproduce both the wave fixed phase pattern and the general dependence of level on the ratio f_2/f_1. The dependence on f_2/f_1 appeared to be determined by the vector summation of the DP across its generating region. For the $2f_2-f_1$ DP component, this was largely self-cancelling in the basal direction for all f_2/f_1. With the $2f_1-f_2$DP, cancellation was significant only when f_2/f_1 was small. The directly emitted model DP peaked at a wider ratio than in the experimental wave fixed DPOAE data [Figs. 3 and 4 (a and b)]. This difference is likely to be due to the precise shape of the phase curve, specifically the phase gradients of f_1 and f_2 in the region at which the DP is generated.

Introduction of reflectors

In stage two of our model the onward-propagating DP signal encounters fixed, randomly spaced partial reflectors. The reflection coefficient has been made a function of the level of gain present so that significant reflection occurs at the DP frequency only in the region leading up to and at the DP frequency place. Summation of these reflected DP signals in the reverse direction gives rise to a DP component at the input which is taken to represent ''place fixed'' DPOAE generation.

Figure 7 (c and d) indicates the level and phase of the reflected DP generated by the model. The general patterns are very similar to the experimental DPOAE data in Figs. 3 and 4 (c and d). The model phase data follow the general pattern of actual DP data, with the upper sideband phase contours being slightly off vertical and the lower sideband contours being curved and deviating more strongly from vertical. The model phase data also reproduce the notchy level characteristic of experimentally obtained place fixed DPOAE, and this originates in interference between contributions from the random distribution of reflectors.

In the model data the place fixed DPOAE intensity is greater in the lower sideband than the upper sideband, whereas in the experimental data it is the other way round. The model DP level also falls more rapidly at wider ratios in the upper sideband in the model data compared to the experimental data. These features are affected by the parameters of the model such as the shape of the stimulus phase curves in the region of DP generation, the location of the gain, and the input/output shape of the nonlinearity.

There is a consistent notch in the lower sideband place fixed model DP level at a ratio of approximately $f_2/f_1 = 1.3$. Detailed inspection of spatial distribution of signal levels in the model indicated that this is caused by cancellation across the DP generation region in the forward direction. This feature is mostly not present in the experimental DPOAE data, which may be because the phase gradients and DP generating regions are less consistent in a real cochlea so the ''notch'' does not always occur at the same frequency ratio so is less obvious. Alternatively the notch may occur at a wider frequency ratio in experimental data at which the place fixed DPOAE has a small amplitude, making the notch less noticeable.

We have extensively varied the parameters and details of the formulation within reasonable limits and found the general features to be robust and comparable with the experimental data. It should be noted that our model is not a co-

chlear model but a means of computing the consequences of the various assumptions and simplifications made in developing a qualitative explanation of observed DPOAE behavior. The models capacity to simulate realistic DPOAE maps serves only to support the authors deductions from their qualitative explanation of the phenomenon observed. It would not be appropriate to use the model to make specific predictions. However, the model does support the contention that DPOAE behavior can be simply understood to a first approximation and the qualitative model is useful for this purpose.

VI. SUMMARY AND CONCLUSIONS

Separate wave and place fixed DPOAE maps with DP frequency and f_2/f_1 have been derived for two human ears using a phase gradient based method of separation. The wave fixed output is maximal over a restricted f_2/f_1 range in the lower sideband (e.g., $2f_1-f_2$), but it remains present at a wide range of f_2/f_1 at much reduced levels. We found slight evidence that a very small wave fixed upper sideband (e.g., $2f_2-f_1$) component may be present but if present is less dependent on f_2/f_1. Unlike the lower sideband, its source may be in the region of the DP frequency place in the cochlea.

The place fixed emission has been found at most stimulus frequency combinations, including the entire upper sideband region and the region in the lower sideband where it is normally obscured by the larger wave fixed component. The place fixed emission level was greatest at the smallest frequency ratios.

The phase of the place fixed DPOAE supports the common view that the source of DP contributing to upper sideband DPOAE is a region related to the DP frequency place for all f_2/f_1 ratios. For lower sideband DPOAE the initial DP source must be related to the f_2 peak excitation region when f_2/f_1 is large. However, the excellent continuity of phase behavior between the lower and upper sidebands suggests that when f_2/f_1 is small the lower sideband place fixed DPOAE is a continuation of the upper sideband DPOAE with its source region associated with the DP frequency place.

The observed result can be qualitatively accounted for on the basis of some reasonable assumptions about traveling wave behavior in the cochlea and the mechanical nonlinear interaction which must take place. These assumptions have been formulated into a transmission line model consisting of a wave fixed DP source in the region of the primaries and randomly spaced place fixed reflections in the region of the DP frequency place. Good agreement between the model and experimental data could be achieved, indicating that the observed DP behavior can be simply understood. Differences between the model and experimental data could probably be further reduced by modifications to the modeling of the phase curve of the traveling wave and to the nonlinear interactions in the region of the stimulus tone peaks but this would add little to our main finding.

The simple model used here is not a physical model. Instead, ''cochlearlike'' transmission characteristics and compressive nonlinearity were built into the model which was then used to predict the reverse propagation of DPOAEs. DPOAE generation inherently resulted in a resultant reverse traveling wave with very similar phase and amplitude characteristics to the ''wave fixed'' experimental data. It was necessary to insert ''reflectors'' in order to simulate the place fixed DPOAE map. We used a relatively small number of discrete reflectors active over the region of significant gain. The detailed reflection model of Zweig and Shera (1995) demonstrates that even if an extensive random distribution had been used, complete cancellation among the reflectors would not occur so that a reflection is always to be expected from such an arrangement. We predict that other more physical cochlear models, which include a place fixed reflection mechanism most effective in the region leading up to the DP frequency peak, would also reproduce the experimental level and phase DPOAE characteristics presented in this study.

In conclusion, this study has demonstrated that the apparently complex behavior of DPOAE phase when primary frequencies are altered can be readily understood and even predicted on the basis of a small number of reasonable assumptions about the cochlea. Furthermore it has demonstrated that the wave and place fixed mechanisms of distortion product emission are generally active over a wider range of stimulation conditions than was indicated by Kemp (1986). Kemp presented data suggesting a bimodal generating mechanism, whereas data presented here demonstrate that both emission modes are simultaneously active, at least for the lower sideband DPs. Our analysis does not support the concept of two DP generator regions but does clearly illustrate that interference across a single distributed DP source controls both the observed primary ratio dependence of wave fixed DPOAEs and also the input to the reflection mechanism which is responsible for the place fixed DPOAE. When they are mixed, interference between direct and indirectly emitted DPOAEs results in additional DPOAE fine structure.

We believe the careful separation of DPOAE components in experimental data, as demonstrated here, simplifies the task of extracting consistent physiologically significant information from DPOAE data by removing interference between wave and place fixed emission. However, we are of the opinion that the two emission modes complement each other and that neither one should be studied in isolation.

ACKNOWLEDGMENTS

This project was financially supported by the Auditory Biophysics Lab. at the Institute of Laryngology and Otology. ILO 88 and ILO 96 equipment was provided by Otodynamics Ltd. A preliminary account of parts of this work has been presented at the ARO conference (Knight and Kemp, 2000b).

APPENDIX A: DESCRIPTION OF MODEL

The phase lag from the input to the model is calculated for each stimulus frequency at each incremental distance along the transmission line. For each model element a parameter determines the phase characteristic at that element

for each frequency. This is termed the ''frequency parameter.'' The ''frequency parameter'' for each element of the model is set up in a logarithmic relationship covering frequencies from 25 kHz down to 600 Hz. The incremental change between each model element is such that each element approximately represents two hair cells. The phase lag from the base of the cochlea is calculated for each stimulus frequency at each incremental distance along the basilar membrane. The phase lag is logarithmically related to the frequency parameter [see Eqs. (B1) and (B2) in Appendix B].

In this formulation, gain is necessary in the model so that the traveling waves grow to form sharp peaks (Neely and Kim, 1986; Geisler, 1993; de Boer, 1983a, b). The gain is configured such that the total maximum linear gain along the basilar membrane (which would be delivered if the traveling wave amplitude was small) increases the amplitude of the traveling wave peak by 25 dB. This value is in comparison to the model operating without gain—i.e., the comparison is to a passive and damped model. The gain is frequency dependent and therefore restricted to a region which we have centred just to the basal side of the traveling wave peak. The gain window covers approximately one-third of an octave (-6 dB) of place frequencies in the model. The gain delivered by each model element is reduced by a factor related to the total stimulus travelling wave amplitude (f_1+f_2) at that point (this also includes self-suppression in the case of the propagating DP). Suppression by the small DP on the stimuli is therefore presumed to be not significant [see Eqs. (B3)–(B5) in Appendix B].

The model included gain only for forward-propagating waves. Inclusion of gain for reverse-propagating waves in this model would have resulted in considerable undesirable growth of place fixed DPOAE in the lower sideband at wide frequency ratios. This was because the DP gain region was suppressed by the stimulus tones when f_2/f_1 was small but not when f_2/f_1 was large, so the place fixed DP would be amplified twice by the gain region at large frequency ratios (as it traveled through the gain region both ways).

The amplitude of the propagating traveling wave is diminished by a damping function. Damping is applied as a function of the phase gradient with a constant added. The phase dependent part controls the sharpness and the location of the apical end of the traveling wave envelope. The additional constant tends to control the general growth of the traveling wave envelope basal to the traveling wave peak, outside of the gain region [Eq. (B6) in Appendix B].

DP levels were obtained by applying the computed levels of f_1 and f_2 to the nonlinear input/output function. The model assumes that a fixed proportion of the DP generated is converted into a traveling wave and therefore that a fixed proportion of the DP is either absorbed or directed in a mode which does not readily set up a traveling wave. The phase of the DP generated at each incremental point along the basilar membrane is defined by the phase of f_1 and f_2 at that point.

The model supports propagation of the DP in both directions along the basilar membrane. The DP generated at each model element is vectorally added to the propagating traveling wave as it passes.

The reverse-traveling part provided the ''wave fixed'' DPOAE component (path 1 described in our Introduction). The onward-propagating component results in a traveling wave which proceeds to the DP frequency place. In order that a DPOAE results from this component, randomly spaced place fixed reflectors have been introduced. Each model element had a 1 in 5 chance of being a reflector, except that adjacent reflectors were not allowed. The reflector strength is linked to the linear gain window associated with f_{DP}—therefore the significant reflectors are limited to the region leading up to the DP traveling wave peak. The reflection is defined by a proportion of the incident traveling wave, rather than being defined by an impedance irregularity.

The reflected DP propagates back toward the start of the model with contributions from each reflector being added to the traveling wave as it passes. The reflected reverse-propagating traveling wave is kept numerically separate from the reverse propagating ''wave fixed'' DP to avoid the difficulties of separating them later. The model does not include reflections back into the cochlea from the stapes, so repeated reflections between the base and the DP frequency place are not supported. Essentially the same procedure is performed to obtain $2f_2-f_1$ DPOAEs, but the model produces very little $2f_2-f_1$ DPOAE via the wave fixed mode. For place fixed DPOAE, the reflectors placed around the DP frequency place were also in the DP generation region so that the distance of apical DP travel before a reflector was encountered was minimal.

APPENDIX B

The main formulas for the model are listed below.

Frequency parameter array

For the purpose of this model, an array of ''frequency parameters'' was ascribed to the model elements. A phase delay of three cycles was incurred by a stimulus in traveling to the place of its corresponding ''frequency parameter.'' At a stimulus level of 70 dB SPL, the traveling wave amplitude peaked at the appropriate ''frequency parameter.'' For lower stimulus levels, the traveling wave peak would be slightly beyond the ''frequency parameter.''

The ''frequency parameter'' of each model element is calculated as follows:

$$\text{Frequency parameter } (f_p) = 25\,000 \times 0.995\,595^{(n-1)}, \tag{B1}$$

where n = the number of the model element. The ''frequency parameter'' is used in the calculation of the phase gradients and the gain characteristics. The gain and damping is balanced such that, at 70 dB SPL, the traveling wave envelope peaks at the frequency parameter corresponding to the stimulus frequency.

Phase curve

The phase curve is:

$$\text{Phase lag (radians)} = \frac{6\pi\{(e^{(2f/f_p)}) - (e^{(2f/25\,111)})\}}{2.3e}, \tag{B2}$$

where f = the stimulus frequency of the traveling wave, and f_p = frequency parameter. The second part of the numerator ensures that the phase lag at the first model element is 0. Where this term is ≈ 0, there are three phase cycles to the frequency place. With high frequencies, the number of phase cycles reduces slightly because part of the phase curve is ''lost'' at the basal end of the model. The phase gradient at the peak of the traveling wave is the same for all frequencies.

Conversion of stimulus level to traveling wave amplitude

The stimulus level is entered into the model in dB SPL, which is converted to pascals. In order to obtain the initial amplitude of the traveling wave the sound pressure in pascals is multiplied by the phase difference between the first two model elements for that frequency and arbitrarily by 100. This calculation is reversed for the 'reverse propagating DP' when it arrives back at the start of the model to obtain a DP level in dB SPL. The traveling wave amplitude is of arbitrary units.

Gain

The nonlinear gain is defined as follows:

$$G = 1 + (g_L / C), \tag{B3}$$

where g_L defines the place envelope of the maximum gain (the gain which would be obtained if the traveling wave amplitude were small). C is a compression function which reduces the gain which is achieved depending on the traveling wave amplitude.
g_L is calculated as follows:

$$g_L = \frac{(0.12(1.1f/f_p)^{10})}{(1 + (1.1f/f_p)^{25})}, \tag{B4}$$

where f = traveling wave frequency, and f_p = frequency parameter. The gain is therefore restricted to a region centered just before the traveling wave peak:

$$C = (3L_T)^{1.2} + 1, \tag{B5}$$

where L_T = total apical-going traveling wave amplitude of f_1 and f_2 (for suppression of DP gain, self-suppression by f_{dp} is also included).

Damping

The damping is based on the traveling wave phase gradient and is therefore proportional to the time spent at that element, with a constant added:

$$\text{Damping (D)} = 0.0003(300\Delta\phi)^2 + 0.001, \tag{B6}$$

where $\Delta\phi$ = phase change between adjacent model elements. The new traveling wave amplitude after gain and damping is then as follows:

$$\text{New amplitude} = \frac{A_0 G}{1 + D}, \tag{B7}$$

where A_0 = traveling wave amplitude input to the present model element, G = nonlinear gain, and D = damping. In each model element the traveling wave is modified by functions of the gain, damping, and phase gradient. The amplitude is divided by the present phase gradient and then multiplied by the next phase gradient in order to re-scale it for the incremental phase gradient change.

DP generation formula level/phase

The generation of DP is calculated by applying the nonlinear gain to the waveform of L_1 and L_2 at each model element. For computational convenience, a table of values was generated and from this an approximate function was created which was used in the model. The phase lag of the DP at each point at which it is generated is $2\phi_1 - \phi_2$ in the case of the $2f_1 - f_2$ DP, or $2\phi_2 - \phi_1$ in the case of the $2f_2 - f_1$ DP, where ϕ_1 and ϕ_2 are the phase lags of f_1 and f_2 at the point of DP generation.

Reflections

Each model element is given a reflectivity factor equal to zero or to the frequency dependent gain multiplied by a constant. Each element has a 1 in 5 random chance of being a reflector, except adjacent reflectors were not allowed. The envelope of the reflections of the fixed reflectors on the basilar membrane is therefore linked to the linear gain window related to the DP frequency. Typically there may be around 10 reflectors within the -6 dB limits of the gain region.

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.

de Boer, E. (**1983a**). ''No sharpening? A challenge for cochlear mechanics,'' J. Acoust. Soc. Am. **73**, 567–573.

de Boer, E. (**1983b**). ''On active and passive cochlear models—Toward a generalized analysis,'' J. Acoust. Soc. Am. **73**, 574–576.

Dreisbach, L. E., and Seigel, J. H. (**1999**). ''Level and phase relationships of distortion-product otoacoustic emission sources with varied primary frequency ratios in humans,'' Assoc. Res. Otorhinolaryngol. Abstracts **22**, 392.

Fahey, P. F., and Allen, J. B. (**1997**). ''Measurement of distortion product phase in the ear canal of the cat,'' J. Acoust. Soc. Am. **102**, 2880–2891.

Fahey, P. F., Stagner, B. B., Lonsbury-Martin, B. L., and Martin, G. K. (**2000**). ''Theory for two new routes of distortion product generation,'' Assoc. Res. Otorhinolaryngol. Abstracts **23**, 478.

Gaskill, S. A., and Brown, A. M. (**1996**). ''Suppression of human acoustic distortion product: Dual origin of $2f1-f2$,'' J. Acoust. Soc. Am. **100**, 3268–3274.

Geisler, C. D. (**1993**). ''A realizable cochlear model using feedback from motile outer hair cells,'' Hear. Res. **68**, 253–262.

Harris, F. P., Lonsbury-Martin, B. L., Stagner, B. B., Coats, A. C., and Martin, G. K. (**1989**). ''Acoustic distortion products in humans: systematic changes in amplitude as a function of f_2/f_1 ratio,'' J. Acoust. Soc. Am. **85**, 220–229.

Heitmann, J., Waldmann, B., Schnitzler, H., Plinkert, P. K., and Zenner, H. (**1998**). ''Suppression of distortion product otoacoustic emissions (DPOAE) near $2f1-f2$ removes DP-gram fine structure—Evidence for a secondary generator,'' J. Acoust. Soc. Am. **103**, 1527–1531.

Kalluri, R., and Shera, C. (**2000**). ''Are DPOAEs a mixture of emissions generated by different mechanisms?'' Assoc. Res. Otorhinolaryngol. Abstracts **23**, 480.

Kemp, D. T. (**1978**). ''Stimulated acoustic emissions from within the human auditory system,'' J. Acoust. Soc. Am. **64**, 1386–1391.

Kemp, D. T. (**1980**). ''Toward a model for the origin of cochlear echoes,'' Hear. Res. **2**, 533–548.

Kemp, D. T. (**1986**). ''Otoacoustic emissions, travelling waves and cochlear mechanisms,'' Hear. Res. **22**, 95–104.

Kemp, D. T. (**1987**). ''Otoacoustic emission characteristics during moderate continuous stimulation-intermodulation,'' 12th International Congress of Acoustics B7-1.

Kemp, D. T. (**1998**). ''Otoacoustic emissions: Distorted echoes of the cochlea's travelling wave,'' in *Otoacoustic Emissions Basic Science and Clinical Applications*, edited by C. I. Berlin (Singular Publishing Group, San Diego), Chap. 1, pp. 1–59, Fig. 1-12.

Kemp, D. T., and Brown, A. M. (**1983**). ''An integrated view of cochlear mechanical nonlinearities observable from the ear canal,'' in *Cochlear Mechanics*, edited by E. DeBoer and M. A. Viergever (Delft University Press, Holland).

Kemp, D. T., and Knight, R. D. (**1999**). ''Virtual DP reflector explains DPOAE wave and place fixed dichotomy,'' Association for Research in Otolaryngology Abstracts, 396.

Knight, R. D., and Kemp, D. T. (**1999**). ''Relationships between DPOAE and TEOAE characteristics,'' J. Acoust. Soc. Am. **106**, 1420–1435.

Knight, R. D., and Kemp, D. T. (**2000a**). ''Indications of different DPOAE mechanisms from a detailed $f1$, $f2$ area study,'' J. Acoust. Soc. Am. **107**, 457–473.

Knight, R. D., and Kemp, D. T. (**2000b**). ''Separation of ''wave'' and ''place'' fixed $2f1-f2$ DPOAE,'' Assoc. Res. Otorhinolaryngol. Abstracts **23**, 987.

Kummer, P., Janssen, T., and Arnold, W. (**1995**). ''Suppression tuning characteristics of the $2f1-f2$ distortion-product otoacoustic emission in humans,'' J. Acoust. Soc. Am. **98**, 197–210.

Lonsbury-Martin, B. L., Martin, G. K., Probst, R., and Coats, A. C. (**1988**). ''Spontaneous otoacoustic emissions in a non-human primate. II. Cochlear anatomy,'' Hear. Res. **33**, 69–94.

Martin, G. K., Stagner, B. B., and Lonsbury-Martin, B. L. (**2000**). ''Harmonic otoacoustic emissions in normal and noise-exposed rabbits,'' Assoc. Res. Otorhinolaryngol. Abstracts **23**, 231.

Martin, G. K., Jassir, D., Stagner, B. B., Whitehead, M. L., and Lonsbury-Martin, B. L. (**1998**). ''Locus of generation for the $2f1-f2$ vs $2f2-f1$ distortion-product otoacoustic emissions in normal-hearing humans revealed by suppression tuning, onset latencies, and amplitude corelations,'' J. Acoust. Soc. Am. **103**, 1957–1971.

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., Stagner, B. B., Jassir, D., Telischi, F. F., and Lonsbury-Martin, B. L. (**1999**). ''Suppression and Enhancement of distortion-product otoacoustic emissions by interference tones above $f2$. I. Basic findings in rabbits,'' Hear. Res. **136**, 105–123.

Mauermann, M., Uppenkamp, S., van Hengel, P. W., and Kollmeier, B. (**1999a**). ''Evidence for the distortion product frequency 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**, 3473–3482.

Mauermann, M., Uppenkamp, S., van Hengel, P. W., and Kollmeier, B. (**1999b**). ''Evidence for the distortion product frequency 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**, 3473–3482.

Moulin, A. (**2000**). ''Influence of primary frequencies ratio on distortion product otoacoustic emissions amplitude. I. Intersubject variability and consequences on the DPOAE-gram,'' J. Acoust. Soc. Am. **107**, 1460–1470.

Neely, S. T., and Kim, D. O. (**1986**). ''A model for active elements in cochlear biomechanics,'' J. Acoust. Soc. Am. **79**, 1472–1480.

Shera, C. A., and Guinan, J. J. (**1999**). ''Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs,'' J. Acoust. Soc. Am. **105**, 782–798.

Shera, C. A., and Zweig, G. (**1993**). ''Noninvasive measurement of the cochlear traveling-wave ratio,'' J. Acoust. Soc. Am. **93**, 3333–3352.

Siegel, J., Borneman, A., and Dreisbach, L. (**2000**). ''Suppressor conditions for optimal separation of distortion product otoacoustic emission sources,'' Assoc. Res. Otorhinolaryngol. Abstracts **23**, 983.

Stover, L. J., Neely, S. T., and Gorga, M. P. (**1996**). ''Latency and multiple sources of distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **99**, 1016–1024.

Sun, X., Schmiedt, R. A., He, N., and Lam, C. F. (**1994**). ''Modeling the fine structure of the $2f1-f2$ acoustic distortion product. I. Model development,'' J. Acoust. Soc. Am. **96**, 2166–2174.

Talmadge, C. L., Long, G. R., 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 Tong, C. (**2000**). ''Modeling the combined effects of basilar membrane nonlinearity and roughness on stimulus frequency otoacoustic emission fine structure,'' J. Acoust. Soc. Am. **108**, 2911–2932.

Tubis, A., Talmadge, C., Long, G. R., Dhar, S., and Tong, C. (**2000**). ''Amplitude and group-delay fine structures of distortion product otoacoustic emissions as functions of primary levels and frequency ratios,'' Assoc. Res. Otorhinolaryngol. Abstracts **23**, 483.

Zweig, G., and Shera, C. A. (**1995**). ''The origin of periodicity in the spectrum of evoked otoacoustic emissions,'' J. Acoust. Soc. Am. **98**, 2018–2047.

Effects of stimulus frequency and complexity on the mismatch negativity and other components of the cortical auditory-evoked potential[a)]

Julia L. Wunderlich[b)] and Barbara K. Cone-Wesson
Audiology and Speech Sciences Unit, Department of Otolaryngology, The University of Melbourne, 384-388 Albert Street, East Melbourne, 3002 Victoria, Australia

(Received 19 June 2000; accepted for publication 20 December 2000)

This study investigated, first, the effect of stimulus frequency on mismatch negativity (MMN), N_1, and P_2 components of the cortical auditory event-related potential (ERP) evoked during passive listening to an oddball sequence. The hypothesis was that these components would show frequency-related changes, reflected in their latency and magnitude. Second, the effect of stimulus complexity on those same ERPs was investigated using words and consonant–vowel tokens (CVs) discriminated on the basis of formant change. Twelve normally hearing listeners were tested with tone bursts in the speech frequency range (400/440, 1500/1650, and 3000/3300 Hz), words (/bæd/ vs /dæd/) and CVs (/bæ/ vs /dæ/). N_1 amplitude and latency decreased as frequency increased. P_2 amplitude, but not latency, decreased as frequency increased. Frequency-related changes in MMN were similar to those for N_1, resulting in a larger MMN area to low frequency contrasts. N_1 amplitude and latency for speech sounds were similar to those found for low tones but MMN had a smaller area. Overall, MMN was present in 46%–71% of tests for tone contrasts but for only 25%–32% of speech contrasts. The magnitude of N_1 and MMN for tones appear to be closely related, and both reflect the tonotopicity of the auditory cortex. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1349184]

PACS numbers: 43.64.Qh, 43.64.Ri, 43.64.Sj [LHC]

I. INTRODUCTION

In order to understand spoken language, the listener must be able to discriminate sounds that vary in frequency, intensity, and temporal characteristics. Frequency information is essential for the accurate perception of vowels and consonants, as well as for the recognition of prosodic features which carry meaning, such as intonation and voice pitch (see Moore, 1995, for a review). The importance of frequency information for speech perception is evident in listeners with cochlear hearing loss, whose impaired frequency selectivity and discrimination contribute to difficulty understanding speech, especially in the presence of background noise (Moore, 1996).

Components of the cortical auditory event-related potential (ERP) can be used to investigate neural processes involved in discrimination of pure tones that vary in frequency or complex signals such as speech, that also vary along some acoustic continuum, such as frequency (Näätänen, 1992). In general, ERPs can provide an ''objective'' measure of sound detection and discrimination, and so can be useful in the study of auditory processing and plasticity (Näätänen, 1995) not only among those with normal hearing, but also in the very young or in other listeners in whom psychophysical measures are difficult to obtain (Csépe and Molnár, 1997). The broad aim of the current study was to employ ERPs as a measure of discrimination of simple and complex sounds that varied in frequency, and so document the presence, amplitude, and timing of cortical neural processes that underlie speech perception.

Components of the ERP, N_1 and P_2, have been used to determine sensitivity for tones that vary in frequency (e.g., Antinoro, Skinner, and Jones, 1969; Rapin *et al.*, 1966) and their latency and amplitude also reflect the audibility of speech sounds in bandpassed noise (Martin, Kurtzberg, and Stapells, 1999). The ERP consists of a series of positive and negative peaks starting about 50 ms after the onset of a stimulus and continuing for around 300 ms. The first negative peak, N_1, occurs around 100 ms and is followed by the positive peak, P_2, at 175–200 ms. These potentials have response characteristics which are determined predominantly by stimulus parameters. Parametric studies of the effect of frequency on the amplitude and latency of these potentials show that both amplitude (Antinoro *et al.*, 1969; Jacobson *et al.*, 1992; Pantev *et al.*, 1995) and latency (Alain, Woods, and Covarrubias, 1997; Jacobson *et al.*, 1992) decrease as frequency increases. Some investigators have used speech sounds varied along the acoustic continuum of voice-onset-time (Tremblay *et al.*, 1997) or other timing differences (e.g., ba/wa by Kraus *et al.*, 1999) and reported differential effects on ERP component latency and amplitude. The current study replicates and extends that work, by measuring the latency and amplitude of ERP components in a parametric fashion, as frequency is varied, and in the same subjects, for complex sounds, i.e., speech tokens, that differ with respect to their frequency components.

The MMN is an ERP component evoked using a stimu-

[a)]Portions of this research were presented at the 22nd Midwinter Research Meeting of the Association for Research in Otolaryngology, St. Petersburg Beach, Florida, 13–18 February 1999.

[b)]Electronic mail: j.wunderlich@medoto.unimelb.edu.au

lus sequence in which an acoustically or perceptually deviant stimulus is occasionally presented in a series of homogeneous stimuli. MMN is an increased negativity occurring around 100–250 ms after stimulus onset in the averaged response evoked by the deviant stimulus relative to the response evoked by the standard. The latency and amplitude of these components are determined mainly by the context in which the stimulus occurs, as well as by behavioral factors and task demands (Näätänen and Picton, 1986; Picton *et al.*, 2000). MMN is thought to index the early stages of auditory information processing, short-term auditory memory and attention (Näätänen, 1992), and precognitive sensory discrimination (Kraus *et al.*, 1999).

MMN can be evoked by simple and complex acoustic signals differing in frequency (e.g., Sams *et al.*, 1985), intensity or level (e.g., Schroger, 1996), duration (e.g., Ponton *et al.*, 1997) and spatial location (e.g., Winkler *et al.*, 1998). There has been little systematic investigation of the effect of frequency, as such, on MMN. It is known that MMN can be recorded at, or near, the psychophysical threshold for small changes in frequency around 1000 Hz (Sams *et al.*, 1985; Tiitinen *et al.*, 1994), but it is not known whether this is also the case at other frequencies. It is not known whether the presence, amplitude, or latency of MMN varies with frequency when relative stimulus deviance is kept constant. It might be expected that MMN would show some degree of frequency dependence, since at least some of its neural generators are located in the auditory cortex on the supratemporal plane (Alho *et al.*, 1996; Csépe *et al.*, 1997; Tiitinen *et al.*, 1993) and, according to magnetoencephalographic (MEG) studies, appear to be tonotopically organized (May *et al.*, 1999; Tiitinen *et al.*, 1993). MMN latency and amplitude, however, has been shown to be dependent upon the discriminability of the stimulus contrast, with larger amplitudes and shorter latencies for more easily discriminable contrasts (e.g., Sams *et al.*, 1985). The current investigation tested the hypothesis that MMN presence, latency, amplitude, and area would vary along a continuum of frequency, when the magnitude of the stimulus contrasts was kept constant.

Another area for investigation was the difference between MMN for simple sounds, varying in only one dimension (frequency) versus complex, meaningful stimuli, such as speech. MMN has been evoked by changes in speech parameters such as formant transition (e.g., Maiste *et al.*, 1995; Martin *et al.*, 1997; Sams *et al.*, 1990), formant structure (e.g., Aaltonen *et al.*, 1997; Näätänen *et al.*, 1997), formant duration (Kraus *et al.*, 1999) and voice onset time (e.g., Tremblay *et al.*, 1997). There are inconsistent findings regarding MMN for simple versus complex sounds. Tervaniemi *et al.* (2000) showed that spectral complexity, which facilitates the discrimination of pitch, also results in larger MMN in comparison to those evoked by pure tones. In contrast, Winkler, Tervaniemi, and Näätänen (1997) showed that that complex, nonspeech sounds evoked smaller-amplitude MMN than simple stimuli. In addition, there is evidence that MMN evoked by complex stimuli such as chords (Alho *et al.*, 1996) or vowel contrasts (Aaltonen *et al.*, 1993) involve different parts of the brain than MMN evoked by tones. Vowels and consonant–vowel tokens (CVs) have been used to investigate various aspects of speech perception and learning (e.g., Aaltonen *et al.*, 1997; Kraus *et al.*, 1993a; Martin *et al.*, 1999; Sharma and Dorman, 1998; Tremblay *et al.*, 1997); however, no studies of MMN have made use of words. In this study words, having CVC structure, and CV tokens were used to evoke MMN.

It is of theoretical and practical importance to investigate the properties of MMN as a function of stimulus parameters such as frequency and complexity. MMN has been used in the research laboratory as an indicator of the earliest stages of auditory processing underlying sound perception and discrimination, and, further, as a tool for investigating the neural bases of speech perception. Thus, it is important to show in a detailed fashion that MMN response properties reflect aspects of sound discrimination such as for frequency or for phonemes, in order to understand how MMN reflects auditory information processing. Review of the recent literature suggests that some assumptions about MMN, particularly with regard to speech perception and discrimination, need further investigation, owing to conflicting findings. Specifically, MMN has been assumed to index discrimination of acoustic cues, and not phonemic boundaries (Maiste *et al.*, 1995; Sharma *et al.*, 1993) but newer studies now show MMN variation with phonemic boundaries (Dehaene-Lambertz, 1997; Dehaene-Lambertz and Baillet, 1998; Näätänen *et al.*, 1997; Sharma and Dorman, 1999). Furthermore, the finding that MMN may be present for speech sound contrasts that are not perceptually discriminated (Dalebout and Stack, 1999; Kraus *et al.*, 1999) suggests that the neural processes underlying MMN generation are not simple. The practical application of these MMN studies is to develop tests of sound discrimination that might be used in clinical settings, supplementing or substituting for behavioral measures of perception and discrimination. Whereas the reliability of the MMN has been studied (Escera *et al.*, 2000; Tervaniemi *et al.*, 1999), the reliability of MMN within individuals for different stimulus contrasts has not previously been investigated. Again, knowledge of MMN response parameters, ascertained from individual and group data, is needed to extend the applicability of this technique in research and clinical contexts. Bearing in mind these theoretical and practical imperatives, the first aim of this study was to investigate frequency effects on MMN, N_1, and P_2 components of the ERP with the hypothesis that there would be frequency-related effects. The second aim was to compare tone-evoked responses to those evoked by words and CVs differing only in the place of articulation of the initial consonant.

II. METHOD

A. Participants

There were seven female and five male participants aged between 22 and 45 years (mean 33.6 years), all with normal hearing as defined by pure tone thresholds below 15 dBHL from 500–4000 Hz.

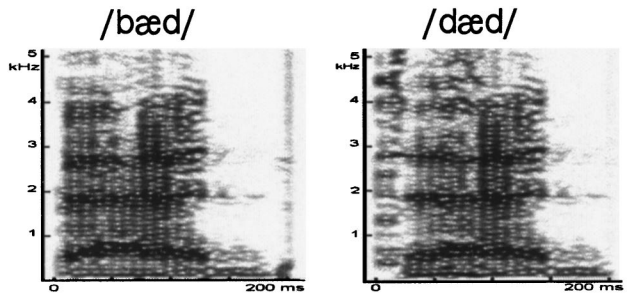


FIG. 1. Spectrograms for the words /bæd/ and /dæd/. Formant frequencies (Hz) were /bæd/: F0 110, F1 673, F2 1769, F3 2677, /dæd/: F0 111, F1 641, F2 1825, F3 2725. The CVs were the same stimuli truncated at 100 ms.

B. Stimuli

Pure-tone and complex stimuli were used to evoke MMN. There were three test conditions contrasting pure tones at frequencies in the low, mid, and high speech-frequency range and two speech test conditions contrasting either words or CVs, which differed by one phoneme. Each participant was tested with all stimuli, with the exception of one participant who was not tested with the CV tokens. All stimuli were 200 ms in total duration except the CVs which were 100 ms; the tones had a 10-ms rise/fall time and were generated by NEUROSCAN® STIM software. The pure-tone conditions contrasted an easily discriminable, 10% change in frequency in either the low (400/440 Hz), mid (1500/1650 Hz), or high (3000/3300 Hz) speech-frequency range. The word condition contrasted ''bad'' /bæd/ with ''dad'' /dæd/, which differ in place of articulation, while the CVs were these words truncated to form /bæ/ and /dæ/. These stimuli were created by Appen Speech Technology, Australia, using a diphone synthesis technique (Isard and Miller, 1986) and the spectrograms for the words are shown in Fig. 1.

All stimuli were presented to the right ear through an EAR® insert earphone at a sensation level of 60 dB. The threshold for each stimulus was determined at the start of the experiment using a standard audiometric threshold seeking technique (Hughson and Westlake, 1944). Stimuli were calibrated using a Brüel & Kjær Modular Precision Sound Analyzer type 2260 with Enhanced Sound Analysis Software BZ 7202, a Brüel & Kjær Artificial Ear 4152, and a 1/2-in. polarizing microphone type 4144.

Stimuli were presented in an oddball sequence, in which the standard stimulus occurred in 84% of presentations and the deviant stimulus in 16% of presentations. The onset- to-onset interstimulus interval (ISI) was 750 ± 50 ms. This slight onset asynchrony was used to decrease temporal probability of both the standard and the deviant stimuli. In order to avoid fatigue, each contrast was tested over three trials; the order in which trials were presented ensured that stimuli from the same condition were never used consecutively. Each trial consisted of 470 stimuli, i.e., 75 deviant stimuli and 395 standard stimuli. Thus, for each stimulus contrast there was a total of 225 deviant stimuli and 1185 standard stimuli presented over the three trials. The first 20 stimuli were always standards; thereafter, the stimuli were presented in a pseudo-random manner so that there were 3–7 standard stimuli between each deviant stimulus. Each stimulus condition was counterbalanced; that is, each stimulus served as a standard sound for three trials, and also as a deviant in another three trials. There was a total of 10 stimulus contrasts (5 stimuli $\times$ 2 tests) for each participant, tested over 30 trials (470 stimulus presentations/trial).

C. Evoked potential recordings

Electrodes were placed at F_z, C_z, P_z, C_4, T_4, C_3, T_3, and left mastoid M_1 with a reference electrode at the right mastoid (M_2) and a ground electrode at F_{pz} (Jasper, 1958). An electrode to record vertical eye movements was placed above the left eye and also referred to M_2. This montage has been found to yield a similar eye-blink waveform to that obtained using a bipolar montage (Kraus *et al.*, 1993b). Recording of the EEG was made between 1 to 100 Hz (24 dB/octave slope) at 2000-Hz sampling rate using a Neuroscan® Synamp amplifier and SCAN 4.0 acquisition software. The recordings were processed off-line to remove eye movement artifacts (Semlitsch *et al.*, 1986); this procedure included reviewing the continuous EEG record, and creating a template of eye-blink artifacts observed in the appropriate recording channel. The EEG was then epoched at 650 ms with 100-ms pre- and 550-ms poststimulus onset. Each epoch was baseline corrected to the prestimulus period, and epochs in which the amplitude of the EEG at either F_z or C_z exceeded ±50 μV were excluded from averaging. After processing there were approximately 210 epochs in response to the deviant and approximately 900 epochs in response to the standard. Averages were computed, as described below, and digitally bandpass filtered between 1–30 Hz (24-dB/octave slope).

Participants were seated comfortably and watched a video of their choice throughout the evoked potentials recording; one participant chose to read. Video sound levels were kept at a low listening level, at least 20 dB below the level of the sound presented to the test ear. Participants were asked to ignore the sounds presented and to minimize movement. Recordings were made over 2–3 test sessions of approximately 2–3 hours each.

D. Behavioral measure of discrimination

On the completion of the evoked potential measurements, each participant performed a behavioral test of stimulus discrimination. Each stimulus pair was presented in a three-alternative, forced-choice test with 750 ± 50 ms between each sound (the same ISI as used in the oddball sequences) and 4 s between trials. There were initially three practice trials and then three test trials for each condition. As expected, all participants were able to discriminate all of the stimuli on all trials ($p<0.05$ of chance performance).

E. Data analysis

Responses evoked by standard and deviant stimuli were averaged (separately) for each of the three trials per contrast and then summed. Since the contrasts were counterbalanced, this resulted in a ''standard'' and ''deviant'' averaged response for each stimulus. The standard average excluded responses to the stimulus which occurred immediately after the

deviant stimulus. Difference waves were calculated by subtracting the response to the stimulus as a standard from the response to the same stimulus as a deviant, thus controlling for acoustic differences. Grand mean standard, deviant, and difference averages were formed using the averaged responses from each individual. Point-to-point *t* tests were used to determine the interval over which the grand mean difference waves were significantly less than zero, i.e., where the deviant average was significantly more negative than the standard average ($p<0.05$, one-tailed) in the manner of Kraus *et al.* (1993b).

All standard and deviant waveforms were initially inspected visually. Baseline-to-peak amplitude and latency measurements were made for ERP components N_1 and P_2. N_1 was defined as the most negative peak occurring 70–180 ms after stimulus onset, and P_2 as the most positive peak between 125–280 ms. In some averages P_2 was below the baseline, i.e., a negative value, in which case the latency of the peak was measured and the amplitude recorded as missing. All measurements reported here are from responses recorded at the F_z electrode since it was at this electrode that MMN was largest (see also Näätänen, 1992).

The statistical technique described by Ponton *et al.* (1997) was adapted to determine the presence of MMN for individual participants. This integral-distribution technique is designed to improve signal-to-noise ratio and provide a means of determining the presence of MMN for individuals without reference to group or normative data. First, subaverages were created for the response to a stimulus when it served as a standard. This was done by pooling all of the accepted standard epochs from the three trials and averaging a random sample of 200. The process was repeated 100 times, always drawing from the total pool of epochs, to produce a group of 100 subaverages to the standard stimulus; all were bandpass filtered between 1–30 Hz. Second, the integral of the waveform was calculated over 100–220 ms for these 100 subaverages. This epoch was chosen as it represented the period in which MMN was evident in the grand mean averages and in which MMN is typically found (Lang *et al.*, 1995). Third, a histogram was created which showed the distribution of the integral values for the response to the standard. Finally, the integral value for the response to the same stimulus when it was serving as a deviant was also calculated and compared to the standard subaverage integral distribution. MMN was defined as present if the integral value for the response to the deviant was below the 10th percentile of the standard distribution.

MMN latency, amplitude, and area measures were made for those tests in which MMN was identified as present according to the integral-distribution test (described above). MMN peak amplitude, measured from baseline to peak, and latency both occurred within the 100–220-ms epoch. MMN onset and offset latency were determined by visual inspection of the difference waves and the MMN area was measured between these latencies. MMN duration was the difference between MMN offset and onset latencies. In order to compare the magnitude of MMN to N_1, the ratio of MMN area to the area of N_1 (between N_1 onset and offset latencies) in the standard average was computed. When additional components, such as P_{3a}, were observed in individual responses, the baseline-to-peak amplitude and latency were also measured.

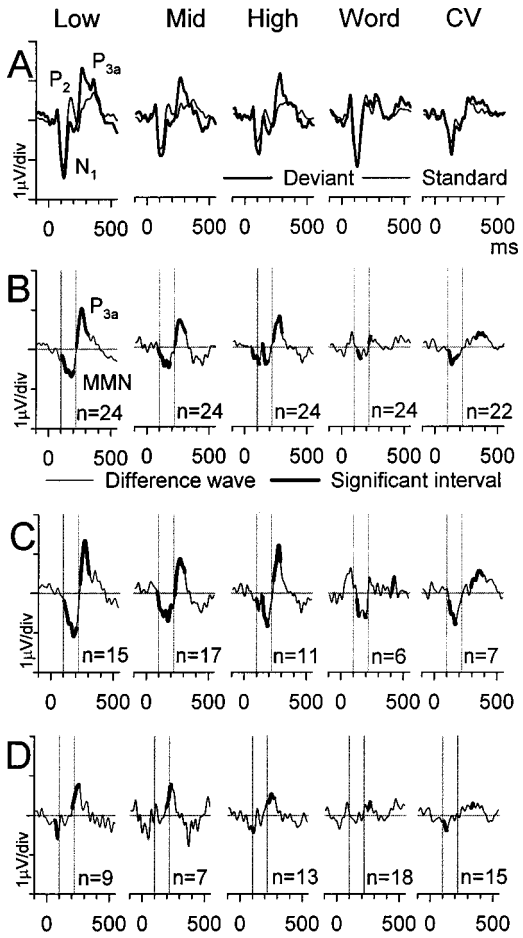


FIG. 2. A. Grand mean averages, standard (thin), deviant (thick) for each condition recorded at F_z. B. Grand mean difference waves (deviant minus standard average) with the thickened portion indicating the interval over which the waveform was significantly below zero (MMN, *t* test $p<0.5$, one-tailed) or above zero (P_{3a}, *t* test $p<0.05$, one-tailed). The interval 100–220 ms within which MMN was measured is marked. C. Grand mean difference waves for participants with MMN present according to the integral-distribution test. D. Grand mean difference waves for participants with MMN absent.

III. RESULTS

A. Stimulus effects on the presence of MMN in individuals

The presence of MMN in individual data was determined by the integral-distribution test (Ponton *et al.*, 1997). The averages which passed this test (MMN present) were grouped and the grand mean average (Fig. 2, panel C) can be compared to the grand mean average of the tests which failed (MMN absent, Fig. 2, panel D). MMN is clearly present in the ''MMN present'' group and clearly absent in the ''MMN absent group,'' which indicates that the integral-distribution test is appropriately discriminating between the two groups and is a valid technique by which to judge the presence of MMN in individual data.

Using the criterion of $p<0.10$ for the integral distribution test, Fig. 3 shows the percentage of tests in individuals

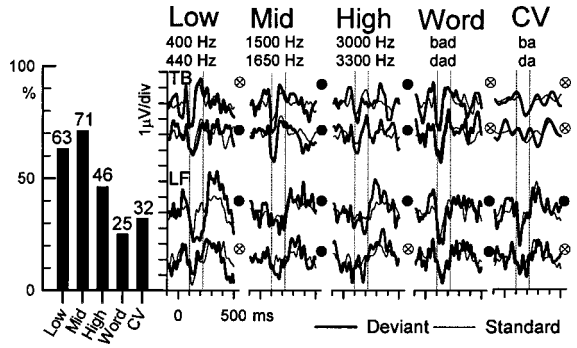


FIG. 3. Percentage of tests with significant MMN present according to the integral-distribution test ($p<0.10$). Typical responses recorded at F_z from two participants are shown. The overlaid averages were evoked by the same stimulus serving as a standard (thin) and as a deviant (thick). The stimulus used to evoke the upper and lower pairs is indicated at the top of each panel. Marks beside each overlaid pair indicate the presence (●) or absence (⊗) of MMN according to the integral-distribution test.

in which MMN was present. Note that there was a total of 24 tests in each condition (12 participants × 2 tests) except for the CV condition, in which there were 22 tests (11 participants × 2 tests). There was a significant difference in the distribution of MMN ($\chi_1^2=9.92$, $p<0.002$). MMN was more frequently present for tones (low 63%, mid 71%, and high tones 46% of tests) than for speech sounds (words 25% and CVs 32%). Figure 3 also shows exemplars of waveforms recorded at F_z from two participants with the result of the integral-distribution test for MMN indicated beside each pair of averages. These waveforms show MMN occurring in the 100–220-ms epoch, as can also be appreciated in Fig. 2.

Given the relatively low number of tests showing speech-evoked MMN, *post-hoc* analyses of these data were carried out. First, alternative averages were created, using several different referencing points, including (1) a common average reference; (2) linked mastoids (M_1+M_2); and (3) each of the electrode sites (F_z, C_z, P_z, C_4, T_4, C_3, T_3). MMN was no more evident in any of these configurations. Second, visual inspection of the difference waveform was carried out to determine if MMN occurred outside of the designated analysis epoch (100–220-ms poststimulus onset). These analyses did not result in an increase in the number of individuals demonstrating MMN.

The results from individual tests of MMN are shown in Table I. For each stimulus condition there are two tests, reflecting the counterbalanced test design, in which each stimulus was used as a standard and as a deviant. (The complex tone test is described below.) There is no pattern to the presence or absence of MMN across different stimuli for individual subjects, except that individuals were more likely to have MMN for low- and midfrequency tone contrasts and less likely to have MMN for words or CVs. The two individuals with no MMN for low-frequency tones (ES and MD) did not have MMN for high frequency tones or for CVs.

B. Stimulus effects on MMN and P_{3a} in group data

Grand mean deviant, standard, and difference averages for all stimulus conditions are shown in Fig. 2. Panel B shows the grand mean difference average for all participants

TABLE I. Presence of MMN in individuals. A filled circle (●) indicates the presence of MMN while (⊗) indicates the absence of MMN according to the integral test. There were two tests for MMN for each stimulus condition, with each item in the stimulus pair used as a standard and then as a deviant. For some individuals, the MMN was present for one stimulus within a condition (e.g., 400-Hz standard, 400-Hz deviant, but not for the other 440-Hz standard, 440-Hz deviant).

Subject	Low	Mid	High	Word	CV	Complex tone
bc	●●	●●	⊗●	⊗⊗	…	●⊗
cw	●●	●⊗	●⊗	⊗⊗	●⊗	⊗⊗
es	⊗⊗	⊗●	⊗⊗	⊗●	⊗⊗	…
hw	●●	●●	●●	⊗●	●●	⊗⊗
jw	⊗●	⊗⊗	⊗⊗	⊗⊗	●⊗	…
lf	●⊗	●●	●⊗	●●	●⊗	…
lm	●●	●●	●●	⊗●	⊗●	…
md	⊗⊗	●●	⊗⊗	⊗●	⊗⊗	⊗●
pv	⊗●	●●	⊗⊗	⊗⊗	⊗⊗	⊗⊗
tb	⊗●	●●	●●	⊗⊗	⊗⊗	⊗⊗
ts	●●	●⊗	●⊗	⊗⊗	⊗⊗	⊗⊗
vr	●⊗	⊗⊗	●⊗	⊗⊗	●⊗	●⊗

in the study while panels C and D are the grand mean difference averages for those with (C) or without (D) MMN as determined by the integral-distribution test. The following description of MMN parameters are only for those subjects who had MMN present according to this criterion (i.e., those represented in panel C).

MMN area decreased as frequency increased, and MMN area for low-frequency tones was larger than that for word or CV stimuli as shown in Fig. 4. There was no difference in the area of MMN whether evoked by words or CVs ($F_{4,51}=3.88$, $p<0.01$ with Student–Newman Keuls *post hoc* pairwise comparisons). There was no significant effect of stimulus on MMN peak amplitude ($F_{4,51}=2.07$, $p=0.10$), peak latency ($F_{4,51}=1.15$, $p=0.34$), onset latency ($F_{4,51}=0.55$, $p=0.70$), or MMN duration ($F_{4,51}=2.41$, $p=0.06$). There was a significant difference for offset latency as a function of stimulus, with low tones having the longest offset latency

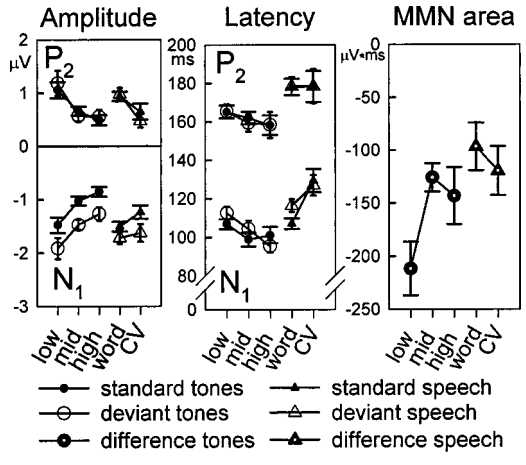


FIG. 4. Mean amplitude (μV) and latency (ms) of N_1 and P_2 at F_z for standard tones (filled circles), deviant tones (open circles), standard words and phonemes (filled triangles), and deviant words and phonemes (open triangles). Mean integral values of the difference waves (μV×ms) calculated over 100–220 ms measured at F_z for tones (dotted circles) and words and phonemes (dotted triangles). Error bars indicate standard error of the mean.

TABLE II. Mean amplitudes (μV), latencies (ms), and area (μV×ms) with standard deviations bracketed for N_1 and P_2, MMN and P_{3a}. N_1, P_2, and P_{3a} measures were made for all participants. MMN measures were made only for those tests passing the integral-distribution test.

	N_1 amp		N_1 lat		P_2 amp		P_2 lat		MMN					P_{3a}	
	Dev	Stand	Dev	Stand	Dev	Stand	Dev	Stand	Area	Peak amp	Peak lat	Onset lat	Offset lat	Peak amp	Peak lat
Low	−1.92	−1.48	112.6	106.8	1.19	1.05	165.3	165.1	−211.8	−1.58	168.9	95.8	224.2	1.90	283.0
	(0.96)	(0.71)	(15.5)	(12.5)	(0.84)	(0.68)	(18.0)	(16.5)	(97.6)	(0.59)	(26.6)	(33.2)	(28.8)	(1.12)	(47.3)
Mid	−1.47	−1.03	104.2	98.8	0.57	0.66	159.3	161.9	−126.0	−1.09	146.1	92.7	196.5	1.35	270.2
	(0.47)	(0.41)	(20.9)	(17.3)	(0.42)	(0.36)	(20.0)	(14.9)	(55.2)	(0.35)	(34.9)	(19.1)	(36.7)	(0.84)	(32.7)
High	−1.27	−0.86	95.3	101.0	0.55	0.49	158.3	158.2	−143.3	−1.29	154.5	104.2	216.3	1.67	278.1
	(0.62)	(0.42)	(15.8)	(19.6)	(0.54)	(0.41)	(29.4)	(20.8)	(89.1)	(0.80)	(37.7)	(41.1)	(28.5)	(0.94)	(34.6)
Words	−1.72	−1.51	116.3	107.0	0.97	0.94	178.4	178.0	−96.7	−1.18	162.1	110.3	184.2	1.29	273.3
	(0.56)	(0.45)	(15.9)	(13.6)	(0.59)	(0.40)	(22.5)	(20.4)	(54.9)	(0.47)	(26.2)	(22.8)	(25.0)	(0.43)	(49.3)
CVs	−1.63	−1.24	126.8	129.2	0.47	0.65	178.4	178.3	−119.7	−0.96	152.9	98.0	218.5	0.86	343.6
	(0.72)	(0.55)	(23.4)	(26.7)	(0.54)	(0.60)	(37.7)	(32.4)	(61.1)	(0.58)	(19.3)	(21.4)	(36.7)	(0.44)	(66.1)

and words the shortest, but there appeared to be no systematic trends ($F_{4,51}=2.69$, $p=0.04$). The mean areas, amplitudes, and latencies of MMN for all stimulus contrasts are presented in Table II.

Visual inspection of the difference waves (such as those shown in Fig. 2) revealed the presence of a positive component at latencies exceeding 220 ms which was identified as P_{3a}. This cognitive potential is thought to signal that attention has been directed to the deviant stimulus and may occur even in a passive listening paradigm (Lavoie *et al.*, 1997; Sams *et al.*, 1985; Winkler, Tervaniemi, and Näätänen, 1997). P_{3a} occurred at a latency around 225–320 ms for tones, 220–235 ms for words, and 320–405 ms for CVs (Table II). As expected, P_{3a} was larger in response to the deviant sound [one-way analysis of variance (ANOVA) $F_{1,158}=38.4$, $p<0.0001$], but it was also observed in response to standard sounds. P_{3a} was present in 80% of averaged waveforms, and for 56% of tests it was present in *both* standard and deviant stimulus waveforms, despite the use of a commonly employed form of distraction (video) and verbal instruction to ignore the sounds. Further analyses revealed that P_{3a} was more often present for standard speech sounds than for standard tones ($\chi^2_1=6.34$, $p=0.01$). P_{3a} was often present for both standard and deviant speech and so it was partly canceled in the difference waveform (see Fig. 2).

C. Stimulus effects on N_1 and P_2 amplitude and latency

N_1 was present in 96% and P_2 in 86% of all averages. The mean amplitudes and latencies of N_1 and P_2 measured at F_z are plotted in Fig. 4 and summarized in Table II. Two-way ANOVA (5 stimulus types × 2 roles, standard or deviant) and *post hoc* Student–Newman–Keuls pairwise multiple comparisons were used to determine the statistical significance of amplitude and latency differences.

N_1 amplitude and latency decreased with frequency. Mid- and high-frequency tones evoked a smaller N_1 than low-frequency tones, words, or CVs ($F_{4,217}=8.48$, $p<0.0001$). N_1 was larger in deviant averages compared to standard averages ($F_{1,217}=21.96$, $p<0.0001$) and there was no interaction between factors ($F_{4,217}=0.30$, $p=0.88$). The latency of N_1 was shorter when evoked by mid- and high-frequency tones than by low-frequency tones. There was no difference between the latency of N_1 evoked by low-frequency tones and words, but CVs evoked a longer latency response than any other stimulus ($F_{4,217}=16.78$, $p<0.0001$). There was no effect of stimulus role ($F_{1,217}=1.07$, $p=0.30$) and no interaction ($F_{4,217}=1.34$, $p=0.26$).

The stimulus effects on P_2 amplitude were like those on N_1. P_2 amplitude decreased with frequency. Mid- and high tones and CVs evoked a smaller response than low tones and words ($F_{4,168}=8.21$, $p<0.0001$). There was no effect of stimulus role ($F_{1,168}=0.004$, $p=0.95$) and no interaction ($F_{4,168}=0.44$, $p=0.78$).

There was no effect of frequency on the latency of P_2 but, as for N_1, the latency of the response evoked by words and CVs was longer than for tones ($F_{4,194}=6.46$, $p<0.0001$). There was no effect of stimulus role ($F_{1,194}=0.01$, $p=0.91$) and no interaction ($F_{4,194}=0.03$, $p=0.99$).

In order to compare the magnitude of MMN relative to N_1, the ratio of MMN area to the area of N_1 in response to standards was computed for each condition. The area ratios (standard deviations in brackets) were 2.51 (1.32), 1.78 (1.29), 2.36 (0.83) for low-, mid-, and high-frequency contrasts, respectively, and 2.54 (3.39) and 0.83 (0.61) for CVs and words, respectively. These amplitude ratios were not significantly different ($F_{4.51}=1.60$, $p=0.19$), but this appeared to be due to the variability for the CV condition. When CVs were excluded and words were compared to tones, the MMN-to-N_1 area ratio for words was significantly smaller compared to tones ($F_{1.47}=7.12$, $p=0.01$). This suggests that for words the determinants of MMN were disassociated from the determinants of N_1, whereas for tones both N_1 and MMN decreased in parallel as pure-tone frequency increased.

D. MMN for complex nonspeech stimuli

The results from the experiments using words and CVs were surprising given the findings of other published reports. We had expected to obtain MMN for discriminable words, as well as for CVs, since there are a number of reports of MMN in response to CVs that are discriminable (e.g., Kraus *et al.*, 1999) and also to those that are not (Dalebout and Fox, 2000; Kraus *et al.*, 1999). To test the hypothesis that the complex-

ity of the stimulus had an effect on MMN, an additional experiment was carried out in which MMN was measured in response to a complex nonspeech stimulus, in which there was an embedded frequency change. An embedded frequency change was chosen to represent a ''formant transition''; that is, a change in frequency for one component of a complex stimulus, with other components held constant.

The stimuli were contrasting 200-ms tone complexes. At onset they were composed of 400 Hz, 3000 Hz, and either 1500 or 1650 Hz. At 80 ms the 1500-Hz component was changed to 1650 Hz (''up''), or, for the second tone complex, the 1650-Hz component was changed to 1500 Hz (''down''), while the other tone components remained constant. Thus, the deviant stimulus had a slightly different initial frequency composition (10% frequency difference for the midfrequency component) as well as a change in frequency for an embedded component (up or down) at 80 ms. Responses were recorded using these two sounds in an oddball paradigm, with a 16% probability of a deviant, and all stimuli presented an ISI of 750 ± 50 ms. Recording and analyses were completed as for the pure-tone and speech tests, and the participants were a subset of 8/12 individuals who also participated in the previous experiments.

Robust obligatory responses were obtained in response to these tone complexes. N_1 was present in 97% of responses and P_2 in 69%. Mean peak amplitude and latency for N_1 in the deviant averages was -1.62 μV and 104.7 ms, and in the standard averages was -1.64 μV and 102.9 ms. Mean peak amplitude and latency for P_2 in the deviant averages was 1.33 μV and 152.7 ms, and in the standard averages was 0.90 μV and 151.2 ms. The presence of MMN in individuals was similar to that for speech sounds, with MMN recorded in only 3/16 (19%) tests. When MMN was present the mean area was similar in magnitude to that found for words and CVs (for these three cases $M=-104.5$ μV$\times$ms, cf. -96.7 and 119.7 μV$\times$ms for words and CVs, respectively) and the mean MMN-to-N_1 area ratio was low, more like the ratio found for words (for these three cases $M=0.66$; for words $M=0.83$).

IV. DISCUSSION

Overall, the results of this study indicate that MMN presence and area are affected by stimulus parameters such as frequency and complexity. It was expected that MMN would be evoked in all conditions since all participants were able to accurately discriminate the contrasting stimuli (Näätänen, 1992). However, we found that discriminable contrasts of tones evoked MMN more often than discriminable words, CVs, or a tone complex. The MMN evoked by low tones was larger in area than that evoked by speech sounds. As well, MMN evoked by tone contrasts showed a magnitude gradient, decreasing in area as frequency increased. This paralleled the gradient seen in the N_1 and P_2 responses for tones which also decreased in amplitude and latency (N_1 only) as frequency increased.

A. Effects of frequency on the amplitude of MMN, N_1, and P_2

These are the first data to describe the effect of frequency on the magnitude of MMN. These data show that MMN area decreased as tone frequency increased. Previous studies have shown that MMN magnitude decreases as stimulus deviance decreases (e.g., Sams *et al.*, 1985; Tiitinen *et al.*, 1994), but these data show, for the first time, that when stimulus deviance is held constant, MMN magnitude decreases as a function of frequency *per se*. The decreasing magnitude of MMN was not due to the degree of stimulus deviance or discriminability, as each contrast represented the same amount of frequency change (10%) and all frequency contrasts were easily discriminated by the listeners with an accuracy of 100%, although more sensitive measures of discrimination, such as reaction time, may reveal subtle differences in perceptual processing that are related to MMN.

The frequency-related magnitude gradient parallels that seen in the amplitudes of the N_1 and P_2 responses. These data support previous studies which have generally found that the amplitude (e.g., Antinoro *et al.*, 1969; Jacobson *et al.*, 1992; Pantev *et al.*, 1995) and latency (Alain *et al.*, 1997; Jacobson *et al.*, 1992) of the obligatory response decreases as stimulus frequency increases. This decrease is attributed to two different effects. First, a larger portion of the basilar membrane is activated by a low-frequency tone, because the wave travels from the basal to apical cochlear turn, compared to a high-frequency tone for which the traveling wave is restricted to the basal turn. It has been suggested that there is some cortical representation of basilar-membrane traveling wave properties resulting in larger cortical responses for low frequency tones (Antinoro *et al.*, 1969; Jacobson *et al.*, 1992). A study using functional magnetic resonance imaging (fMRI) supports this view, showing that more of the transverse temporal gyrus was activated by a 1-kHz tone compared to a 4-kHz tone (Strainer *et al.*, 1997). However, a study using positron emission tomography (PET) (Lauter *et al.*, 1985) showed that stimulation by low-frequency tones (500 Hz) produced the same change in regional cerebral blood flow as high-frequency tones (4000 Hz) at 50 dB SL, suggesting that low tones do not cause more cortical activation than high tones. Additional support for this comes from the findings of MEG studies which have shown that there is no frequency related difference in the amplitude of N_{1m} (the magnetic counterpart of N_1) (May *et al.*, 1999; Verkindt *et al.*, 1995) or the equivalent current dipole moment (Pantev *et al.*, 1988), suggesting that the number of neurones responding to each test frequency is the same across the range 250–4000 Hz (Pantev *et al.*, 1988). Finally, there is evidence that the tonotopic organization of the auditory cortex represents pitch rather than frequency and is therefore different from that seen in the cochlea (Pantev *et al.*, 1989; Pantev *et al.*, 1991).

The second alternative is that the frequency gradients for N_1 (and MMN) reflect the tonotopic organization of the auditory cortex, and the location of the neural generators relative to the scalp recording electrodes. Both N_1 and MMN have generators in the primary auditory cortex (Alho, 1995; Kropotov *et al.*, 1995; Näätänen and Picton, 1987). Studies

in humans using MEG (e.g., Elberling *et al.*, 1982; Pantev *et al.*, 1988; Pantev *et al.*, 1991; Romani *et al.*, 1982), PET (Lauter *et al.*, 1985), fMRI (Strainer *et al.*, 1997) and chronic microelectrode techniques (Howard *et al.*, 1996) have shown that areas with neurones responding best to high-frequency sounds are located deeper in the brain, in the more posterior and medial portions of the auditory cortex (transverse gyri of the supratemporal plane), whereas those areas with neurones responding best to low-frequency sounds are located more superficially, in the anterior and lateral areas. Therefore, the larger amplitudes of the obligatory evoked potentials for low-frequency sounds are attributed to the smaller distance between the neural generator and the site of recording for low-, compared to high-frequency tones.

The same explanation may account for the frequency-related MMN gradient described in this study. Just as for N_1, MMN evoked by low tones may be larger because it is generated in areas of the auditory cortex which are closer to the scalp. There already exists some evidence from MEG studies that, like N_1, the generators of MMN are tonotopically organized. Tiitinen *et al.* (1993) measured the MMN_m dipole for 10% changes in frequency for tones at either 250 Hz, 1 kHz, or 4 kHz using very similar stimulation parameters to those employed in this experiment. They reported a systematic and equivalent shift in the angles of the dipole in the sagittal plane for both N_{1m} and MMN_m and proposed that this was due to a systematic change in the location of the dipole reflecting the tonotopic organization of the auditory cortex generating the response. May *et al.* (1999) also found that the angles of both the N_{1m} and MMN_m dipole were frequency dependent. Studies investigating the magnetic equivalent of N_1 (May *et al.*, 1999; Pantev *et al.*, 1988; Verkindt *et al.*, 1995) and MMN (Tiitinen *et al.*, 1993) did not, however, indicate a frequency gradient in the magnitude of the response, such as was found in this study. This is most likely because, unlike electric fields, magnetic fields are not attenuated by the skull (Brandt, 1992). Consequently, the magnetic field resulting from stimulation with high-frequency sounds, which originates medially in the auditory cortex, may be measured at the scalp surface without significant attenuation, but the electric field arising from the same area may be diminished to such an extent that it is indistinguishable from noise. The results of this study therefore support the findings from previous MEG studies (May *et al.*, 1999; Tiitinen *et al.*, 1993) that the generators of MMN in the auditory cortex are arranged tonotopically.

The current study revealed that N_1 and MMN showed parallel changes in magnitude as tone frequency was manipulated. The ratio of MMN area to N_1 area was constant across frequency, indicating that both these components of the ERP changed proportionally. This study suggests that, for tone contrasts, the generators of MMN and the generators of N_1 are closely linked, since the magnitude of MMN and N_1 varied in essentially the same way.

B. MMN to speech stimuli

We found that discriminable words and CVs infrequently evoked MMN, with only 25%–32% of tests in individuals demonstrating a response which met a lenient ($p < 0.10$) statistical criterion. This finding for speech-like stimuli was unexpected given the wealth of reports in which MMN has been evoked by various synthesized vowels (e.g., Aaltonen *et al.*, 1997; Näätänen *et al.*, 1997) and CV syllables (e.g., Kraus *et al.*, 1999; Maistei *et al.*, 1995; Martin *et al.*, 1999; Sharma and Dorman, 1999) as well as to other complex acoustic signals (Alho *et al.*, 1996; Alho *et al.*, 1998; Tervaniemi *et al.*, 2000; Winkler *et al.*, 1997). Indeed, the results of the *post hoc* experiment with the complex nonspeech sounds were similarly disappointing. The results for speech and complex nonspeech stimuli were similar, in that N_1 and P_2 evoked by these sounds were robust, but MMN was not. MMN was less often present than for pure tones generally and of smaller magnitude than for low-frequency tones.

There are a number of possible explanations for these findings. First, this study showed that discriminable speech sounds evoked a smaller MMN than low tones which may reflect a more medial neural generator for the speech stimuli and so a greater distance between the generators and scalp recording electrodes. There exists evidence that generators of MMN_m evoked by complex acoustic signals are located more medially in the auditory cortex than those for simple tones. Generators for MMN_m evoked by CV contrasts (/ba/ vs /da/) have been found at around 4.2 cm from the midline along the axis drawn between preauricular points (Diesch and Luce, 1997) and for a musical chord contrast at about 4.6 cm from midline (Alho *et al.*, 1996). By comparison, pure tones generated MMN_m more laterally at 5.6–5.7 cm from midline (Alho *et al.*, 1996; Tiitinen *et al.*, 1993). If the generators of MMN for the words and CVs used in this study are also located more medially than those for low tones, then a smaller MMN for speech sounds is expected.

Another aspect worth considering is the extent to which P_{3a} was evoked by speech sounds. It appears that different attentional processes were engaged in response to speech sounds compared to tones. The cognitive P_{3a} component was more often present in *both* the deviant and standard waves when evoked by speech sounds (words or CVs), suggesting that these sounds engaged the listener's attention, albeit involuntarily. Because P_{3a} was evoked by both standard and deviant speech tokens, it was small or nonexistent in the difference waveform (deviant minus standard), but its presence for both standard and deviant speech stimuli may explain why MMN was less robust. A large P_{3a} with an onset latency and duration partly overlapping that of the MMN could obscure the presence and decrease the magnitude of MMN. Recently published studies of MMN which used speech-like stimuli demonstrate that the interpretation of MMN as a preattentive electrophysiological marker for fine acoustic differences (Kraus *et al.*, 1999) is not sufficient to explain the current results for speech. Szymanski, Yund, and Woods (2000) showed that listener attention to the deviant stimulus and phonemic cues of the speech stimuli influence the amplitude of MMN. Although we did not manipulate attention as a variable in this study, the presence of P_{3a} for word stimuli would be consistent with the explanation that attentional processes were engaged by speech

stimuli, and that this affected (decreased) the MMN.

Another explanation relates to the linguistic nature of the speech sounds and the possible engagement of additional brain areas or processes by these stimuli. There is evidence that several parts of the brain, not just the auditory cortex, are involved in the generation of MMN (Giard, Perrin, and Pernier, 1991; Giard *et al.*, 1990), and that this is also true when different stimuli are used. For example, findings in a small group of aphasic patients with lesions in the left temporoparietal region showed that this area is required for the generation of MMN to speech sounds but not to tones (Aaltonen *et al.*, 1993). Szymanski *et al.* (2000) showed that phonemic differences, rather than acoustic differences, influence the magnitude of MMN, similar to recent findings by Sharma and Dorman (1999) and contradicting previous work (e.g., Sharma *et al.*, 1993; Maistei *et al.*, 1995). There is also evidence that the linguistic relevance of otherwise meaningless CVs enhances MMN (Dehaene-Lambertz, 1997; Näätänen *et al.*, 1997). The words and CVs used in this study had the same acoustic characteristics (at least for the first 100 ms) but differed in linguistic content; that is, the words were recognized and understood as words by all of the listeners. By extension, one might expect that MMN evoked by words would be even further enhanced in comparison to CVs. However, we found that words were no more effective in evoking MMN than CVs (25% vs 31%) and when they did evoke a significant MMN there was no difference in its magnitude. There is a need for focused and comprehensive study of word versus CV contrasts, using, for example, words that are also CVs (e.g., /bai/ and /dai/).

Another possible explanation comes from the work of May *et al.* (1999). They suggested that stimulation of the auditory system with pure tones in an oddball sequence may result in a specific pattern of adaptation and lateral inhibition across parts of the auditory cortex and hypothesized that MMN to frequency change is the result of activity produced by a population of neurones responding to the deviant tone under the inhibitory influence exerted by the standard. This theory may help to explain the current findings for speech-based contrasts. Spectrally complex stimuli, such as words and CVs, are likely to result in a broad spread of activation over the auditory cortex (O'Leary *et al.*, 1998; Ottaviani *et al.*, 1997). As a consequence, repetitive presentation of the standard stimulus would set up a broad pattern of adaptation and lateral inhibition, affecting most of the areas which are subsequently activated by the deviant sound. The amount of overlap in the two populations would be considerable and while the deviant stimulus would generate somewhat different activity from that of the standard, it may not be sufficiently dissimilar to evoke a difference which is measurable by scalp recording electrodes.

The relative difficulty in obtaining MMN for speech stimuli was considered in view of the published literature and the methodology used. The contrast of /b/ and /d/ in a CV syllable has been used in other studies in which MMN has been recorded (e.g., Dehaene-Lambertz, 1997; Martin *et al.*, 1999; Sams *et al.*, 1990). We chose recording parameters, including electrode montage, stimulus level and rate, from those regularly used by researchers in this field, with the aim being to optimize the measurement of MMN. Unlike other studies that focused on either tones or speech, the experimental design involved recording responses to all types of stimuli in the same participants, thus reducing the variability that might otherwise be found. The large amplitude of the N_1 and P_2 responses to speech sounds indicates that the electrode montage was appropriate for the task of measuring ERP components to these sounds, as well as to tones. Rigorous and appropriate criteria for determining the presence of MMN were used and the method of analysis used a number of standard measures of MMN (peak amplitude and latency, area, onset, and offset latency and duration). With these experimental controls in mind, the finding of decreased magnitude and presence of MMN for words is a new finding.

C. MMN in individuals

These data show that MMN is not always apparent in all individuals able to discriminate the contrasting sounds. We used a statistical technique designed to improve the signal-to-noise ratio of the response, which provided a precise probability that an individual response constituted an MMN (probability that H_0 was true was generously set at $p<0.10$ rather than 0.05) and which accounted for individual differences rather than relying on normative or group data (Ponton *et al.*, 1997). In particular, we found that words and CVs evoked MMN in only 25%–32% of tests, similar to results found by Dalebout and Fox (2000), who report MMN in only 40% of listeners who discriminated CV tokens /da/ vs /ga/. These results are in contrast to findings of researchers such as Kraus and colleagues, who found that all (Kraus *et al.*, 1993a) or most (75%–83%) (Kraus *et al.*, 1999) of school-age children had MMN evoked by easily discriminable CVs (/da/ vs /ga/ or /ba/ vs /wa/). McGee, Kraus, and Nicol (1997) illustrated the importance of using well-constructed statistical criteria when determining the presence of MMN in individuals. In exploring the most efficient detection criteria, they found that a combination of area and onset latency gave the best d' value, but in this case the hit rate was 71%, corresponding to the best results obtained in this study, at least when tone contrasts were used. They found that when another common detection technique was used, visual inspection, the number of MMN responses in individuals increased to 94%, but this was associated with a disappointing 63% false alarm rate. It would appear that some speech sounds are more effective in evoking MMN than others and that this may change in different populations [compare the findings of Kraus *et al.* (1999), who reported MMN present in 75%–83% of children with those of Dalebout and Fox (2000) who found MMN in only 40% of adults]. The reasons for these differences warrant careful investigation. Clinical applications of MMN will be dependent upon objective detection criteria that are applicable for individuals, and some progress in this area has been made by Ponton *et al.* (2000), whose methods we used in this study.

Even when simple stimuli are used, as opposed to complex sounds or speech, the choice of stimulus feature being contrasted deserves consideration. Recently, Escera *et al.* (2000) showed that changes in the duration of a tone were highly effective in evoking MMN. They also found that the

test–retest reliability of the response was good, over the span of a month provided that appropriate measures of MMN were used (mean amplitude over 100–200 ms rather than peak amplitude or latency). In addition, Tervaniemi *et al.* (1999) showed that the test–retest reliability of MMN when it was evoked by a frequency change was not as high as when it was evoked by a change in duration. We chose rigorous statistical criteria to determine whether a waveform showed the MMN component and made appropriate MMN measurements. Using easily discriminable frequency contrasts, carefully selected measurement criteria, and appropriate statistical techniques it is reasonable to expect that, at best, 70%–75% of tests will be positive. Other stimulus contrasts may yield different results.

Because the MMN has a low signal-to-noise ratio in comparison to other ERPs (Picton *et al.*, 2000), it is important to use optimized stimulus and recording paradigms, and to control for artifact, such as from eye movements. In this study, a template-matching algorithm was used to subtract the eye-blink artifacts from the EEG records, prior to epoching and averaging. In addition, eye-blink artifacts exceeding $\pm 50\ \mu$V, in the F_z or C_z channel, caused an epoch to be rejected. As with any method used for reducing artifact, there is always the possibility that the artifact-rejection or -correction method will also reject or reduce the ''response'' that is present with the artifact. It is unlikely that the eye-movement artifact-reduction methods used in this study were a factor in the results. The same artifact-reduction methods were used across all subjects and all test conditions, and approximately equal numbers of epochs for deviants (>200) were averaged for each stimulus condition. There was no evidence that the different stimulus conditions resulted in a greater number of eye blinks to be corrected or rejected.

V. CONCLUSIONS

The results of this study indicate that MMN is influenced not only by the discrimination ability of the individual but by the characteristics of stimulus frequency and complexity. All of our participants were able to discriminate all of the contrasts with 100% accuracy, but the presence of MMN varied according to stimulus. MMN showed frequency-related effects similar to those seen in the N_1 and P_2 response. In addition, MMN was smaller for contrasts of discriminable words and CVs despite large-amplitude N_1 and P_2 responses.

There are two implications of the frequency dependency of MMN for experimental design. First, it seems advisable to use relatively low-frequency stimuli, at or below about 1500 Hz, when frequency contrasts are used to evoke MMN. Also, frequency differences of 10% produce easily discriminable differences, and robust MMN. We found that when testing normal adults who were able to discriminate the sounds, MMN was present in up to 71% of tests. It would appear that we cannot expect to evoke MMN consistently to some CV or CVC tokens discriminable on the basis of the place-of-articulation speech feature. In light of the difficulty in consistently obtaining MMN for easily discriminable speech sounds in normally hearing, cooperative adults, in well-controlled experimental conditions, the practical significance of MMN must be scrutinized. Words that are discriminated on the basis of acoustically based phonemic differences cannot be relied upon to evoke MMN. Neural mechanisms, such as lateral inhibition, adaptation, and attention may prove to be critical in the interpretation of MMN and its theoretical significance. The lability of MMN may limit its use as a tool to mark neurophysiological events involved in sound discrimination for individuals; however, investigation of the factors involved in this lability will certainly lead to insights regarding sound and speech perception.

ACKNOWLEDGMENTS

This work was supported by the Bionic Ear Institute Hearing Research Fund, the Royal Victorian Eye and Ear Hospital Research Institute, and the Beth Maclaren Smallwood Foundation. Many thanks to our colleagues, Mark Harrison, Anthony Burkitt, Ian Bruce, Rob Shepherd, Colette McKay, and the participants in this study.

Aaltonen, O., Eerola, O., Hellström, A., Uusipaikka, E., and Lang, A. H. (**1997**). ''Perceptual magnet effect in the light of behavioral and psychophysiological data,'' J. Acoust. Soc. Am. **101**, 1090–1105.

Aaltonen, O., Tuomainen, J., Laine, M., and Niemi, P. (**1993**). ''Cortical differences in tonal versus vowel processing as revealed by an ERP component called mismatch negativity (MMN),'' Brain Language **44**(2), 139–152.

Alain, C., Woods, D., and Covarrubias, D. (**1997**). ''Activation of duration-sensitive auditory cortical fields in humans,'' Electroencephalogr. Clin. Neurophysiol. **104**, 531–539.

Alho, K. (**1995**). ''Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes,'' Ear Hear. **16**(1), 38–51.

Alho, K., Tervaniemi, M., Huotilainen, M., Lavikainen, J., Tiitinen, H., Ilmoniemi, R. J., Knuutila, J., and Näätänen, R. (**1996**). ''Processing of complex sounds in the human auditory cortex as revealed by magnetic brain responses,'' Psychophysiology **33**, 369–375.

Alho, K., Winkler, I., Escera, C., Huotilainen, M., Virtanen, J., Jääskeläinen, I., Pekkonen, E., and Ilmoniemi, R. (**1998**). ''Processing of novel sounds and frequency changes in human auditory cortex: Magnetoencephalographic recordings,'' Psychophysiology **35**, 211–224.

Antinoro, F., Skinner, P., and Jones, J. (**1969**). ''Relation between sound intensity and amplitude of the AER at different stimulus frequencies,'' J. Acoust. Soc. Am. **46**, 1433–1436.

Brandt, M. E. (**1992**). ''Topographic mapping of brain electromagnetic signals: A review of current technology,'' Am. J. Physiol. Imag. **3**, 160–174.

Csépe, V., and Molnár, M. (**1997**). ''Towards the possible clinical application of the mismatch negativity component of event-related potentials,'' Audiol. Neuro-otology **2**, 354–369.

Csépe, V., Pantev, C., Hoke, M., Ross, B., and Hampson, S. (**1997**). ''Mismatch field to tone pairs: Neuromagnetic evidence for temporal integration at the sensory level,'' Electroencephalogr. Clin. Neurophysiol. **104**, 1–9.

Dalebout, S. D., and Fox, L. G. (**2000**). ''Identification of the mismatch negativity in the response of individual listeners,'' J. Am. Acad. Audiol. **11**, 12–22.

Dalebout, S. D., and Stack, J. W. (**1999**). ''Mismatch negativity to acoustic differences not differentiated behaviorally,'' J. Am. Acad. Audiol. **10**, 388–399.

Dehaene-Lambertz, G. (**1997**). ''Electrophysiological correlates of categorical phoneme perception in adults,'' NeuroReport **8**(4), 919–924.

Dehaene-Lambertz, G., and Baillet, S. (**1998**). ''A phonological representation in the infant brain,'' NeuroReport **9**, 1885–1888.

Diesch, E., and Luce, T. (**1997**). ''Magnetic mismatch fields elicited by vowels and consonants,'' Exp. Brain Res. **116**, 139–152.

Elberling, C., Bak, C., Kofoed, B., Lebech, J., and Sarmark, K. (**1982**). ''Auditory magnetic fields,'' Scand. Audiol. **11**, 61–65.

Escera, C., Yago, E., Polo, M. D., and Grau, C. (**2000**). ''The individual replicability of mismatch negativity at short and long interstimulus intervals,'' Clin. Neurophysiol. **111**, 546–551.

Giard, M.-H., Perrin, F., and Pernier, J. (**1991**). ''Scalp topographies dissociate attentional ERP components during auditory information processing,'' Acta Oto-Laryngol., Suppl. **491**, 168–175.
Giard, M.-H., Perrin, F., Pernier, J., and Bouchet, P. (**1990**). ''Brain generators implicated in the processing of auditory stimulus deviance: A topographic event-related potential study,'' Psychophysiology **27**(6), 627–640.
Howard, M., Volkov, I., Abbas, P., Damasio, H., Ollendieck, M., and Granner, M. (**1996**). ''A chronic microelectrode investigation of the tonotopic organization of human auditory cortex,'' Brain Res. **724**, 260–264.
Hughson, W., and Westlake, H. (**1944**). ''Manual for program outline for rehabilitation of aural casualties both military and civilian,'' Transcript of the American Academy of Ophthalmology and Oto-laryngology, 48 (Supplement), pp. 1–15.
Isard, S. D., and Miller, D. A. (**1986**). ''Diphone synthesis techniques (speech),'' Paper presented at the International Conference on Speech Input/Output; Techniques and Applications, London, UK.
Jacobson, G., Lombardi, D., Gibbens, N., Ahmad, B., and Newman, C. (**1992**). ''The effects of stimulus frequency and recording site on the amplitude and latency of multichannel cortical auditory evoked potential (CAEP) component N1,'' Ear Hear. **13**(5), 300–306.
Jasper, H. H. (**1958**). ''The ten twenty electrode system of the international federation,'' Electroencephalogr. Clin. Neurophysiol. **10**, 371–375.
Kraus, N., Koch, D. B., McGee, T., Nicol, T., and Cunningham, J. (**1999**). ''Speech-sound discrimination in school-age children: Psychophysiological and neurophysiologic measures,'' J. Speech, Language Hear. Res. **42**, 1042–1060.
Kraus, N., McGee, T., Carrell, T., Sharma, A., Micco, A., and Nicol, T. (**1993a**). ''Speech-evoked cortical potentials in children,'' J. Am. Acad. Audiol. **4**, 238–248.
Kraus, N., McGee, T., Micco, A., Sharma, A., Carrell, T., and Nicol, T. (**1993b**). ''Mismatch negativity in school-age children to speech stimuli that are just perceptibly different,'' Electroencephalogr. Clin. Neurophysiol. **88**, 123–130.
Kropotov, J. D., Näätänen, R., Sevostianov, A. V., Alho, K., Reinikainen, K., and Kropitova, O. V. (**1995**). ''Mismatch negativity to auditory stimulus change recorded directly from the human temporal cortex,'' Psychophysiology **32**, 418–422.
Lang, A., Eerola, O., Korpilahti, P., Holopainen, I., Salo, S., and Aaltonen, O. (**1995**). ''Practical issues in the clinical application of mismatch negativity,'' Ear Hear. **16**(1), 118–130.
Lauter, J., Herscovitch, P., Formby, C., and Raichle, M. (**1985**). ''Tonotopic organization in human auditory cortex revealed by positron emission tomography,'' Hear. Res. **20**, 199–205.
Lavoie, M. E., Robaey, P., Stauder, J. E. A., Glorieux, J., and Lefebvre, F. (**1997**). ''A topographical ERP study of healthy premature 5-year-olds in the auditory and visual modalities,'' Electroencephalogr. Clin. Neurophysiol. **104**, 228–243.
Maiste, A., Weins, A., Hunt, M., Scherg, M., and Picton, T. (**1995**). ''Event-related potentials and the categorical perception of speech sounds,'' Ear Hear. **16**, 68–90.
Martin, B., Sigal, A., Kurtzberg, D., and Stapells, D. (**1997**). ''The effects of decreased audibility produced by high-pass noise masking on cortical event-related potentials to speech sounds /ba/ and /da/,'' J. Acoust. Soc. Am. **101**, 1585–1599.
Martin, B. A., Kurtzberg, D., and Stapells, D. R. (**1999**). ''The effects of decreased audibility produced by high-pass noise masking on N1 and the mismatch negativity to speech sounds /ba/ and /da/,'' J. Speech, Language, Hear. Res. **42**, 271–286.
May, P., Tiitinen, H., Ilmoniemi, R. J., Nyman, G., Taylor, J., and Näätänen, R. (**1999**). ''Frequency change detection in human auditory cortex,'' J. Comput. Neurosci. **6**, 99–120.
McGee, T., Kraus, N., and Nicol, T. (**1997**). ''Is it really a mismatch negativity? An assessment of methods for determining response validity in individual subjects,'' Electroencephalogr. Clin. Neurophysiol. **104**, 359–368.
Moore, B. C. J. (**1995**). *Perceptual Consequences of Cochlear Damage* (Oxford University Press, New York).
Moore, B. C. J. (**1996**). ''Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids,'' Ear Hear. **17**(2), 133–161.
Näätänen, R. (**1992**). *Attention and Brain Function* (Erlbaum, Hillsdale, NJ).
Näätänen, R. (**1995**). ''The mismatch negativity: A powerful tool for cognitive neuroscience,'' Ear Hear. **16**(1), 6–18.
Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Ilvonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., and Alho, K. (**1997**). ''Language-specific phoneme representations revealed by electric and magnetic brain responses,'' Nature (London) **385**, 432–434.
Näätänen, R., and Picton, T. (**1986**). ''N2 and automatic versus controlled processes,'' Electroencephalogr. Clin. Neurophysiol. **38**, 169–186.
Näätänen, R., and Picton, T. (**1987**). ''The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure,'' Psychophysiology **24**(4), 375–425.
Pantev, C., Bertrand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuierer, G., and Elbert, T. (**1995**). ''Specific tonotopic organization of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings,'' Electroencephalogr. Clin. Neurophysiol. **94**, 26–40.
O'Leary, D. S., Andreasen, N. C., Hurtig, R. R., Hichwa, R. D., Watkins, G. L., Boles Ponto, L. L., Rogers, M., and Kirchner, P. T. (**1996**). ''A positron emission tomography study of binaurally and dichotically presented stimuli: effects of level of language and directed attention,'' Brain Lang. **53**, 20–39.
Ottaviani, F., Di Girolamo, S., Briglia, G., De Rossi, G., Di Guida, D., and Di Nardo, W. (**1997**). ''Tonotopic organization of human auditory cortex analyzed by SPET,'' Audiology **36**, 241–248.
Pantev, C., Hoke, M., Lehnertz, K., Lütkenhöner, B., Anogianakis, G., and Wittkowski, W. (**1988**). ''Tonotopic organization of the human auditory cortex revealed by transient auditory evoked magnetic fields,'' Electroencephalogr. Clin. Neurophysiol. **69**, 160–170.
Pantev, C., Hoke, M., Lütkenhöner, B., and Lehnertz, K. (**1989**). ''Tonotopic organization of the auditory cortex; pitch versus frequency representations,'' Science **246**, 486–488.
Pantev, C., Hoke, M., Lütkenhöner, B., and Lehnertz, K. (**1991**). ''Neuromagnetic evidence of functional organization of the auditory cortex in humans,'' Acta Oto-Laryngol., Suppl. **491**, 106–115.
Picton, T. W., Alain, C., Otten, L., Ritter, W., and Achim, A. (**2000**). ''Mismatch negativity: Different water in the same river,'' Audiol. Neuro-otol. **5**(3–4), 111–139.
Ponton, C. W., Don, M., Eggermont, J. J., and Kwong, B. (**1997**). ''Integrated mismatch negativity (MMNi): A noise-free representation of evoked responses allowing single-point distribution-free statistical tests,'' Electroencephalogr. Clin. Neurophysiol. **104**, 143–150.
Ponton, C. W., Eggermont, J. J., Don, M., Waring, M. D., Kwong, B., Cunningham, J., and Trautwein, P. (**2000**). ''Maturation of mismatch negativity: Effects of profound deafness and cochlear implant use,'' Audiol. Neuro-otol. **5**(3–4), 167–185.
Rapin, I., Schimmel, H., Tourk, L. M., Krasnegor, N. A., and Pollak, C. (**1966**). ''Evoked responses to clicks and tones of varying intensity in waking adults,'' Electroencephalogr. Clin. Neurophysiol. **21**, 335–344.
Romain, G. L., Williamson, S. J., Kaufman, L., and Brenner, D. (**1982**). ''Characterization of the human auditory cortex by the neuromagnetic method,'' Exp. Brain Res. **47**, 381–393.
Sams, M., Aulanko, R., Aaltonen, O., and Näätänen, R. (**1990**). ''Event-related potentials to infrequent changes in synthesized phonetic stimuli,'' J. Cogn. Neurosci. **2**(4), 344–357.
Sams, M., Paavilainen, P., Alho, K., and Näätänen, R. (**1985**). ''Auditory frequency discrimination and event-related potentials,'' Electroencephalogr. Clin. Neurophysiol. **62**, 437–448.
Schroger, E. (**1996**). ''The influence of stimulus intensity and inter-stimulus interval on the detection of pitch and loudness changes,'' Electroencephalogr. Clin. Neurophysiol. **100**, 517–526.
Semlitsch, H., Anderer, P., Schuster, P., and Presslich, O. (**1986**). ''A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP,'' Psychophysiology **23**(6), 695–703.
Sharma, A., and Dorman, M. (**1998**). ''Exploration of the perceptual magnet effect using the mismatch negativity auditory evoked potential,'' J. Acoust. Soc. Am. **104**, 511–517.
Sharma, A., and Dorman, M. (**1999**). ''Cortical auditory evoked potential correlates of categorical perception of voice-onset time,'' J. Acoust. Soc. Am. **106**, 1078–1083.
Sharma, A., Kraus, N., McGee, T., Carrell, T., and Nicol, T. (**1993**). ''Acoustic versus phonetic representation of speech as reflected by the mismatch negativity event-related potential,'' Electroencephalogr. Clin. Neurophysiol. **88**, 64–71.
Strainer, J. C., Ulmer, J. L., Yetkin, F. Z., Haughton, V. M., Daniels, D. L., and Millen, S. J. (**1997**). ''Functional MR of the primary auditory cortex: An analysis of pure tone activation and tone discrimination,'' Am. J. Neuroradiol. **18**(4), 601–610.

Szymanski, M. D., Yund, E. W., and Woods, D. L. (**2000**). ''Phonemes, intensity and attention; differential effects on the mismatch negativity (MMN),'' J. Acoust. Soc. Am. **106**, 3492–3505.

Tervaniemi, M., Lehtokoski, A., Sinkkonen, J., J., V., Ilmoniemi, R. J., and Näätänen, R. (**1999**). ''Test–retest reliability of mismatch negativity for duration, frequency, and intensity changes,'' Clin. Neurophysiol. **110**, 1388–1393.

Tervaniemi, M., Schroger, E., Saher, M., and Näätänen, R. (**2000**). ''Effects of spectral complexity and sound duration on automatic complex-sound pitch processing in humans — a mismatch negativity study,'' Neurosci. Lett. **290**(1), 66–70.

Tiitinen, H., Alho, K., Huotilainen, M., Ilmoniemi, R. J., Simola, J., and Näätänen, R. (**1993**). ''Tonotopic auditory cortex and the magnetoencephalographic (MEG) equivalent of the mismatch negativity,'' Psychophysiology **30**, 537–540.

Tiitinen, H., May, P., Reinikainen, K., and Näätänen, R. (**1994**). ''Attentive novelty detection in humans is governed by pre-attentive sensory memory,'' Nature (London) **372**(6501), 90–92.

Tremblay, K., Kraus, N., Carrell, T., and McGee, T. (**1997**). ''Central auditory system plasticity: Generalization to stimuli following listening training,'' J. Acoust. Soc. Am. **102**, 3762–3773.

Verkindt, C., Bertrand, O., Perrin, F., Eschallier, J.-F., and Pernier, J. (**1995**). ''Tonotopic organization of the human auditory cortex: N100 topography and multiple dipole model analysis,'' Electroencephalogr. Clin. Neurophysiol. **96**, 143–156.

Winkler, I., Tervaniemi, M., and Näätänen, R. (**1997**). ''Two separate codes for missing-fundamental pitch in the human auditory cortex,'' J. Acoust. Soc. Am. **102**, 1072–1082.

Winkler, I., Tervaniemi, M., Schroger, E., Wolff, C., and Näätänen, R. (**1998**). ''Preattentive processing of auditory spatial information in humans,'' Neurosci. Lett. **242**(1), 49–52.

Monaural masking release in random-phase and low-noise noise[a)]

David A. Eddins[b)]
Psychoacoustics Laboratory, Department of Communicative Disorders & Sciences, Center for Hearing & Deafness, State University of New York at Buffalo, Buffalo, New York 14214

(Received 1 October 1998; accepted for publication 2 January 2001)

Two masking-release paradigms thought to involve across-channel processing are comodulation masking release (CMR) and profile analysis. Similarities between these two paradigms were explored by comparing signal detection in maskers that varied only in degree of envelope fluctuation. The narrow-band-noise maskers were 10 Hz wide and their envelope fluctuations were manipulated using the low-noise noise algorithm of Pumplin [J. Acoust. Soc. Am. **78**, 100–104 (1985)]. Masking conditions included the classic CMR conditions of an on-frequency band, multiple (five) incoherent bands, or multiple coherent bands. Detection was compared using both random-phase noise (RPN) and low-noise noise (LNN) maskers. In one set of conditions, the signal was identical to the on-frequency masker, yielding an intensity discrimination task. Conditions that included RPN maskers and tonal signals resembled the classic CMR paradigm, whereas conditions including LNN and noise signals more closely resembled the classic profile analysis paradigm. Other conditions may be considered hybrids. This combination of conditions provided a wide variety of within- and across-channel cues for detection. The results suggest that CMR and profile analysis could be based upon the same set of stimulus cues and perhaps the same perceptual processes. [DOI: 10.1121/1.1352083]

PACS numbers: 43.66.Ba [JWH]

I. INTRODUCTION

In recent years, a number of studies have examined the way in which the detection of a narrow-band signal may be influenced by background stimuli spectrally remote from that signal. Two experimental paradigms that have been extensively applied to this issue are profile analysis and comodulation masking release (CMR). Under the typical conditions of both paradigms, the standard stimulus is composed of multiple *bands* that are separated in frequency such that they are likely to be spectrally resolved by the auditory system. In the context of critical band theory (e.g., Fletcher, 1940), this implies that these separate bands stimulate separate auditory channels or critical bands. Individual *bands* are typically composed of a single sinusoidal component in the profile analysis paradigm and a narrow-band noise in the CMR paradigm. It is generally assumed that information obtained from these separate bands may be combined or compared in some fashion that facilitates the detection of the signal. Thus, a theme common to both paradigms is across-channel or across-frequency processing of stimulus information. As explained below, however, profile analysis is often considered to be spectrally based whereas CMR is considered to be temporally based. One goal of the present experiment was to investigate the roles of various stimulus cues in profile analysis and CMR experiments including within-channel and across-channel cues, short-term and long-term cues, and short-term cues based upon level and modulation pattern.

While it has often been noted that the CMR and profile analysis paradigms are conceptually similar (e.g., Hall *et al.*, 1984), relatively few studies have compared the paradigms directly. This likely reflects differences among the typical stimuli used in the two paradigms (narrow-band noises versus tonal complexes). The present investigation was motivated, in part, by a rather unique CMR experiment reported by Hall and Grose (1988) and subsequent studies by Fantini and colleagues (Fantini *et al.*, 1993; Fantini and Moore, 1994a, b) in which the stimulus conditions employed had elements common to both the profile analysis and the CMR paradigm. The conditions of interest are most easily described in terms of the typical ''CMR'' paradigm, in which signal detection is measured in one of three masker conditions: on-frequency, incoherent, and coherent maskers. Briefly, the *on-frequency* masker is a single narrow-band noise centered at the signal frequency. An *incoherent* masker consists of the on-frequency band and one or more spectrally remote narrow-band noises whose temporal envelopes are statistically independent from that of the on-frequency band. A *coherent* masker consists of the on-frequency band and one or more spectrally remote masker bands that are constructed to have temporal envelopes identical to the on-frequency band. Thresholds for a tonal signal in the coherent condition typically are lower than in the on-frequency or incoherent conditions. This threshold difference, or masking release, generally is attributed to across-channel temporal cues derived from maskers having coherently modulated temporal envelopes. The addition of a tonal signal to the on-frequency noise band provides a spectrally local increase in level, change in the modulation depth, and change in modulation pattern that may be compared to maskers remote

[a)]Portions of this research were presented at the 129th meeting of the Acoustical Society of America, Indianapolis, Indiana, June 1996 [J. Acoust. Soc. Am. Suppl. 1 **99**, 2471 (1996)].
[b)]Electronic mail: deddins@buffalo.edu

in frequency to the on-frequency stimuli. Any or all of these cues could serve as the basis for a masking release.

In the typical profile analysis paradigm, the detection of an increment in a single component of a broad-spectrum, multicomponent tonal complex is estimated. It is generally believed that the auditory system processes stimulus level in different spectral regions and that the presence of a signal increment is cued by a change in the overall spectral shape or spectral profile. Under typical experimental conditions, stimuli used in profile analysis have no short-term temporal variations, and thus differential thresholds across stimulus conditions are discussed in terms of the available spectral cues. Furthermore, the overall stimulus level is randomly varied from interval to interval of an experimental trial to reduce the salience of within-channel changes in level introduced by the signal increment and to encourage the use of across-channel level cues for detection. Thus, the three most prominent distinguishing features between the two paradigms are the stable versus dynamic contexts for signal detection, the type of signal to be detected, and the use of roving level to control for within-channel level cues.

The stimuli used by Hall and Grose (1988) had features of both profile analysis and CMR paradigms. In their first experiment, the on-frequency and spectrally remote maskers were composed of narrow-band noise, similar to other CMR experiments, but the signal was an increment in the on-frequency band, similar to studies of profile analysis. This eliminated the availability of across-channel modulation depth and pattern cues, yet provided robust across-frequency level cues. From the profile analysis perspective, the novel feature of these conditions was the variable context provided by the narrow-band noise stimuli. The resulting masking release (on-frequency minus coherent threshold or incoherent minus coherent threshold) indicated that across-channel changes in modulation depth or pattern were not required for obtaining a masking release and that masking release based upon across-channel changes in level does not require a stable stimulus context. In a second experiment, the signal was a pure-tone rather than an increment in the on-frequency band and the level of the two spectrally remote narrow-band noise maskers was randomly varied across frequency. The random-level maskers limited the use of across-channel level cues, and the resulting masking release demonstrated that across-channel level cues are not required to obtain a masking release. Hall and Grose concluded that the masking release obtained with tonal or noise signals could be based upon across-channel comparisons of envelope amplitude (experiment I) or across-channel comparisons of envelope depth or pattern (experiment II).

Building upon the results of Hall and Grose (1988), Fantini and colleagues (Fantini *et al.*, 1993; Fantini and Moore, 1994a, b) noted that the results of Hall and Grose were consistent with decisions based upon changes in the short-term spectral shape rather than across-channel comparisons of envelope amplitudes or envelope modulation. Changes in the long-term spectral shape, they argued, would not provide differential thresholds in the incoherent and coherent conditions. In a series of investigations, Fantini and colleagues explored a number of similarities and differences among profile and CMR experiments including the use of stimulus manipulations to restrict the availability of various within-channel and across-channel cues. Collectively, these studies have shown that short-term across-channel spectral cues are less effective in isolation than when combined with across-channel cues related to differences in envelope pattern or depth.

The present study expands upon previous investigations by directly comparing signal detection in a profile analysis paradigm, featuring stimuli with stable across-channel spectro-temporal patterns, a CMR paradigm, featuring noise maskers with variable spectro-temporal patterns, and hybrid conditions with various degrees of temporal fluctuation. By comparing masked thresholds obtained with random-phase noise and low-noise noise, the present investigation introduces a novel method for direct comparison of the context variability characteristics typically associated with profile analysis and CMR experiments. The specific experimental design was similar to the study of Hall and Grose (1988), in which the typical CMR paradigm was modified by using as the signal an increment in the on-frequency noise masker. The approach employed here was to include one set of conditions in which the fluctuations in the temporal envelopes were minimized, thereby reducing the salience of any across-frequency envelope cue, while leaving the long-term magnitude spectrum unaltered. By using the low-noise noise algorithm provided by Pumplin (1985) and advanced by Hartmann and Pumplin (1988, 1991), it is possible to obtain a large set of minimal fluctuation noises having magnitude spectra statistically identical to random, ''Gaussian'' noises. The two classes of noise differ only in their phase spectra. In particular, the algorithm selects a phase spectrum that minimizes the fourth moment of the envelope for any given magnitude spectrum. Hartmann and Pumplin (1988, 1991) have shown the fourth moment of the envelope to be a very useful metric for the degree of envelope fluctuation. These two noise types, denoted random phase noise (RPN) and low-noise noise (LNN) here, are discussed in more detail in the Methods section below.

The present study included one set of conditions in which the signal was an increment in the level of the on-frequency noise band and a second set in which the signal was a sinusoid. In addition, detection thresholds were measured under conditions in which masker levels were fixed or were randomly varied about a mean level from interval to interval. Thus, the use of RPN maskers, characterized by marked fluctuations in the temporal envelope, and tonal signals, which introduce changes in overall level and modulation depth and pattern, parallels the typical CMR task. Likewise, the use of LNN maskers, characterized by very stable temporal envelopes, a level increment in the on-frequency band as the signal, and random level variations from interval to interval, parallels the typical profile analysis task. Evaluating signal detection in each of the hybrid conditions (independently varying noise type, signal type, and presence or absence of random level variation) allows a direct comparison of profile analysis to CMR. Furthermore, comparisons of increment detection in LNN and RPN permit an evaluation of the role of stimulus fluctuations in a simple intensity dis-

crimination task previously investigated using variable bandwidth RPNs (e.g., Bos and de Boer, 1966). In evaluating the stimulus cues available in each of the various stimulus conditions, the tendency to associate profile analysis with spectral cues and CMR with temporal cues will be avoided. Rather, the distinction made here will be between long-term versus short-term cues and within-channel versus across-channel cues, similar to the distinction made by Fantini and Moore (1994a).

II. METHOD

A. Subjects

Four listeners participated, ranging in age from 22 to 29 years. All had pure-tone thresholds of 15 dB HL or better at octave frequencies from 250 to 8000 Hz and normal middle-ear function based on a screening tympanogram (Y, 226 Hz). Each of the four listeners was highly practiced on a variety of psychoacoustic listening tasks including the detection of tones in single and multiple bands of noise. None had practice on intensity discrimination tasks with noise or tonal signals and none received extensive practice on the intensity discrimination tasks reported here. Except for the author, each listener was paid an hourly wage and received a 10% bonus upon completion of the study.

B. Procedures

An adaptive, two-interval, two-alternative forced-choice method with a three-down, one-up tracking algorithm was used to estimate the signal level corresponding to 79.4%-correct detection (Levitt, 1971). The initial step size of each 60-trial block was 5 dB and was reduced to 2 dB after the first three reversals of the adaptive track. Threshold was estimated by averaging the signal level at each of the last even number of reversals, excluding the first three. The final threshold for a condition was taken as the average threshold across six such estimates. Each trial consisted of two 400-ms observation intervals separated by 400 ms, marked by LEDs, and followed by visual feedback. Stimuli were presented monaurally via Sennheiser (HD-250) earphone and testing was conducted in a sound-attenuating chamber. The order of experimental conditions was randomized and listeners completed all six threshold runs for a given condition before continuing to the next condition. Informal retesting at the end of the experiment indicated no evidence of threshold improvement over the course of testing.

C. Stimulus conditions

Two masker types were used: random-phase noise (RPN) and low-noise noise (LNN). Each RPN masker was composed of five sinusoidal components spaced 2.5 Hz apart with Rayleigh-distributed magnitudes and uniformly distributed phases (0 to 2π). Similarly, LNN maskers consisted of five sinusoidal components with Rayleigh-distributed magnitudes and component phases chosen by an algorithm designed to minimize the variance of the instantaneous power of the waveform, or the fourth moment (Pumplin, 1985). Each 10-Hz band of RPN or LNN had a spectrum level of 53 dB SPL. Samples of RPN and LNN waveforms are shown in the upper left and lower left panels of Fig. 1, respectively. The panels on the right are discussed below.

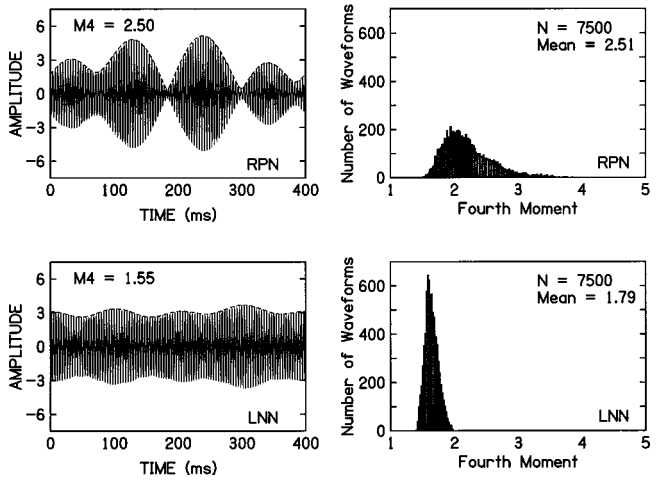


FIG. 1. The left column shows sample waveforms for RPN (upper left) and LNN (lower left) stimuli. The dashed line superimposed upon the waveforms represents the Hilbert envelope. Note the difference in amplitude fluctuation between RPN and LNN. In each case amplitude is in arbitrary units. The right column shows the distribution of fourth moment values computed from 7500 samples of RPN (upper right) and LNN (lower right) waveforms.

For each masker type, there were two reference conditions. The *on-frequency* reference consisted of a single, 10-Hz narrow-band-noise masker centered at 2500 Hz. The *incoherent* reference consisted of five 10-Hz narrow-band-noise maskers, each having independent component magnitude and phase values, and thus incoherent temporal envelopes. The five noise bands were centered at 1500, 2000, 2500, 3000, and 3500 Hz. Finally, the *coherent* comparison condition consisted of five 10-Hz narrow-band noises, each having identical component magnitude and phase values, and thus coherent temporal envelopes. These noise bands also were centered at 1500, 2000, 2500, 3000, and 3500 Hz.

Two signal types were used: a *sinusoidal signal*, which was simply an in-phase increment in the level of the central component of the masker centered at 2500 Hz, and a *noise signal*, which was an in-phase increment in the level of the entire 10-Hz noise band centered at 2500 Hz. In total, the experimental design included two noise types (RPN, LNN) by two signal types (NBN, pure-tone) by two presentation levels (roving level, no-rove) by three masker conditions (on-frequency, incoherent, coherent), for a total of 24 conditions.

D. Stimulus generation

All stimuli were generated in the frequency domain by filling portions of two 8192-point buffers with the appropriate magnitudes and phases. A subsequent inverse fast Fourier transform (FFT) on the complex buffer pair yielded the desired noise waveform. Though component magnitudes were randomly chosen, the total intensity of each individual masker band was normalized to a constant value. The signal was created by incrementing specific magnitude components in the frequency domain. The digital stimulus was output via

a 16-bit D/A converter (TDT DA3) with a sampling period of 24.4 μs. In all conditions, the signal and maskers had simultaneous onsets and offsets for a total stimulus duration of 400 ms, including 10-ms $\cos^2$ rise/fall envelopes. The median masker spectrum level was 53 dB SPL. However, in a subset of conditions, on each stimulus presentation the overall level of the masker plus signal was selected randomly (in 1-dB steps) from a uniform distribution with a mean of 53 dB SPL and range of 20 dB. Maskers without random level variation will be termed ''no rove'' and maskers with such variation will be termed ''rove.''

Samples of low-noise noise were generated using the algorithm described by Pumplin (1985) and Hartmann and Pumplin (1988). The interested reader is referred to those papers for a general discussion of the algorithm and a detailed description of the necessary computations. Briefly, a sample of low-noise noise is obtained by minimizing the fluctuations in the power of the waveform. Beginning with a magnitude spectrum, a directed search of the multidimensional space of possible component phases results in the set of phases that produces minimal variance in the power of the waveform, or the fourth moment. In practice, this algorithm is extremely computationally intensive. In the present study, the algorithm was implemented using the MATLAB (4.0) software package and a personal computer. Due to processing limitations, the algorithm was not practical for component numbers in excess of 5. A total of 7500 LNN samples was generated and their magnitude and phase spectra were stored on disk. Minimal fourth-moment values of the individual noise samples depended upon their specific, randomly selected magnitude spectra. The distribution of minimal fourth-moment values obtained from the 7500 noises is shown in the lower right panel of Fig. 1. For comparison, the distribution of fourth-moment values from 7500 samples of random-phase noise is shown in the upper-right panel. Although the two distributions overlap somewhat, the mean of the LNN distribution is about 1.5 standard units below the mean of the Gausian-noise distribution. The actual LNN stimuli used in this experiment included only the 375 LNN samples with the lowest fourth moments (<1.60). The LNN waveform in the lower-left panel of Fig. 1 is representative of this smaller set. Sample magnitude and phase spectra were chosen from this set of 375 LNN stimuli without replacement within an interval (i.e., the incoherent masker was composed of five completely independent noises) and with replacement across intervals and trials.

III. RESULTS AND DISCUSSION

Threshold values in each experimental condition were consistent across the four listeners (the standard error across all subjects and conditions was 0.65 dB) and thus performance is well represented by mean thresholds across listeners. The results for the narrow-band noise signals will be considered first, followed by the results for tonal signals. In each case, potential stimulus cues available in each condition will be discussed. Shown in Fig. 2 are the mean thresholds (upper panel) and corresponding estimates of masking release (lower panel) for the narrow-band-noise signals. Labels on the abscissa designate the specific stimulus conditions as

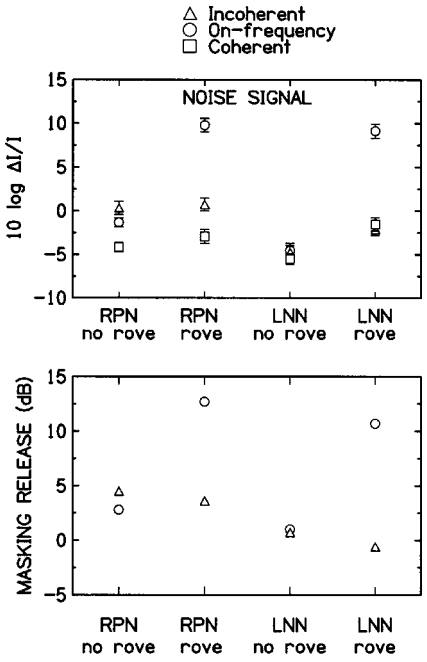


FIG. 2. Increment detection for a noise signal. Thresholds in units of $10\log(\Delta I/I)$ are shown in the upper panel and masking release is shown in the lower panel. Labels on the abscissa indicate the masker type (RPN or LNN) and the presence or absence of roving level from interval to interval. Symbols indicate the masker condition: on-frequency band (circles), incoherent bands (triangles), and coherent bands (squares). Masking release is shown in the lower panel for each of the noise types computed using either an on-frequency reference (circles) or an incoherent reference (triangles).

either RPN or LNN. ''Rove'' indicates a random variation in overall level from interval to interval of ± 10 dB about a median spectrum level of 53 dB SPL. ''No rove'' indicates a fixed masker spectrum level of 53 dB SPL. All thresholds are reported in units of $10\log(\Delta I/I)$ and, unless otherwise stated, all thresholds noted from previous studies are in units of $10\log(\Delta I/I)$ as well.

A. Narrow-band-noise signals

Mean increment-detection thresholds are shown for the on-frequency (circles), incoherent (triangles), and coherent (squares) masker conditions. Bars on each point show the average within-subject standard error. First, consider thresholds for the RPN masker with no rove in overall level (column 1). The on-frequency condition (circles) represents increment detection for a 10-Hz noise centered at 2500 Hz, for which the mean threshold, expressed as $10\log(\Delta I/I)$, was -1.3 dB. This threshold is slightly smaller than those reported by Fantini and Moore (1994a, 1.85 dB), Hall and Grose (1988, 4.57 dB), or Bos and de Boer (1966, 5.25 dB), for 10-Hz maskers centered at 2000 Hz (also see Table I).[1] This 3- to 6-dB difference likely results from a combination of factors, which may include differences in stimulus generation, individual differences among listeners, or practice effects. Signal detection in this spectrally-local increment detection condition is likely determined by within-channel level cues.

With the addition of incoherent flanking bands to the on-frequency band, thresholds in this second reference condition increased to 0.3 dB (triangle). Thus, incoherent flank-

ing bands produced a slight (1.6-dB) interference in increment detection. This result is commonly reported in studies of CMR and appears to be accentuated by using burst (gated) rather than continuous maskers (e.g., Moore *et al.*, 1990; Eddins and Wright, 1994; Fantini *et al.*, 1993). Incoherent noise bands are characterized by pronounced short-term level variations across frequency due to the moment-by-moment variability in the relative amplitude of the individual bands. Within this variable context, level changes introduced by the signal would need to be substantial to produce robust across-channel short-term level cues for detection. A better detection cue might be the same within-channel cue used in the on-frequency condition or a long-term across-channel level comparison. The fact that thresholds were worse in this condition than the on-frequency condition suggests that either the flanking bands interfered with the efficient use of such cues or that short-term across-channel cues were used.

In the comparison condition, the additional flanking bands were coherent with the on-frequency band, and threshold decreased to -4.2 dB. By comparing this threshold to either reference threshold, a masking release can be computed, as shown in the lower panel of Fig. 2. Relative to the on-frequency reference, the masking release was 2.8 dB and relative to the incoherent reference, the masking release was 4.5 dB. Note that the latter reflects both improved detection due to across-frequency coherence in the comparison condition and impoverished detection due to across-frequency incoherence in the reference condition. In general, these results are consistent with those of Hall and Grose (1988), who reported on-frequency and incoherent CMRs of 4.8 and 5.0 dB, respectively.

Long-term cues would not be different for the incoherent and coherent conditions, so the difference in thresholds obtained in the two conditions indicates the use of short-term cues. Furthermore, the fact that threshold improved with the addition of coherent flanking bands to the on-frequency band indicates the use of across-channel rather than or in addition to within-channel cues. Because the noise signal does not change the modulation depth or pattern of the on-frequency band, one may conclude that short-term, across-channel level cues provide the basis for lower thresholds in the coherent rather than the incoherent or on-frequency condition.

To encourage the use of across-channel cues for detection in the multiband conditions, a random level variation of ± 10 dB was introduced. The resulting thresholds are shown in the second column of Fig. 2. Roving the overall level produced a mean threshold in the on-frequency condition (circle) of 9.8 dB, some 11.1-dB higher than the corresponding no-rove condition. The rove would disrupt the long- or short-term within-channel level cues presumably used in the no-rove condition above. Fantini and Moore (1994a, their Appendix) showed that threshold based upon the change in overall level of the stimulus should be 10.1 dB for the ± 10-dB rove used here, in close agreement with the 9.8-dB threshold actually obtained. The addition of incoherent flanking bands (triangle) reduced the increment threshold to 0.7 dB, an improvement of 9.1 dB relative to the roving-level on-frequency condition. This improvement is consistent with the influence of either long- or short-term across-channel level cues. Although short-term level cues would be quite variable over time, their variability may have less of a detrimental influence on threshold than that imposed by the rove.

When the flanking bands were coherent with the on-frequency band, threshold decreased to -2.9 dB (square), revealing a 12.7-dB improvement relative to the on-frequency condition and a 3.7-dB improvement relative to the incoherent condition. Thus, the coherent condition led to a greater masking release than the incoherent condition. The use of long-term across-channel cues would not yield differential thresholds in the incoherent and coherent conditions, but a short-term across-channel cue would. Likewise, the noise signal would not introduce changes in the modulation depth or pattern, and thus one can conclude that the cue for detection was a short-term across-channel level cue. It seems likely that the same cue would serve as the basis for detection in the incoherent condition (as opposed to a long-term cue); however, the short-term analysis would be subject to marked variability.

The last two conditions listed on the abscissa of Fig. 2 include LNN maskers. In this case, the temporal envelope of each noise band is characterized by minimal fluctuations in amplitude, and is closer to the envelope of a pure tone than to a typical narrow-band noise. Thus, the LNN stimulus complex more closely resembles stimuli commonly used in a standard profile analysis paradigm, consisting of tonal complexes. Because the envelope of an LNN has so little fluctuation, there is little difference between incoherent and coherent LNN maskers. In the case of no rove, the increment threshold for the on-frequency condition was -4.5 dB, which is very close to estimates of intensity discrimination for a single pure tone. Summarizing the results from 15 studies of pure-tone intensity discrimination, and correcting for differences in duration, Green (1988) reported an average threshold in units of $20\log(\Delta A/A)$ of -17.7 dB, which is equal to $10\log(\Delta I/I)$ of -5.5 dB. The addition of incoherent or coherent flanking bands to the on-frequency band produced only a slight decrease in threshold (0.2 and 1.0, respectively). Following the analysis of the no-rove conditions above, the use of short-term across-channel level cues would be consistent with the lack of a substantial difference in detection threshold for the two masker types, given the similarity of incoherent and coherent LNN stimuli (note the slightly lower thresholds in the coherent than incoherent condition, possibly reflecting the benefit of across-channel coherence). If it is assumed that the small threshold differences among the three LNN-no-rove conditions simply reflects variability in threshold estimates, then thresholds in all three conditions could be explained in terms of within-channel level cues.

In the fourth condition, which consisted of LNN maskers with random level variation from interval to interval, threshold again was quite high for the on-frequency masker (9.1 dB), consistent with degradation of within-channel level cues by the rove. The addition of flanking bands produced approximately 11 dB of threshold improvement, regardless of the coherence among the noise bands. This improvement again illustrates a robust across-channel level cue in the absence of any substantial temporal cues.

Likewise, the absence of a multiband CMR (incoherent minus coherent threshold) reflects minimal differences between incoherent and coherent LNN conditions. The addition of roving level resulted in a threshold increase in the multiband conditions of 4 dB, which is a rather large change relative to the masking release obtained in this and other studies. No corresponding change was evident for the RPN conditions. This threshold difference with LNN may simply reflect the switch from within-channel level cues without rove to across-channel level cues with rove. Although it is not entirely clear why the 4-dB change occurred, it may be related to practice and order effects (see Sec. IV). Since long-term across-channel cues would be the same for incoherent (or coherent) RPN and LNN stimuli, the slightly lower thresholds (2 dB) for incoherent LNN than incoherent RPN would be consistent with the use of short-term level cues that are less variable for LNN than RPN. Finally, the fact that the no-rove LNN thresholds are the lowest of all thresholds may simply reflect the sensitivity of within-channel intensity discrimination to any level variations, whether inherent in the noise itself (RPN) or imposed by rove.

To summarize the results in Fig. 2, detection in the on-frequency conditions is consistent with the use of within-channel, short-term level cues. For RPN, detection in incoherent and coherent conditions is consistent with the use of short-term across-channel level cues. Under conditions of high level uncertainty (rove), short-term across-channel level cues provide a robust release from masking. For LNN, detection in the no-rove condition is consistent with a within-channel level cue, regardless of the presence of multiple flanking bands. With the addition of roving level, however, across-channel cues provide a robust masking release, yet the across-channel cues are not used as effectively as they are in RPN.

B. Pure-tone signals

The waveforms produced by combining pure-tone signals with RPN or LNN maskers differ substantially in terms of their characteristic modulation depth or the degree of amplitude fluctuation. It will be important to keep this difference in mind as thresholds are compared for these two noise types. With both masker types, the addition of a tonal signal to a 10-Hz masker produces an increase in overall level and more regular zero crossings in the waveform fine structure. For an RPN masker, the signal plus masker yields a reduction in the degree of amplitude fluctuation or modulation depth relative to the masker alone. For an LNN masker, however, signal-plus-masker waveform has progressively greater amplitude fluctuations as the signal level is increased from subthreshold to suprathreshold levels. At signal levels well above threshold, the degree of amplitude fluctuation will actually decrease with increasing signal level as the signal-plus-masker envelope becomes more dominated by the sinusoidal signal. As a result, near threshold the characteristic modulation depth decreases for signal-plus-RPN masker and increases for signal-plus-LNN masker.

The results for pure-tone signals are shown in Fig. 3. Recall that the addition of a pure-tone signal to the on-frequency masker was simply an in-phase increment in the central component of the five-component masker band centered at 2500 Hz. For RPN, with no rove (column 1), threshold in the on-frequency condition was −2.9 dB. This is 1.6 dB lower than the threshold for the noise signal, and is consistent with the notion that changes in modulation depth/pattern as well as changes in overall level within the on-frequency channel provide better detection than within-channel changes in level alone. The addition of incoherent flanking bands increased threshold slightly, to −0.6 dB, consistent with the interference effect discussed above. The addition of coherent flanking bands decreased thresholds to −6.9 dB. As shown in the lower panel, the masking release relative to the on-frequency reference was 4.0 dB and relative to the incoherent reference was 6.4 dB. The fact that thresholds were lower in the coherent condition than in both the on-frequency and incoherent conditions is consistent with the use of a short-term across-channel cue. Furthermore, threshold was 2.8 dB lower for the tonal (Fig. 3) than the noise (Fig. 2) signal. For both signal types, short-term across-channel level cues would be available, but for the tonal signal the use of additional short-term across-channel modulation depth/pattern cues would explain lower threshold values.

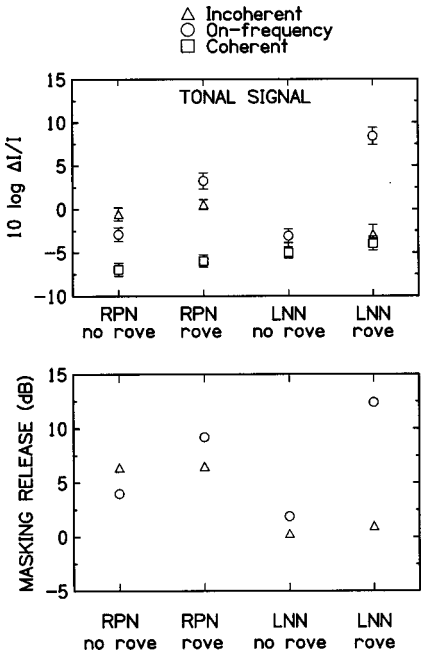


FIG. 3. Increment detection for a tonal signal. Same as Fig. 2.

The introduction of roving level (column 2) shifted the on-frequency threshold from −2.9 to 3.3, a change of 6.2 dB. This contrasts with the 11.1-dB change noted for noise signals and is again consistent with the notion that the within-channel change in modulation depth or pattern introduced by the tonal signal provided an important cue for detection not available with a noise signal. When incoherent flanking bands were added (column 2), threshold decreased to 0.5 dB. This 2.8-dB improvement may be interpreted as a masking release in the *incoherent* condition due to either short- or long-term across-channel level cues. In general, attention to short-term across-channel level cues would explain the threshold differences between incoherent and on-

frequency or coherent noise bands. With coherent flanking bands, threshold decreased to −6.0 dB, providing a masking release of 6.5 dB relative to the incoherent reference and 9.2 dB relative to the on-frequency reference (lower panel). Furthermore, lower thresholds for tonal (Fig. 3) than noise (Fig. 2) signals in this condition may reflect the additional use of short-term across-channel modulation depth/pattern cues with tonal signals not available with noise signals.

With LNN stimuli and no rove (column 3), thresholds for tonal signals were clearly lower for the multiband maskers than for the on-frequency masker, producing masking release of about 2 dB for both the coherent maskers (lower panel) and the incoherent maskers. The distinction between incoherent and coherent noise again is negligible with LNN maskers, consistent with the results for noise signals above. The masking release seen for tonal (Fig. 3) and not noise (Fig. 2) signals likely reflects the use of short-term across-channel modulation depth cues rather than across-channel level cues, since both signal types would lead to the latter. The use of short-term across-channel modulation depth/pattern cues in conditions of masking release is supported by the results of Hall and Grose (1988) and Fantini and Moore (1994b), who randomized the level of flanking bands to render short-term across-channel level cues unreliable and still reported masking release for coherently fluctuating maskers.

Roving level in the LNN condition (column 4) shifted the on-frequency threshold from −3.1 to 8.4 dB, a change of 11.5 dB. This is somewhat smaller than the 13.7-dB change noted for the noise signal, consistent with the use of a within-channel modulation cue introduced by the tonal signal. In the multiband conditions, thresholds in LNN decreased substantially relative to the on-frequency condition, with little advantage for coherent over incoherent maskers, similar to the no-rove condition. In general, roving level produces higher thresholds but this is not always the case in the coherent conditions. For example, while rove increased thresholds by 4 dB in the LNN-coherent condition with a noise signal, rove did not affect threshold for a tonal signal in the LNN-coherent condition or thresholds for either signal type with RPN. These results contrast with those of Fantini and Moore (1994b), who used AM tonal stimuli and consistently found that thresholds in coherent conditions were worse in the presence of roving level. Because roving level disrupts the use of within-channel cues, one interpretation is that such cues are being used in conditions in which rove does introduce a substantial threshold change.

IV. GENERAL DISCUSSION

A. Comparison of RPN and LNN maskers

While the pattern of results across masking conditions was similar in many ways for the two signal and two noise types, a few noteworthy differences provide further insight into the nature of the stimulus cues used in these detection tasks. First, consider threshold differences between RPN and LNN maskers of noise signals. In the on-frequency condition without rove, thresholds were substantially lower in LNN than RPN (3.2 dB). In both cases, the only available cue for detection was a within-channel level cue. A long-term analysis of stimulus level would not be different for the two noise types. A short-term analysis of level, however, would be sensitive to the large moment-to-moment variability in level of the standard RPN stimulus (due to inherent temporal fluctuations) and the greater stability of the LNN stimuli, which likely serves as a basis for lower thresholds with LNN. With the addition of a ±10-dB rove, the random level variations introduced to RPN and LNN were substantially larger than the level variations characteristic of RPN without rove. The unstable stimulus level from interval to interval led to similar thresholds in both cases (9.1 and 9.8 dB). The greater influence of rove for LNN maskers is consistent with the result of Fantini and Moore (1994a), who reported a larger effect of rove on threshold for a noise with a bandwidth of 64 Hz than a bandwidth of 8 Hz. In this case, the wider band is perceived as having a smoother temporal envelope than the narrower band, presumably as a result of limited auditory temporal resolution.

For the incoherent noise, no-rove conditions, thresholds were 5.1 dB lower for LNN than RPN. Above, it was suggested that thresholds in the incoherent LNN no-rove condition could be explained on the basis of either short-term within-channel or across-channel cues, since thresholds were the same for LNN on-frequency and incoherent conditions. Poorer detection in the RPN conditions would be consistent with a reduced ability to make use of those short-term level cues, likely resulting from the moment-to-moment variability of RPN. The fact that threshold was actually 2.4 dB worse in the incoherent than the on-frequency RPN condition is evidence that short-term processing was occurring but may be inefficient in the face of stimulus level uncertainty due to the moment-by-moment variability inherent in narrow-band noise.

Finally, in the coherent condition, thresholds were again lower for LNN than RPN. Although the short-term across-channel level cues available for the RPN condition provide a substantial improvement in threshold relative to the RPN on-frequency condition, the within- or across-channel cues used in the LNN-coherent condition lead to lower thresholds than for RPN. The addition of rove has little effect on thresholds in the RPN coherent condition but a substantial (4.0 dB) effect in the LNN coherent condition. If detection in the LNN-coherent condition were based primarily upon a within-channel level cue in the absence of rove (consistent with nearly equal on-frequency, incoherent, and coherent thresholds), then rove would require the listener to begin using a relatively less effective across-channel level cue, leading to poorer signal detection. The notion of a dominant within-channel cue in conditions generally considered to promote across-channel processing is not novel. Fantini and Moore (1994b) measured the detection of an increment in the central component of an AM tone in the presence of coherent AM flanking tones and suggested that within-channel cues might account for similar threshold improvements with increasing depth of AM for both on-frequency and coherent conditions.

When comparing thresholds for tonal signals in RPN and LNN maskers, it is important to keep in mind that the

addition of a tonal signal near threshold levels tends to reduce the moment-by-moment amplitude variations for RPN maskers while increasing the moment-by-moment amplitude variations for LNN maskers. Thus, in both cases there is a change in degree of amplitude variations but in different directions. In light of this, the lack of difference between thresholds in RPN and LNN in the no-rove on-frequency condition is somewhat surprising, especially given several previous investigations in which the detection of tonal signals is substantially better for LNN than RPN (e.g., Eddins and Barber, 1998; Hall *et al.*, 1998; Kohlrausch *et al.*, 1997). For both noise types, thresholds were approximately −3 dB. In the previous studies, threshold in the RPN condition was 2–4 dB higher than obtained here. Thus, one explanation for the lack of a threshold difference is relatively good performance in the RPN condition. As discussed below, lower thresholds here may reflect the fact that all 10-Hz noise bands in the present study were adjusted to have the same overall level in each observation interval. Most other masking investigations using RPN have maintained the same average level across bands, resulting in random level variations across bands in any given observation interval due to the random statistics of noise. Interestingly, thresholds for noise signals in the no-rove on-frequency conditions were substantially lower for LNN than RPN maskers, a result consistent with the use of short-term within-channel level cues for detection and the deleterious role of the inherent variability in RPN.

In several conditions, the comparison of thresholds for tonal signals in RPN and LNN can be used to highlight the relatively poor use of within and/or across-channel short-term level cues in LNN. For example, with roving level, thresholds for tonal signals in the on-frequency masker condition were 5 dB lower for RPN than LNN and thresholds for tonal signals were 5 dB lower than for noise signals in either RPN or LNN. Because roving level limits the usefulness of within-channel long-term level cues introduced by the signal, these results demonstrate the importance of the within-channel short-term level cues produced by tonal signals in RPN maskers and illustrate the relatively poor use of those cues in LNN. In fact, Eddins and Barber (1998) showed, via computer simulations, that if detection were based purely on the changes in within-channel short-term level cues produced by a tonal signal at threshold, then thresholds for a tonal signal would be approximately 28 dB *lower* for LNN than RPN in the on-frequency masker condition. In addition to the poor use of within-channel short-term level cues present in LNN, thresholds in the coherent conditions reveal poor use of across-channel short-term level cues in LNN. Regardless of rove, thresholds for tonal signals in the coherent conditions were about 2 dB better for RPN and LNN, despite the fact that short-term across-channel level cues should be stronger in LNN than RPN (according to Eddins and Barber, 1998). The 2-dB difference in the opposite direction could be related to the way in which such across-channel cues are processed. In RPN, the low-amplitude portions of the *maskers* (dips) could cue the times at which the signal-to-noise ratio is maximal, in which case differential weighting favoring across-channel-level cues at those times would provide an advantage for detection (e.g., Buus, 1985). Although the same across-channel level cues may exist in LNN, the flat envelopes of the coherent LNN flanking bands would provide no amplitude dips for such a weighting scheme.

The RPN–LNN comparison reflects points near the two extremes on the continuum of fluctuation depth. Based upon this comparison and similar RPN–LNN comparisons in studies of the masking level difference (MLD) [Eddins and Barber, 1998; Hall *et al.*, 1998; Kohlrausch *et al.*, 1997], it appears that masking release decreases with decreasing fluctuation depth. A number of CMR investigations, using sinusoidally amplitude modulated maskers, have demonstrated a systematic effect of decreasing modulation depth on the magnitude of masking release (e.g., Eddins and Wright, 1994; Fantini and Moore, 1994b; Grose and Hall, 1989). Masking release also decreases with increasing fluctuation rate, as modified by sinusoidal amplitude modulation rate and/or noise bandwidth (e.g., Eddins and Wright, 1994; Fantini *et al.*, 1993; Fantini and Moore, 1994a). In each of these cases, the decrease in masking release depends in part upon improvement in detection in the baseline or reference condition, limiting the possibility for masking release.

B. Effects of stimulus variability

One distinction between typical profile analysis and CMR paradigms is the stable versus dynamic context in which signal detection occurs and the influence of that context on detection thresholds and on masking release. Fantini and Moore (1994a) suggested that across-channel comparisons of level are only useful when the stimulus level is highly variable. When the present conditions are taken into consideration, one may conclude that within-channel as well as across-channel short-term level comparisons are useful when the stimulus level is highly variable. The most robust examples related to within-channel cues and across-channel cues will be used to illustrate this point. For the on-frequency RPN masker with roving level, threshold for the tonal signal is 5 dB lower than threshold for the noise signal (see the circle, column 2, upper panel, Figs. 2 and 3). This difference could only be due to the reduction in short-term amplitude variations resulting from the addition of the tonal signal, thereby providing a within-channel cue related to modulation depth. In terms of short-term across-channel level cues, the difference in detection threshold for noise signals in incoherent and coherent RPN maskers (see the triangles and squares, column 2, Fig. 2) is consistent with the use of a short-term across-channel level comparison. In addition, the stimulus context does not necessarily have to be highly variable for across-channel cues to be useful. For example, a modest masking release was obtained for a tonal signal in LNN without roving level, reflecting the use of across-channel cues related to the change in short-term amplitude fluctuations introduced by the combination of a tonal signal and LNN masker.

A second distinction between CMR and profile experiments is that while CMR experiments generally report thresholds for both on-frequency and incoherent reference conditions, and the influence of roving level is reasonably well known, the typical profile analysis experiment does not

report thresholds for on-frequency conditions, and measurements without roving level are rare. Both of the latter conditions were included in the profile investigation by Spiegel *et al.* (1981), who reported that the addition of sinusoidal flanking bands produced a 3-dB threshold increase relative to the on-frequency sinusoidal reference condition (i.e., negative masking release). Conversely, Green *et al.* (1984) reported a 1.5–2.0-dB improvement as the number of spectral components in the profile stimulus increased from 1 to 3. Green and Mason (1985) later showed that the use of within- and across-channel level cues with deterministic stimuli appears to be highly dependent upon prior listening experience, suggesting an explanation for the discrepant results of Green *et al.* (1984) and Spiegel *et al.* (1981). Green and Mason reported that a group of listeners who were well practiced on multicomponent profile tasks with roving level achieved thresholds of −2.5 dB (signal level *re*: standard level) in a single-component intensity discrimination condition and −5.7 dB (signal level *re*: standard level) in a multicomponent profile condition. For a group of listeners who were well practiced on intensity discrimination for tonal signals in isolation, thresholds were −5.2 dB in the single-component intensity discrimination condition and −3.0 in the multicomponent profile condition. Remarkably, the performances of the two groups were mirror images of each other. Thus, prior experience specifically using within-channel level and across-channel level cues in listening tasks can strongly influence the use of those cues in other related tasks. Prior experience does not seem to be so critical under conditions of high stimulus level uncertainty. Recall that with roving level, thresholds were always lower for the incoherent and coherent conditions than the on-frequency condition and that, even in the absence of rove, thresholds in RPN were consistently lower in the coherent condition than the on-frequency or incoherent condition. The latter result is undisputed in numerous studies of CMR using listeners with varying degrees of prior experience. The issue of prior experience, however, raises the question of whether short-term across-channel level cues would have been more beneficial in the LNN-coherent condition (low stimulus uncertainty) had listeners not had prior experience listening in RPN conditions.

C. Comparisons with previous results

One question of interest is whether the benefits gained from short-term across-channel level cues and short-term across-channel modulation cues are additive. Fantini and Moore (1994b) addressed this question systematically using AM tonal maskers with varying modulation depths and conditions of roving level across intervals and across flanking bands. Their results indicated a large masking release due to short-term across-channel level cues (5–6 dB) and a smaller masking release due to short-term across-channel modulation cues (1–3) dB. The two cues summed to produce 6–9-dB masking release when both types of cues were available. The results of Hall and Grose (1988), using narrow-band noise maskers, showed that both cues give rise to masking release and that combining the cues produced a larger masking release than either cue alone, but that the individual masking releases did not sum. The key to these comparisons is that the overall level of individual flanking bands within the same observation interval was varied. In the present study, this question can only be addressed indirectly by determining whether across-channel level and across-channel modulation cues might combine to produce lower thresholds than one of the cues alone. Consider the RPN-coherent conditions for noise and tonal signals. The noise signal provides across-channel level cues while the tonal signal provides both across-channel level and across-channel modulation cues. The fact that larger masking release was obtained for the tonal signal might be interpreted as suggesting that the two types of cue combine to produce a larger masking release than obtained by level cues alone.

A second question of fundamental concern is the extent to which results obtained with noise maskers can be generalized to other stimulus conditions such as tonal maskers. Fantini and Moore (1994b), who used AM tonal maskers, expressed the same concern in their comparisons of ''profile-like'' and ''CMR-like'' cues. The above discussion indicates that the additivity of stimulus cues such as across-channel level and across-channel modulation cues may be strongly influenced by stimulus type; however, there is strong support from experiments with tonal and noise maskers that the two cues combine to produce larger masking release than either cue alone. Roving level seems to have a much larger influence on tonal than noise maskers in conditions in which there are coherent amplitude fluctuations across frequency. Fantini and Moore (1994b) showed that rove consistently produced increased thresholds in coherent AM tone conditions, whereas the present conditions indicated a large effect for the tone-like LNN masker but not for RPN maskers. The use of RPN conditions provides an added dimension to the notion that high uncertainty relative to stimulus level plays an important role in detection and discrimination. Nevertheless, the results obtained here with RPN continue to support the conclusion of Fantini and Moore (1994b) that across-channel comparisons of level are only useful when within-channel level cues are unreliable. With their tonal stimuli, however, there was no possibility of across-channel cues related to the change in modulation pattern. The masking release obtained for tonal signals in LNN maskers suggests that in the absence of dynamically varying maskers, across-channel *modulation* cues can lead to a masking release. In general, however, comparisons of the present results with noise maskers to those of Fantini and Moore (1994b) using tonal maskers suggests that the results obtained do generalize across stimulus types.

The pattern of thresholds in the on-frequency, incoherent, and coherent conditions with RPN maskers and the size of the masking release obtained is similar to previous investigations using very narrow-band maskers (e.g., Hall and Grose, 1988; Fantini *et al.*, 1993; Fantini and Moore, 1994a, b). However, the actual threshold values reported differ substantially across these various studies. In fact, for noise signals, thresholds here were about 4 dB lower than previously reported. To illustrate threshold differences between these investigations, Table I displays increment detection thresholds (noise signals added in phase) grouped by condition for

TABLE I. Increment detection thresholds for band-pass noise (10-Hz) centered at 2000 Hz (2500 Hz for Eddins) expressed using four different metrics.[2] In each case, the narrowband signal was identical to and added in-phase to the narrowband noise standard. The studies listed in column 1 are Eddins (present data), Fantini and Moore (1994a), Hall and Grose (1988), and Bos and de Boer (1966). Note that the data from Fantini and Moore are actually the average thresholds from 8 and 16 Hz bandwidth conditions. Also listed in column 1, in parentheses, are the masker condition: O=on-frequency, C=coherent, I=incoherent. Asterisks indicate values reported in the original papers.

Study/Condition	$10\log(I/I)$	Signal dB SPL	I in dB (L)	$20\log(A/A)$
Eddins, RPN (O)	−1.32	50.06	2.40	−9.94
Fantini and Moore (O)	1.85*	55.43	4.03	−4.57
Hall and Grose (O)	4.57	59.70*	5.87*	−0.30
Bos and de Boer (O)	5.25*	60.71	6.39	0.71
Eddins, RPN (C)	−4.15*	44.94	1.41	−15.06
Fantini and Moore (C)	−0.20*	52.00	2.91	−8.00
Hall and Grose (C)	−0.20	52.00*	2.91*	−8.00
Eddins, RPN (I)	0.34*	52.92	3.18	−7.08
Fantini and Moore (I)	4.25*	59.21	5.64	−0.79
Hall and Grose (I)	4.84	60.10*	6.07*	0.10

TABLE II. Detection of a tonal signal in band-pass noise. The noise bandwidth was 10 Hz and the center frequency was 2000 Hz (2500 Hz for Eddins). Thresholds are expressed using four different metrics. The studies listed in column 1 are Eddins (present data), Hall and Grose (1988), and Bos and de Boer (1966). Also listed in column 1, in parentheses, are the masker conditions: O=on-frequency, C=coherent, I=incoherent. In the present experiment, tonal signals were added in phase to the central component of the noise standard. In the three previous experiments, thresholds were measured for incoherent addition. To facilitate comparison, the present thresholds (column 2) were converted to threshold values that would have been obtained had the signal been added in random phase with respect to the masker. In parentheses are actual values of 10 log(I/I) for in-phase addition.

Study/Condition	$10\log(I/I)$	Signal dB SPL	I in dB (L)	$20\log(A/A)$
Eddins, RPN (O)	−0.70(−2.90)	59.30	2.67	−2.90
Hall and Grose (O)	1.60	61.60	3.88	1.60
Bos and de Boer (O)	2.00	62.00	4.12	2.00
Eddins, RPN (C)	−6.80(−6.93)	53.20	0.82	−6.93
Hall and Grose (C)	−6.40	53.60	0.90	−6.40
Eddins, RPN (I)	2.70(−0.55)	62.70	4.57	−0.55
Hall and Grose (I)	2.40	62.40	4.37	2.40

the present study as well as the investigations of Bos and de Boer (1966), Hall and Grose (1988), and Fantini and Moore (1994a). In previous studies of increment detection, a variety of metrics has been used with little agreement on a single standard metric. Since each metric uses a slightly different logarithmic scale, it is difficult to make threshold comparisons across studies. To allow comparisons of the present data to other relevant studies of intensity discrimination, the thresholds in Table I are reported using four different metrics.[2] For the interested reader, Grantham and Yost (1982) and Green (1988) have provided detailed reviews of the relations among these measures. Each of the four studies mentioned included a condition in which increment detection was measured for a 10-Hz band of noise (8 and 16 Hz in Fantini and Moore, 1994a), centered at 2000 Hz (2500 Hz in the present study), and having a standard intensity of 60 dB SPL.

The thresholds from the present study are generally the lowest of Table I, followed by those of Fantini and Moore (1994a), Hall and Grose (1988), and Bos and de Boer (1966). In the ''on-frequency'' reference condition, the range of thresholds, in units of $10\log(\Delta I/I)$, was 6 dB across the four studies. Differences among listeners within each of these studies were quite small relative to the differences across the four studies. In addition to the results shown in Table I, Fantini *et al.* (1993) reported increment detection thresholds for similar conditions (10-Hz bandwidth centered at 2000 Hz, 60 dB SPL) that were generally 4 to 10 dB higher than the highest thresholds reported in Table I. Reasons for these high thresholds are not known, although the results of Green and Mason (1985) discussed above suggest that prior experience in intensity discrimination and related tasks may dramatically alter performance in specific stimulus conditions. If prior experience is a critical variable, then one might suspect that rate of learning plays an important role as well. Either of these factors may explain the wide range of values obtained for intensity discrimination for noise signals.

In total, the thresholds from these five studies spanned a range of 12.3 dB on a scale of $10\log(\Delta I/I)$. Careful review did not reveal any major differences from one study to the next, and there is little uncertainty as to the actual measure of intensity discrimination reported in each case. The fact that the present thresholds are the lowest of the five studies may be related to stimulus generation. Although the noise samples were random from trial to trial, the overall level of the 400-ms noise in each interval was adjusted to be 60 dB SPL. In most experiments, however, the long-term average noise level is adjusted to be 60 dB SPL but the actual level during the finite-duration stimulus interval may vary substantially from that average. Thus, in the no-rove conditions, the standard stimulus in the present experiment was more stable than typically encountered when using narrow-band-noise stimuli. This may have led to an improvement in increment detection in the present experiment. Such an explanation cannot account for the substantial differences among the other four studies.

In addition to increment detection, the investigations by Bos and de Boer, Hall and Grose, and the present study also measured thresholds for the detection of a tonal signal in a 10-Hz band of noise centered at 2000 Hz (2500 Hz in the present study). These thresholds are shown in Table II using the same format as Table I. Thresholds are notably more consistent for these tonal signals than for the noise signals from the same experiments (i.e., Table I). This provides some support for the notion that the large variability shown in Table I is specific to increment detection for noise signals.

In the above discussion of potential within- and across-channel cues, it has been assumed that any benefits gained from the presence of flanking bands in the incoherent and coherent conditions were due to across-channel processing. In these multiband conditions, however, one might question whether the combination of on-frequency and flanking bands could be processed by a single auditory channel. The fre-

quency spacing between the center frequencies of the masking bands just below and above the signal band was 500 Hz. With noise bandwidths of 10 Hz, the spacing between bands was 490 Hz, or approximately 1.7 to 1.8 ERB units (Glasberg and Moore, 1990). Clearly, some energy from adjacent masking bands is passed by the auditory channel centered on the signal; however, it is not at all clear whether changes in threshold can be attributed strictly to processing within a single auditory channel. Recently, Verhey and Dau (1996) addressed this issue by comparing behavioral thresholds obtained in two different CMR paradigms to thresholds predicted by a model incorporating a single peripheral channel followed by a multichannel modulation filter bank. Using the bandwidening conditions of Haggard *et al.* (1990) and Schooneveldt and Moore (1989), the single-channel model accounted for much of the observed masking release. This suggests that across-channel processing is not necessary to obtain a robust masking release. Interestingly, using multiband conditions like those of the present study, the single-channel model failed to predict the observed masking release. Only when multiple peripheral channels were included in the model did the predicted thresholds approximate thresholds observed behaviorally. To the extent that the Verhey and Dau filter-bank model accurately reflects auditory processing, the latter finding suggests that the multiband measurements such as those reported here (i.e., with reasonably wide spacing between maskers) are based in large measure on across-channel processes.

V. SUMMARY

Discussions of masking release often focus on potential acoustic cues that might serve as the basis for masking release and on mechanisms or processes by which listeners may take advantage of those cues. Stimulus conditions of the type presented here blur the distinction between profile analysis and CMR paradigms, the acoustic cues typically associated with them, and perceptual mechanisms proposed to underlie them. This is particularly evident in comparison of the tonal versus noise signal. The tonal signal clearly introduced a temporal variation that is more characteristic of the typical CMR paradigm, whereas the noise signal produced a change similar to the increment in the typical profile analysis paradigm. It is easiest to conceptualize typical CMR stimuli in terms of temporal characteristics and typical profile stimuli in terms of spectral characteristics. In reality, the temporal and spectral domains both contain all the relevant information, but it is not clear whether the neural representation of the stimulus is best modeled in terms of a frequency- or time-domain representation of the acoustic stimulus. The principal distinguishing features between the profile analysis and CMR paradigms are the stable (profile analysis) vs dynamic (CMR) contexts for signal detection, the use of random level variations to control for within-channel level cues (profile analysis), and the type of signal to be detected (level increment in the on-frequency component in profile analysis; addition of a tonal signal in CMR). In the present study, each of these features was independently varied by varying the masker type (LNN or RPN), the signal type (level increment or tonal signal), or the presence or absence of interval-by-interval level randomization (rove). The results warrant the following conclusions:

(1) Detection in all conditions reported here can be attributed to short-term within- or across-channel processing. It is unclear whether long-term cues are involved in the any of the present conditions. In conditions in which long-term cues might explain detection (e.g, detection of an increment in a single band of RPN), changing masker or signal type resulted in thresholds that could only be explained on the basis of short-term cues.
(2) Changes in detection threshold produced by the introduction of roving level suggest that both short-term within-channel level cues and short-term within-channel modulation depth/pattern cues may serve as the basis for detection when available. Studies of profile analysis typically control for within-channel level cues; however, studies of CMR rarely attend to such cues though they likely influence detection in many presumed ''across-channel'' conditions.
(3) Comparisons of the detection of tonal signals in LNN and RPN maskers indicate that short-term within- and across-channel cues are not used efficiently in LNN. Reasons for this are not well understood and deserve further investigation. One possibility is that listeners would benefit from specific training on the use of such cues, as is the case for intensity discrimination with tonal signals (e.g., Green and Mason, 1985).
(4) Stimulus uncertainty plays a critical role in demonstrations of masking release. A robust masking release was only observed for stimuli having substantial variability in stimulus level produced either by the randomization of the overall stimulus level or by the inherent level variations characteristic of narrow-band noise.
(5) Systematic manipulation of a variety of experimental variables has led to the identification of multiple cues for detection. The results obtained with stimuli typical of profile analysis and CMR paradigms are mutually consistent when appropriate comparisons can be made. Comparisons with previous profile analysis, CMR, and hybrid experiments also support this conclusion. Collectively these studies imply that the phenomenon known as profile analysis and CMR are based upon the same set of stimulus cues and perhaps the same perceptual processes.

ACKNOWLEDGMENTS

The author wishes to thank the two reviewers, Dr. Emily Buss and Dr. Deborah Fantini, as well as the Associate Editor, Dr. Joseph W. Hall III, for their many helpful comments and suggestions. This work was supported in part by Grant No. RO1 DC04403 from NIH NIDCD.

[1]The data of Hall and Grose (1988) were originally expressed as signal level in dB SPL (their Table II) and ΔI in dB (their Table III). For comparison with the present data and to those of Fantini and Moore (1994a) and Bos and de Boer (1966), the thresholds reported by Hall and Grose (1988) have been transformed to units of $10 \cdot \log_{10}(\Delta I/I)$. This scale corresponds to Table 3-1, column 4 of Green (1988) and Table II, column 4 of Grantham and Yost (1982).

[2]Table I values were computed using the following equations taken from Grantham and Yost (1982) assuming a signal voltage, V_s, a masker voltage, V_m, and a correlation, ρ, between signal and masker (here, $\rho = 1.0$):

$$10 \cdot \log_{10}(\Delta I/I) = 10 \cdot \log_{10}(V_s^2/V_m^2 + 2V_s/V_m \cdot \rho),$$

$$\text{Signal dB SPL} = 20 \cdot \log_{10}(V_s),$$

$$\Delta I \text{ in dB}(\Delta L) = 10 \cdot \log_{10}(V_s^2/V_m^2 + 2V_s/V_m \cdot \rho + 1),$$

$$20 \cdot \log_{10}(\Delta A/A) = 20 \cdot \log_{10}(V_s^2/V_m^2).$$

Bos, C. E., and de Boer, E. (**1966**). ''Masking and discrimination,'' J. Acoust. Soc. Am. **39**, 708–715.

Buus, S. (**1985**). ''Release from masking caused by envelope fluctuations,'' J. Acoust. Soc. Am. **78**, 1958–1965.

Eddins, D. A., and Barber, L. A. (**1998**). ''The influence of stimulus envelope and fine structure on the binaural masking level difference,'' J. Acoust. Soc. Am. **103**, 2578–2589.

Eddins, D. A., and Wright, B. A. (**1994**). ''Comodulation masking release for single and multiple rates of envelope fluctuation,'' J. Acoust. Soc. Am. **96**, 3432–3442.

Fantini, D. A., Moore, B. C. J., and Schooneveldt, G. P. (**1993**). ''Comodulation masking release as a function of type of signal, gated or continuous masking, monaural or dichotic presentation of flanking bands, and center frequency,'' J. Acoust. Soc. Am. **93**, 2106–2115.

Fantini, D. A., and Moore, B. C. J. (**1994a**). ''Profile analysis and comodulation detection differences using narrowbands of noise and their relation to comodulation masking release,'' J. Acoust. Soc. Am. **95**, 2180–2191.

Fantini, D. A., and Moore, B. C. J. (**1994b**). ''A comparison between the effectiveness of across-channel cues available in comodulation masking release and profile analysis tasks,'' J. Acoust. Soc. Am. **96**, 3451–3462.

Fletcher, H. (**1940**). ''Auditory patterns,'' Rev. Mod. Phys. **12**, 47–65.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.

Grantham, D. W., and Yost, W. A. (**1982**). ''Measures of intensity discrimination,'' J. Acoust. Soc. Am. **72**, 406–410.

Green, D. M. **(1988)**. *Profile Analysis: Auditory Intensity Discrimination* (Oxford, New York).

Green, D. M., and Mason, C. R. (**1985**). ''Auditory profile analysis: Frequency, phase, and Weber's law,'' J. Acoust. Soc. Am. **77**, 1155–1161.

Green, D. M., Mason, C. R., and Kidd, G. (**1984**). ''Profile analysis: Critical bands and duration,'' J. Acoust. Soc. Am. **75**, 1163–1167.

Grose, J. H., and Hall, J. W., III (**1989**). ''Comodulation masking release using SAM tonal complex maskers: Effects of modulation depth and signal position,'' J. Acoust. Soc. Am. **85**, 1276–1284.

Haggard, M. P., Hall, J. W., III, and Grose, J. H. (**1990**). ''Comodulation masking release as a function of bandwidth and test frequency,'' J. Acoust. Soc. Am. **88**, 113–118.

Hall, J. W., III, and Grose, J. H. (**1988**). ''Comodulation masking release: Evidence for multiple cues,'' J. Acoust. Soc. Am. **84**, 1669–1675.

Hall, J. W., III, Grose, J. H., and Hartmann, W. M. (**1998**). ''The masking-level difference in low-noise noise,'' J. Acoust. Soc. Am. **103**, 2573–2577.

Hall, J. W., III, Haggard, M. P., and Fernandes, M. A. (**1984**). ''Detection in noise by spectro-temporal pattern analysis,'' J. Acoust. Soc. Am. **76**, 50–56.

Hartmann, W. M., and Pumplin, J. (**1988**). ''Noise power fluctuations and the masking of sine signal,'' J. Acoust. Soc. Am. **83**, 2277–2289.

Hartmann, W. M., and Pumplin, J. (**1991**). ''Periodic signals with minimal power fluctuations,'' J. Acoust. Soc. Am. **90**, 1986–1999.

Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A. J., and Puschel, D. (**1997**). ''The detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations,'' Acta Acust. (China) **83**, 659–669.

Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Moore, B. C. J., Glasberg, B. R., and Schooneveldt, G. P. (**1990**). ''Across-channel masking and comodulation masking release,'' J. Acoust. Soc. Am. **87**, 1683–1694.

Pumplin, J. (**1985**). ''Low-noise noise,'' J. Acoust. Soc. Am. **78**, 100–104.

Schooneveldt, G. P., and Moore, B. C. J. (**1989**). ''Comodulation masking release (CMR) as a function of masker bandwidth, modulator bandwidth, and signal duration,'' J. Acoust. Soc. Am. **85**, 273–281.

Spiegel, M. F., Picardi, M. C., and Green, D. M. (**1981**). ''Signal and masker uncertainty in intensity discrimination,'' J. Acoust. Soc. Am. **70**, 1015–1019.

Verhey, J., and Dau, T. (**1996**). ''Modeling comodulation masking release (CMR) effects with a single-channel analysis,'' J. Acoust. Soc. Am. **100**, 2625(A).

Measurement of auditory temporal processing using modified masking period patterns[a)]

David A. Eddins[b)]
Department of Communicative Disorders and Sciences, Center for Hearing and Deafness, State University of New York at Buffalo, Buffalo, New York 14214

(Received 18 January 1998; revised 3 January 2001; accepted 8 January 2001)

A common metric of auditory temporal processing is the difference in the threshold for a pure-tone signal masked by either unmodulated or amplitude-modulated noise. This technique may be viewed as a modification of the masking period pattern technique. Such measurements have been proposed as an efficient means of estimating auditory temporal resolution in a clinical setting, although in many cases threshold differences may reflect additional spectro-temporal processes. The primary purpose of the present experiment was to examine interactions among signal frequency and masker bandwidth and the effects of modulation frequency on modified masking period patterns. The results revealed unmodulated–modulated threshold differences that increased with increasing masker bandwidth and decreased with increasing modulation frequency. There was little effect of signal frequency for narrow-band noise maskers that were equal in absolute bandwidth across frequency. However, unmodulated−modulated threshold differences increased substantially with increasing signal frequency for bandwidths proportional to the signal frequency and for wideband maskers. Although the results are interpreted in terms of a combination of both within-channel and across-channel cues, the specific contributions of these cues in particular conditions are difficult to ascertain. Because modified masking period patterns depend strongly upon a number of specific stimulus parameters, and because it is difficult to determine with any precision the underlying perceptual processes, this technique is not recommended for use as a clinical measure of auditory temporal processing. [DOI: 10.1121/1.1356024]

PACS numbers: 43.66.Ba [JWH]

I. INTRODUCTION

Auditory temporal resolution is the ability of the auditory system to follow or resolve the temporal patterns characteristic of acoustic stimuli and is a fundamental aspect of auditory perception. In 1980, Zwicker proposed a simple, efficient test of temporal resolution that might be suitable for clinical use. He suggested a modification of the masking period pattern paradigm in which threshold for a long-duration tonal signal (e.g., 500 ms) is measured in the presence of either unmodulated or sinusoidally amplitude-modulated, octave-band masking noise. It was reasoned that the decibel difference between threshold measured in these two masking conditions reflects the modulation depth preserved by the auditory system, and thus temporal resolution. On this basis, it was concluded that temporal resolution could be derived at any frequency region with two threshold measurements (i.e., masked threshold for a sinusoid in unmodulated and amplitude modulated noise). From a clinical point of view, this technique is attractive because of its efficiency and the important role temporal processing plays in the perception of speech and most other sounds. In order for the modified masking period pattern to be useful clinically, however, there must be a clear understanding of how of changes in the most basic acoustic properties of the stimuli influence the resulting measures for normal-hearing listeners (e.g., signal frequency, masker bandwidth, modulation frequency). In addition, the validity of any procedures to be used clinically will be dependent upon identification of the perceptual process(es) being measured and the establishment of appropriate normative data. These requirements have not been achieved to date.

Despite numerous investigations of modified masking period patterns over the last two decades, the influence of stimulus frequency and masker bandwidth on the threshold differences remains unclear. For example, Buunen (1976) and Eddins (1993b) used broadband noise maskers and reported much smaller differences between unmodulated and modulated thresholds at low signal frequencies than high signal frequencies. Zwicker (1980), on the other hand, used octave-band noises and reported no influence of stimulus frequency. Using narrow-band noise maskers with equal bandwidths centered at different frequencies, some investigations have reported frequency effects (e.g., Haggard *et al.*, 1990; Bacon *et al.*, 1997; Bacon and Lee, 1997) while others have not (e.g., Hall *et al.*, 1993; Eddins, 1993b). Contributing to the differences among these and other investigations of the detection of a tone in unmodulated and amplitude-modulated noise are differences in modulated type, modulator frequency, and masker level. In addition, many pervious investigations included only a few stimulus conditions, limiting the ability to draw general conclusions. The wide range of results has been accompanied by an equally wide range of explanations. The present investigation provides a parametric evaluation of the influence of signal frequency and masker

[a)]Portions of this research were presented at the 131st meeting of the Acoustical Society of America, State College, Pennsylvania, June, 1997 [J. Acoust. Soc. Am. Suppl. 1 **101**, 3148 (1997)].
[b)]Electronic mail: deddins@buffalo.edu

bandwidth on modified masking period patterns using stimuli presented at moderate sensation levels. At these levels, parametric studies of temporal resolution using the amplitude modulation detection and temporal gap detection techniques have indicated little influence of stimulus frequency region on estimates of temporal acuity for stimuli presented at moderate sensation levels (Eddins *et al.*, 1992; Eddins, 1993a, 1999; Strickland and Viemeister, 1997).

Though originally proposed as a measure of auditory temporal resolution, many authors have noted that, in addition to temporal acuity, the threshold differences obtained in the modified masking period pattern technique may reflect across-channel temporal processes such as those underlying comodulation masking release (CMR). Within the framework of critical band theory, masked thresholds are determined primarily by the masker energy within a narrow range of frequencies, or auditory channel, surrounding the signal. Hall *et al.* (1984) demonstrated that this general result does not hold for amplitude-modulated (AM) noises, and concluded that improved detection with increasing AM masker bandwidth was a result of across-channel temporal processes. Investigations using the modified masking period pattern technique typically have included masker bandwidths considerably wider than estimates of a single critical band, in which case the resulting masking patterns likely reflected both within-channel and across-channel detection processes. This brings into question the validity of the technique as a measure of temporal resolution. To the extent that across-channel temporal processing contributes to the modified masking period pattern, the technique also reflects frequency resolution and more complex spectro-temporal analyses.

Thus, the purpose of the present study was to investigate auditory temporal processing as reflected in modified masking period patterns with a specific focus on the influence of stimulus frequency and masker bandwidth. In an attempt to resolve the apparent discrepancy within and among numerous investigations to date, stimulus parameters included a range of modulation frequencies, comparisons of equivalent absolute and equivalent relative masker bandwidths across frequency, and subcritical as well as supracritical bandwidths at each frequency.

II. METHOD

A. Subjects

Five listeners participated, ranging in age from 19 to 25 years. All had pure-tone thresholds of 15 dB HL or better at octave frequencies from 250 to 8000 Hz and normal middle-ear function based on a screening tympanogram (Y, 226 Hz). None of the listeners had previous experience in psychoacoustic listening tasks. Each listener was paid an hourly wage and received a 10% bonus upon completion of the study.

B. Stimuli

Pure-tone signals were 500, 1000, 2000, and 4000 Hz. Noise maskers were either unmodulated or sinusoidally amplitude modulated with a modulation depth of $m=1.0$ and modulation frequencies of either 4, 16, or 64 Hz. Masker conditions included four fixed-bandwidth conditions, 200, 400, 800 Hz wide or low-pass filtered at 6000 Hz (denoted wideband), and two relative bandwidth conditions, 1 and 3 ERBs wide, where ERB is the equivalent rectangular bandwidth estimated by Eq. 3 of Glasberg and Moore (1990) to be 79, 133, 241, and 456 Hz at 500, 1000, 2000, and 4000 Hz, respectively. Amplitude-modulated maskers had the restriction that the modulation frequency not exceed one-half of the stimulus bandwidth. Thus, for the 1-ERB (79 Hz) masker at 500 Hz, only modulation frequencies of 4 and 16 Hz were included.

All stimuli were generated using a digital signal processing board (TDT AP2) and 16-bit D/A converters (TDT DA3) with a sampling period of 50 μs. Pure-tone signals were generated in the time domain and were in random phase with respect to the masker. Noise maskers of the desired bandwidth were generated in the frequency domain by filling portions of two 16384-point buffers with the appropriate Rayleigh-distributed magnitudes or uniformly distributed phases (0 to 2π). For amplitude-modulated maskers, sidebands were created in a similar manner, such that when added to the carrier, amplitude modulation was produced. For each Fourier component of the carrier, upper $(f+fm)$ and lower $(f-fm)$ side components were produced with magnitudes 0.5 times the magnitude of the corresponding carrier component (corresponding to $m=1.0$). The side-component phase equaled that of the corresponding carrier. A subsequent inverse FFT (fast Fourier transform) on the complex buffer pair yielded a noise waveform having the desired temporal and spectral characteristics. The above technique has been described in detail by Strickland and Viemeister (1997) and is commonly used in this laboratory in studies of amplitude modulation detection where the standard is a QFM noise (e.g., Eddins, 1999). All maskers had a pressure-spectrum level of 50 dB SPL within the nominal stimulus passband. The 400-ms signal was temporally centered within the 800-ms masker. Both were gated on and off with 10-ms $\cos^2$ rise-fall envelopes.

C. Procedures

An adaptive, two-interval, two-alternative forced-choice method with a three-down, one-up tracking algorithm was used to estimate the signal level corresponding to 79.4% correct detection (Levitt, 1971). The initial step size of each 60-trial block was 5 dB and was reduced to 2 dB after the first three reversals of the adaptive track. Threshold was estimated by averaging the signal levels of each of the last even number of reversals, excluding the first three. Three such estimates were obtained for each condition. A fourth threshold estimate was obtained if the standard deviation among the first three exceeded 2.5 dB. A fourth run was collected on 6.5% (31 of 475) of the conditions. The final threshold for a condition was taken as the average threshold across the three or four estimates. Each trial consisted of two 810-ms observation intervals separated by 400 ms, marked by LEDs, and followed by visual feedback. Stimuli were presented via Etymotic ER-2 insert earphones and testing was conducted in a sound-attenuating chamber.

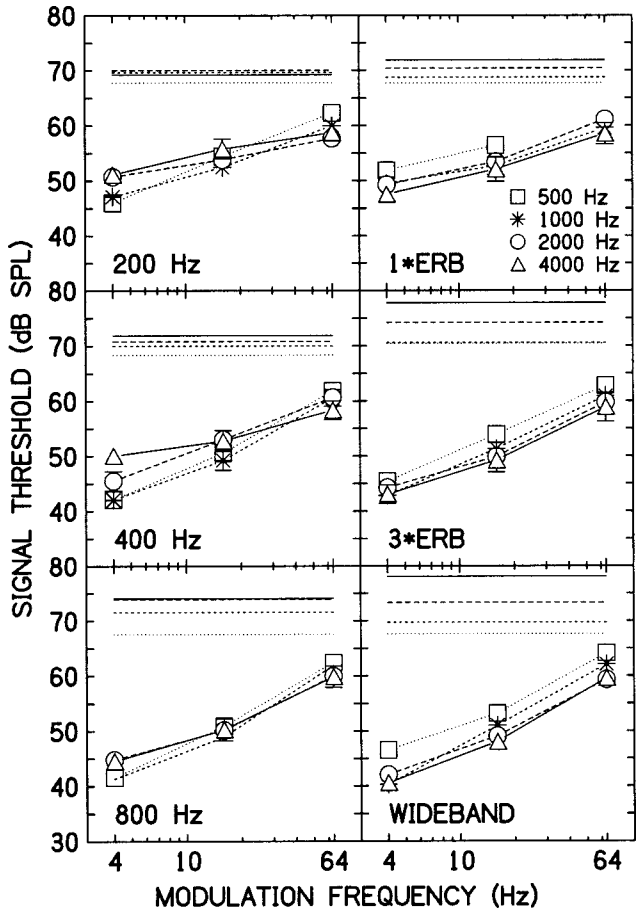


FIG. 1. Mean signal threshold for five normal-hearing listeners as a function of masker modulation rate. Each panel represents a different bandwidth condition. Signal thresholds in unmodulated maskers are indicated by horizontal lines for each signal frequency. Thresholds for modulated maskers ($m=1.0$) at four signal frequencies are indicated by the symbols.

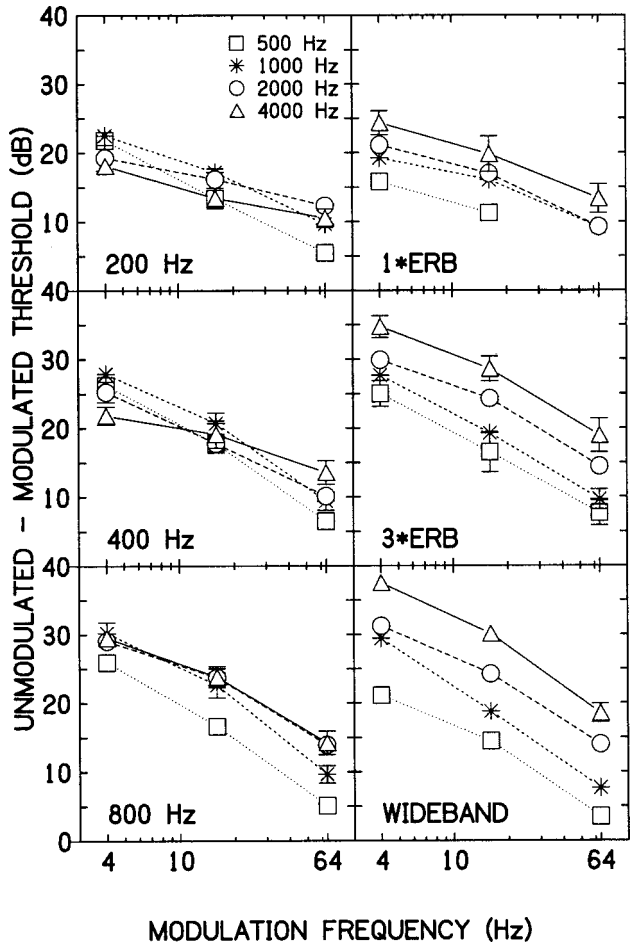


FIG. 2. Modified masking period patterns. Unmodulated minus modulated threshold differences plotted against masker modulation frequency. Each panel represents one of the six bandwidth conditions. Symbols indicate the four signal frequencies.

III. RESULTS

Mean signal thresholds for five normal-hearing listeners are shown in Fig. 1 as a function of masker modulation frequency. Each panel represents a different bandwidth condition. For the unmodulated maskers, signal thresholds are indicated by horizontal lines for each signal frequency. For modulated maskers, thresholds at the four signal frequencies are indicated by symbols. First note the general pattern of results in each panel. Thresholds for the unmodulated conditions (horizontal lines) increase with increasing signal frequency. Since the masker spectrum level was 50 dB SPL in each condition, this systematic increase largely reflects the increasing masker energy within the auditory channel centered on the signal frequency (e.g., Fletcher, 1940). Thresholds in the modulated conditions show a clear and systematic increase with increasing modulation frequency, consistent with many other studies of the detection of a tone masked by sinusoidally amplitude-modulated noise (e.g., Rodenburg, 1972; Buunen, 1976; Carlyon *et al.*, 1989). For a given absolute masker bandwidth, thresholds in modulated maskers did not vary systematically across the four signal frequencies. For the 1- and 3-ERB conditions, thresholds were slightly higher at 500 Hz than the three higher frequencies. Thresholds generally decreased with increasing masker bandwidth, particularly at low modulation frequencies.

Next consider the modified masking period patterns shown in Fig. 2, which are derived from the data presented in Fig. 1. Here, unmodulated minus modulated threshold is shown on the ordinate with modulation frequency on the abcissa. As in Fig. 1, each panel represents a different bandwidth condition and symbols indicate the four signal frequencies. The unmodulated−modulated difference depends strongly on modulation frequency, being greatest for low modulation frequencies and decreasing with increasing modulation frequency. Again this is consistent with many previous studies (Rodenburg, 1972; Buunen, 1976; Weber, 1977; Scott and Humes, 1990; Carlyon *et al.*, 1989; Eddins, 1993b). For the narrow-band, fixed-absolute bandwidth conditions (left column), there is little systematic variation in these modified masking period patterns with signal frequency, with the exception of the 800-Hz masker centered at 500 Hz (lower left panel), which produced slightly smaller differences at all modulation frequencies. A three-way analysis of variance computed for the fixed, narrow-band maskers (200, 400, 800 Hz bandwidth) indicated significant main effects of signal frequency ($F_{2,144}=55.6$, $p=0.01$), masker bandwidth ($F_{3,144}=14.6$, $p<0.01$), and modulation fre-

quency ($F_{2,144}=420.5$, $p<0.01$). *Post hoc* analyses (Tukey HSD) revealed, however, that unmodulated-modulated threshold differences were not significantly different at 1000, 2000, and 4000 Hz, but that threshold differences were significantly smaller at 500 Hz than at higher signal frequencies. Similar *post hoc* analyses revealed significant differences among each of the bandwidth conditions and each different modulation frequency. Specifically, the unmodulated-modulated difference increased with increasing masker bandwidth. There appears to be little influence of stimulus frequency on the modified masking period patterns for narrow-band maskers equal in absolute bandwidth.

A different pattern of results was obtained for the 1-ERB, 3-ERB, and wideband conditions (Fig. 2, right column). In each case, there is a systematic increase in the size of the unmodulated-modulated difference with increasing signal frequency from 500 to 4000 Hz. This separation by signal frequency becomes more pronounced as masker bandwidth increases. A three-way analysis of variance computed for the wideband and two relative bandwidth conditions indicated significant main effects of signal frequency ($F_{2,96}=84.8$, $p<0.01$), masker bandwidth ($F_{3,96}=80.1$, $p<0.01$), and modulation frequency ($F_{1,96}=132.5$, $p<0.01$). Here only the 4- and 16-Hz modulation frequencies were included since the 64-Hz frequency was not tested in the 1-ERB condition at 500 Hz. *Post hoc* analyses (Tukey HSD) revealed significant differences among all signal frequency comparisons, bandwidth comparisons, and modulation frequency comparisons. Note that the strong signal frequency effect for the 1-ERB, 3-ERB, and wideband conditions is dominated by the dependence of threshold in the unmodulated masker conditions upon signal frequency, as seen in Fig. 1 (right column). Changes in threshold with signal frequency in the modulated masker conditions are smaller and less systematic than in the unmodulated masker conditions. Implications of this relation with respect to estimates of across-channel processing are addressed later.

IV. DISCUSSION

A. Temporal resolution

Zwicker (1980) suggested that unmodulated-modulated threshold differences (e.g., Fig. 2 above) may be interpreted as a measure of temporal resolution, reflecting the ability of the ear to preserve the modulation depth of the amplitude-modulated masker. Following the statistical analyses reported here for the 200-, 400-, and 800-Hz bandwidth conditions, one might conclude that temporal resolution depends little on signal frequency between 1000 and 4000 Hz and is slightly poorer below 1000 Hz. Such a conclusion would be generally consistent with estimates of temporal resolution based upon amplitude modulation detection (e.g., Eddins, 1999; Strickland and Viemeister, 1997). Of course, consideration of the broadband and relative bandwidth conditions would lead to just the opposite conclusion.

Rather than estimating temporal resolution on the basis of the threshold difference between the two masker conditions (e.g., Zwicker, 1980), other investigators have computed estimates of temporal resolution from modified masking period patterns by deriving modulation transfer functions from the threshold data (e.g., Scott and Humes, 1990). This three-step process often involves (1) deriving a modulation transfer function by computing the modulation depth preserved by the auditory system at each modulation frequency, based upon the unmodulated-modulated threshold difference, (2) obtaining the best-fitting low-pass filter function to the derived modulation transfer function, and (3) computing a time constant corresponding to the fitted low-pass function. This form of data reduction is useful when a sufficient number of modulation frequencies are used to define the transfer function, but produces unstable results when only three modulation frequencies per condition are used. Since the three-point modulation transfer functions derived from modified masking period patterns in Fig. 2 were highly linear, computing the slope of the best-fitting straight line provides a similar form of data reduction. In this case, a steeper slope resulting from a linear fit would correspond to a longer time constant based upon a low-pass filter fit. For each masker bandwidth, the obtained slopes were similar for signal frequencies of 2000 and 4000 Hz but were substantially steeper as signal frequency decreased from 2000 to 500 Hz. A similar result was obtained for the wideband and 3-ERB maskers, whereas the slope for the 1-ERB masker varied little below 4000 Hz. Based upon these results, one might conclude that estimates of temporal resolution derived from modified masking period patterns improve with increasing signal frequency. Is seems that estimates of temporal resolution based upon modified masking period patterns depend upon how one interprets the threshold data.

B. Traditional CMR interpretation

A number of investigators have suggested that masked signal detection often depends upon *both* within-channel temporal processing and across-channel processing. Conditions in which the masker bandwidth is less than or equal to the ERB at the signal frequency of interest primarily reflect within-channel processes, whereas conditions wider than an ERB may reflect a combination of within and across-channel processes. Thus, modified masking period patterns may represent more complex spectro-temporal processing than would be involved in a temporal resolution task. Carlyon *et al.* (1989) suggested that to estimate the contribution of across-channel temporal processes, one could subtract the unmodulated-modulated differences for a 1-ERB-wide masker from the unmodulated-modulated difference for a wider bandwidth condition. Presumably the remainder would reflect across-channel processes. Carlyon *et al.* termed the remainder a comodulation masking release (CMR), referring to the coherence of the amplitude modulation across auditory channels and to the release from masking as the masker bandwidth is increased. Note that these estimates are derived from differences in unmodulated-modulated differences, and thus depend upon thresholds in both masker conditions. A number of investigations of CMR have used a similar rationale (e.g., Grose *et al.*, 1993; Hall *et al.*, 1993; Schooneveldt and Moore, 1989). To compare the present results to previous investigations, a similar analysis is shown in Fig. 3. For convenience, the ordinate is labeled CMR. Each row repre-

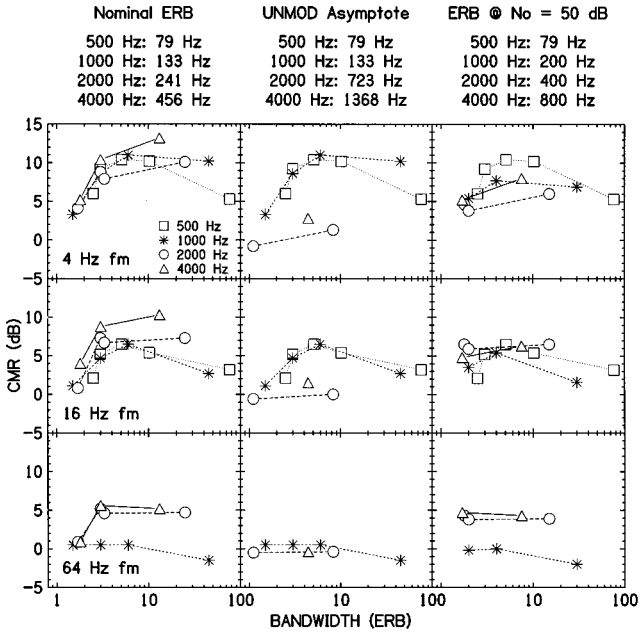


FIG. 3. Across-channel processing, denoted CMR, as a function of masker bandwidth in ERB units. CMR was estimated by subtracting the unmodulated−modulated difference for a masker approximately equal to one equivalent rectangular bandwidth (1 ERB) from the unmodulated −modulated difference for maskers wider than 1 ERB (given on the abscissa). The data presented in each column represent different methods of estimating ERBs (see text). ERB values are given at the top of each column. Each row represents a different masker modulation frequency and symbols indicate signal frequency.

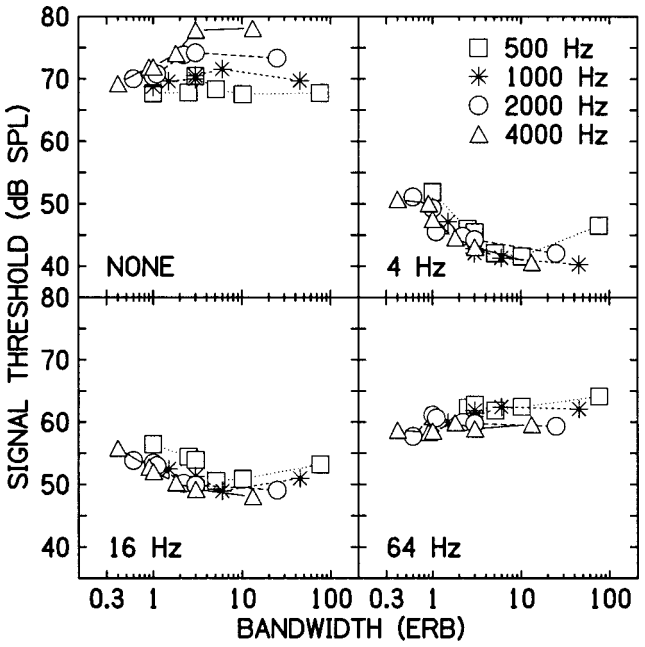


FIG. 4. Signal thresholds, replotted from Fig. 1, with masker bandwidth in ERB units (as specified in Sec. II) on the abscissa and signal frequency indicated by the symbols. Each panel represents a different masker modulation frequency.

sents a different modulation frequency and symbols within a panel indicate the four signal frequencies. Notice that the abscissa represents masker bandwidth in ERB units. First consider the three panels in the left column of the figure (the center and right columns are discussed below). Here, CMRs are based upon the nominal ERBs stated in Sec. II, which were based upon Eq. (3) of Glasberg and Moore (1990). To illustrate specific values on the abscissa, consider the left column and a 500-Hz signal frequency. In this case, a value of 1 ERB corresponds to a masker bandwidth of 79 Hz and a value of 10 ERB corresponds to a masker bandwidth of 790 Hz. The wideband conditions are shown as 6000 Hz divided by the ERB at a particular frequency (e.g., in left column, at 500 Hz, 6000/79=75.9 ERBs).

The CMRs indicate that across-channel contributions to detection depend upon masker bandwidth, masker modulation frequency, and signal frequency. In terms of the masker bandwidth, the general pattern of results in each panel in the left column reveals a rapid increase in the across-channel contribution with increasing masker bandwidth, reaching a maximum at bandwidths roughly three to six times the ERB. A comparison across the three modulation frequencies (upper, middle, and lower panels) reveals an inverse relation between modulation frequency and the magnitude of across-channel contribution. This inverse relation between modulation frequency and CMR is consistent with a number of previous investigations using various masker stimuli (e.g., Hall *et al.*, 1988; Eddins and Wright, 1994), although several studies have not reported a systematic change in CMR with modulation frequency (e.g., Carlyon *et al.*, 1989; Schooneveldt and Moore, 1989). Finally, the CMRs in the left column are similar across signal frequency, although the maximum masking release tends to be largest at the 4000-Hz signal frequency, most notably for the 16-Hz modulation rate (middle panel). Curiously, at the 64-Hz modulation rate (lower panel), there was no apparent across-channel contribution for a signal frequency of 1000 Hz (open circles). While there was a substantial threshold difference between unmodulated and modulated maskers at 1000 Hz, that difference was nearly the same for all bandwidths from 1 ERB (130 Hz) to 6000 Hz wide.

Following the format used in Fig. 3, the signal thresholds displayed in Fig. 1 may be replotted with signal threshold on the ordinate and masker bandwidth, in ERB units, on the abscissa. Figure 4 displays thresholds in such a manner with each panel representing a different modulation frequency (none, 4, 16, and 64 Hz) and the symbols indicating the signal frequency. In Fig. 4, several interesting features in the data are apparent. For example, notice in the upper left panel of Fig. 4 (unmodulated maskers) that signal thresholds at 2000 and 4000 Hz (circles and triangles) clearly increase with increasing bandwidth beyond one (nominal) ERB. An interpretation along the lines of the traditional critical band experiment of Fletcher (1940) would indicate that actual critical bands (or their ERBs) may be substantially wider than the nominal ERBs of 241 and 456 Hz at 2000 and 4000 Hz, respectively. If that is the case, then the CMRs shown in the left column of Fig. 3 may underestimate the contribution of within-channel cues at these frequencies, and a more accurate CMR estimate might be obtained by using the asymptotic threshold levels in unmodulated noise. The ERBs so estimated are 723 and 1368 Hz at a signal frequencies of 2000 and 4000 Hz (in each case, corresponding to the nominal 3-ERB bandwidth). In the middle column of Fig. 3,

CMRs have been recomputed using the latter ERB estimates for 2000 and 4000 Hz. For comparison, CMRs for the signal frequencies of 500 and 1000 Hz are replotted from the left column. At 2000 and 4000 Hz, we see only modest CMRs of 1–3 dB for the 4- and 16-Hz modulation frequencies and no CMRs at a modulation frequency of 64 Hz. Thus the signal-frequency effects are more robust than noted in the left column. ERBs based upon asymptotic threshold levels suggest a much greater influence of within-channel cues (and less influence of across-channel cues) than ERBs based upon the estimates specified earlier.

The notion that the nominal ERBs specified in Sec. II underestimate within-channel cues available in the present experiment is supported by previous studies, such as Rosen and Baker (1994), that have shown ERB estimates to depend upon masker spectrum level. In fact, the ERB estimates provided by Glasberg and Moore [1990, Eq. (3)], which were the basis for the 1- and 3-ERB bandwidths in this experiment, were based upon masker spectrum levels of 30 dB SPL, whereas the masker spectrum level used here was 50 dB SPL. To compensate for the differences in level, ERB estimates have been recomputed to reflect this higher spectrum level following the procedure outlined by Glasberg and Moore (see their Eq. 5). The level-corrected ERBs are 102, 178, 338, and 674 Hz for signal frequencies of 500, 1000, 2000, and 4000 Hz. CMRs based on the level-corrected ERBs are shown in the right column of Fig. 3. For the 4-Hz modulation frequency, CMRs are slightly smaller at 2000 Hz than other signal frequencies. At 16 Hz, the frequency effect is not as obvious, except for the wideband conditions with a signal frequency of 500 and 1000 Hz. At 64 Hz, CMRs are again lower at 1000 Hz than 2000 or 4000 Hz, similar to the left-most column. Looking across columns, it is clear that regardless of how the ERB is defined, CMR estimates clearly depend upon modulation frequency and signal frequency.

The masker bandwidth effects revealed in the modified masking period patterns may reflect changes in either the unmodulated or modulated masker conditions. When the masker bandwidth is less than 1 ERB at all signal frequencies, the effective masker energy and inherent masker fluctuation rate are essentially the same at all signal frequencies, both being determined by the masker level and bandwidth. Thus, threshold should change little with signal frequency for unmodulated or modulated maskers. Increasing the subcritical bandwidth should result in higher thresholds with the unmodulated masker, due to increased masking energy. A similar increase in bandwidth for the modulated masker would result in increased masking energy but also might result in better coding of the imposed amplitude modulation, which would tend to lower masked threshold. Support for the latter claims may be seen in Fig. 4 in comparison of thresholds for the 200- and 400-Hz bandwidths centered at 4000 Hz, where both bandwidths are less than 1 ERB for each of the estimates above. Threshold increased by 3 dB with increasing bandwidth for the unmodulated masker noise (presumably due to increased masker energy), whereas threshold decreased by 1.5 dB for the 4-Hz modulated noise. Despite increased masker energy, threshold for modulated noise decreased, presumably due to improved coding of amplitude modulation. This notion is supported by studies of amplitude modulation detection that have shown improved modulation detection thresholds with increasing noise bandwidth for subcritical bandwidths (e.g., Dau *et al.*, 1997; Eddins, 1999; Strickland and Viemeister, 1997).

For masker bandwidths equal to 1 ERB, increases in signal frequency result in increases in absolute masker bandwidth. Thresholds obtained with unmodulated maskers should increase with increasing signal frequency, again due to increased masking energy. Similar increases in signal frequency and masker bandwidth for the modulated masker would result in increased masking energy but perhaps better coding of the imposed amplitude modulation, leading to either a small increase or a decrease in masked threshold. Therefore, thresholds for unmodulated and modulated maskers would diverge with increasing signal frequency, resulting in larger unmodulated-modulated threshold differences with increasing signal frequency. In fact, these observations would hold for any maskers equal in relative bandwidth across frequency and equal to or less than 1 ERB wide, and appear to hold for each of the three ERB estimates outlined above. For the present data, we see in Fig. 4 that thresholds in unmodulated maskers for bandwidths from 1 ERB (minimum estimate of true ERB) to roughly 3 ERBs (maximum estimate of true ERB) increase with increasing signal frequency. For the same bandwidths, we see that thresholds in modulated maskers remain constant or decrease with increasing signal frequency. Thus, in Fig. 2, the unmodulated-modulated threshold differences for the 1- and 3-ERB maskers increase with increasing frequency from 500 to 4000 Hz. These observations by definition are considered within-channel effects.

When the masker is wider than 1 ERB, modified masking period patterns may reflect **both** within-channel effects and across-channel temporal processing. Here, it is assumed that any increase in the unmodulated-modulated difference with increasing bandwidth beyond 1 ERB reflects the similarity between the temporal features of the stimulus extracted from spectral regions proximal to and distal from the signal frequency. Psychophysically, this is most often demonstrated in CMR paradigms (e.g., Hall *et al.*, 1984). For the bandwidths of 200, 400, and 800 Hz, a greater proportion of the masker falls within the ERB centered on the signal as signal frequency increases, in which case a smaller proportion can contribute to across-channel processes. The masking patterns for these bandwidths (left column, Fig. 2) are consistent with the interpretation that this reciprocal relation between the availability of within- and across-channel energy yielded similar unmodulated-modulated differences at different signal frequencies, and thus little signal frequency effect. This interpretation is also consistent with measures of the detection of amplitude modulation using bandlimited noise. Eddins (1993a, 1999) and Strickland (2000) have shown that the low-pass characteristic of the temporal modulation transfer function is independent of stimulus frequency when the noise carrier bandwidth is held constant across audio frequency and stimuli are presented at moderation sensation levels. Eddins explained these results in terms of an integration of the temporal variations across multiple auditory chan-

nels that involves a trade between the width of those channels, which determines the preserved modulation depth within each channel, and the number of auditory channels stimulated. Recently, Dau *et al.* (1997) replicated the results of Eddins (1993a) and presented a model of auditory processing that accurately predicted such spectral integration of temporal information.

This apparent trading relation did not hold for the wideband maskers of the present experiment. One possible explanation is that the influence of within-channel processes continued to increase as the signal frequency and ERB increased, whereas improvement in detection due to across-channel processes reached an asymptote at some masker bandwidth. Combined, this would result in unmodulated-modulated differences that continually increase with increasing signal frequency, Referring to the data in Fig. 3, the estimated CMRs asymptote (or peak) at roughly three to six times the ERB at a given test frequency. This asymptotic performance is consistent with previous investigations of CMR using bandwidening techniques (e.g., Schooneveldt and Moore, 1989) and is consistent, in principle, with investigations of CMR using multi-band maskers, which have shown asymptotic detection thresholds with increasing numbers of coherently modulated masking bands (e.g., Hall *et al.*, 1990).

Following the above reasoning, the strong frequency effect shown for wideband stimuli may be explained by simply combining within- and across-channel processes. At 4000 Hz, the unmodulated-modulated threshold difference was largest, benefitting from strong within-channel effects due to a wide ERB, as well as across-channel processing spanning approximately 5 ERBs, or nearly 2000 Hz. At 500 Hz, within-channel effects were reduced by the narrow ERB and across-channel processing of the modulated stimulus again spanning about 5 ERBs, or about 400 Hz. Thus, the frequency effects may be determined by within-channel cues.

The derived estimates of CMR, or across-channel temporal processing, reported here are consistent with many previous investigations (e.g., Hall *et al.*, 1984, 1993; Schooneveldt and Moore, 1989). By definition, these estimates are dependent in part upon masked thresholds for the unmodulated maskers, which in turn are dependent upon the noise bandwidth, the signal frequency, the listener's frequency selectivity, and changes in the overall effective masker level with these parameters. An alternative measure of across-channel temporal processing would be to compare thresholds for modulated maskers of different bandwidths, rather than comparing unmodulated-modulated threshold differences. In this case, a suitable metric for across-channel temporal processing, restricted to AM maskers, might be the subtraction of threshold for the 1-ERB condition from threshold for wider bandwidth conditions. This comparison is analogous to CMR estimates derived by subtracting thresholds for an ''on-frequency'' masker band from a multi-band, comodulated masker. A disadvantage of this metric is that there is no control for increases in the overall masker level with increasing bandwidth. Nevertheless, CMR estimated from the AM conditions alone reveals a maximum CMR for a 4-Hz modulation frequency which varied only slightly from 500 to 4000 Hz (decreasing from 10 to 7 dB). The CMR estimates were smaller for the 16-Hz modulation frequency, and again decreased slightly with increasing signal frequency (6, 4, 4, and 3 dB at 500, 1000, 2000, and 4000 Hz, respectively). There was essentially no CMR at the 64-Hz modulation rate. The small signal frequency effect is in the opposite direction to that seen in the estimates of across-channel processing shown in Fig. 3. Clearly the choice of reference condition has a substantial influence on estimates of across-channel processing, both in terms of the magnitude of the effect and the pattern of results across stimulus conditions.

While it is clear that both within-channel and across-channel cues are available in temporal processing tasks such as the present one, it is very difficult to specify the range of bandwidths over which those cues operate. In describing changes in amplitude modulation detection thresholds with frequency region, Strickland (2000) noted that the effective bandwidth resulting from auditory filtering may depend on the perceptual task at hand, and may be substantially wider in certain temporal processing tasks than in tasks such as notched-noise measures of frequency selectivity. This notion of dynamic auditory filtering is consistent with the present results.

C. Within-channel processing based upon statistical analyses

Following the traditional CMR interpretation above, within-channel contributions were assumed to be present in all maskers. The magnitude of the within-channel effect was estimated to be the same as the unmodulated-modulated threshold difference for a masker 1 ERB wide, where the definition of 1 ERB was open to interpretation. Is this a reasonable estimate of the within-channel cues available given wideband maskers? Schooneveldt and Moore (1989) illustrated one approach to evaluate this question. They examined the distribution of envelope amplitudes of maskers with and without the addition of a threshold-level sinusoidal signal. To examine within-channel cues, the stimuli were filtered by a simulated auditory filter centered at 2000 Hz. For unmodulated noise maskers, the signal had little influence on the envelope amplitude distribution. For noise maskers amplitude modulated with a low-pass noise (12.5-, 50-, or 200-Hz corner frequency), the introduction of a signal to the masker led to a small but consistent increase in the mean value of the envelope distributions. This shift was greatest for narrow masker bandwidths and declined with increasing masker bandwidth. The small but consistent changes in the *within-channel* envelope distribution for wide masker bandwidths indicated a potential within-channel cue that was present in wideband maskers but was smaller than for narrow-band maskers.

Because the signal level at threshold varied with masker bandwidth in the Schooneveldt and Moore simulations, one cannot determine whether the magnitude of the within-channel envelope cue remained the same with increasing masker bandwidth beyond 1 ERB. To evaluate this possibility and to extend the envelope simulations to include sinusoidal amplitude modulation, the simulations were repeated for the conditions employed here. For a threshold-level sig-

nal in sinusoidally amplitude modulated noise, the results were essentially the same as those reported by Schooneveldt and Moore for noise modulators. Furthermore, signals presented at a constant level across masker bandwidth conditions indicated that the mean of the envelope distributions changed only slightly with increasing bandwidth. This is consistent with the notion of Schooneveldt and Moore that there is an available within-channel cue given maskers wider than 1 ERB. Since this cue diminished only slightly with increasing masker bandwidth, these simulations indicate that the within-channel cue varies little with increasing bandwidth. These results support the practice of subtracting the unmodulated-modulated difference for 1-ERB maskers from wider maskers to reduce the contribution of within-channel cues to CMR estimates.

A second approach to estimating the contribution of within-channel cues to modified masking period pattern and traditional CMR paradigms (e.g., Hall *et al.*, 1984) was described by Verhey *et al.* (1999). They replicated a number of CMR investigations including the bandwidening experiments of Hall *et al.* (1984) and Schooneveldt and Moore (1989) and reported simulated thresholds for the same conditions using a single-channel version of the modulation filter bank of Dau *et al.* (1997). This model was composed of several stages, including a peripheral filter bank and a modulation filter bank. To focus on potential within-channel cues, the model was restricted to only a single peripheral channel centered on the signal frequency. The model was able to account for most of the masking release obtained behaviorally, regardless of masker bandwidth, suggesting that much of the masking release in these bandwidening experiments may be the result of within-channel rather than across-channel processes. To the extent that this model accurately captures auditory perception, it should also be able to account for a sizable portion of the unmodulated-modulated threshold differences reported here. Given the results of Verhey *et al.*, it seems that masking release in wideband amplitude-modulated noise may be explained in large part by within-channel processes, although the model predictions alone do not eliminate the possibility that across-channel processes play an important role.

D. Signal frequency effects in modified masking period patterns

One goal of the present study was to investigate the influence of stimulus frequency on the modified masking period pattern. Unlike previous investigations, the present experiment showed little influence of signal frequency for relatively narrow maskers ($\leqslant$800 Hz) with constant bandwidths across frequency. On the other hand, modified masking period patterns clearly shifted to higher values with increasing signal frequency when masker bandwidths proportional to the signal frequency or when wideband maskers were used. The clear signal frequency effect with fixed relative bandwidth and wideband maskers is consistent with the results of Bacon *et al.* (1997) and Bacon and Lee (1997), who reported slightly smaller unmodulated-modulated threshold differences than those reported here but a systematic increase in those differences for signal frequencies from 250 to 4000 Hz. Interestingly, for a modulation frequency of 8 Hz and narrow-band maskers with fixed absolute bandwidths, Bacon *et al.* (1997) reported unmodulated-modulated differences that were greater at 2000 than 500 Hz, in contrast to the present results. One explanation for this difference may be differences in the masker spectrum levels used the two studies (40 versus 50 dB SPL). Although the difference in spectrum level was only 10 dB, in a separate experiment Bacon *et al.* illustrated that changes in spectrum level from 0 to 40 dB have strong influences on the resulting modified masking period patterns. Thus, based upon the present results, and previous results in the literature, one may conclude that modified masking period patterns may or may not depend upon signal frequency, depending upon specific choices of masker bandwidths across frequency, masker levels, and perhaps modulation frequency and type of modulation (e.g., sinusoidal, square-wave, or low-pass noise). Likewise, the dependance on signal frequency of CMRs estimated from unmodulated-modulated differences may be determined by specific choices of masker bandwidth, level, type of modulation, modulation frequency, and the choice of reference condition.

V. CONCLUSIONS

The present study was designed to explore the influence of signal frequency and masker bandwidth on modified masking period patterns. The motivation for exploring these parameters arose from discrepancies regarding the influence of these parameters in previous reports using this technique and from the suggestion by several authors that the technique might be useful as a clinical tool to measure auditory temporal resolution. The present data clearly indicate that modified masking period patterns do not vary with signal frequency under certain masker conditions, but do vary with signal frequency under other conditions. Specifically, when maskers have relatively narrow bandwidths ($<$800 Hz) and the bandwidth is held constant (in Hz) across frequency, the masking period patterns reported here varied little for signal frequencies from 1000 to 4000 Hz. For wideband noises and noises with bandwidths that are proportional to the signal frequency, unmodulated-modulated threshold differences increased in magnitude with increasing frequency from 500 to 4000 Hz. When considered in the context of comodulation masking release, one interpretation of frequency effect reported here involves a combination of within-channel cues, that increase with increasing signal frequency, and across-channel cues that reach a maximum or upper limit when the masker bandwidth exceeds roughly three to six times the ERB or roughly 60% of the center frequency. As pointed out by Verhey *et al.* (1999), it may not be necessary to invoke across-channel processes to explain lower thresholds in many of the amplitude-modulated conditions employed here. The extent to which either type of processing contributes to the modified masking period patterns reported here remains uncertain.

Based upon the present results and those reported in other investigations, it can be concluded that modified masking period patterns, based upon threshold differences in unmodulated and amplitude-modulated noise maskers (1) are

inversely related to masker modulation frequency; (2) increase with masker level; (3) increase with masker modulation depth; (4) increase with masker bandwidth; (5) may depend upon both within and across-channel processes; and (6) may vary with signal frequency in different ways, depending upon the masker parameters noted here.

While several authors have advocated the use of modified masking period patterns to estimate temporal resolution in clinical and laboratory settings, the present results suggest that the technique may not provide a valid estimate of temporal resolution. Instead, it may be tapping into the more complex spectro-temporal processing thought to be involved in traditional CMR tasks. Furthermore, the results of the modified masking period pattern technique are dependent upon a number of masker parameters that may include maskers bandwidth, modulation frequency, modulation depth, masker sensation level, and the signal frequency. The interactions among these and perhaps other stimulus variables renders this paradigm a poor choice for clinical measurements of auditory temporal resolution or temporal processing.

ACKNOWLEDGMENTS

The author wishes to thank the two anonymous reviewers and Associate Editor Dr. Joseph W. Hall III for their many helpful comments and suggestions. This work was supported in part by Grant No. RO1 DC04403 from NIH NIDCD.

Bacon, S. B., and Lee, J. (**1997**). ''The modulated-unmodulated difference: Effects of signal frequency and masker modulation depth,'' J. Acoust. Soc. Am. **101**, 3617–3624.

Bacon, S. B., Lee, J., Peterson, D. N., and Raine, D. (**1997**). ''Masking by modulated and unmodulated noise: Effects of bandwidth, modulation rate, signal frequency, and masker level,'' J. Acoust. Soc. Am. **101**, 1600–1610.

Buunen, T. J. F. (**1976**). ''On the perception of phase differences in acoustics,'' doctoral thesis, Technical University, Delft, The Netherlands.

Carlyon, R. P., Buus, S., and Florentine, M. (**1989**). ''Comodulation masking release for three types of modulator as a function of modulation rate,'' Hear. Res. **42**, 37–46.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). ''Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration,'' J. Acoust. Soc. Am. **102**, 2906–2919.

Eddins, D. A. (**1993a**). ''Amplitude modulation detection in narrowband noise as a function of carrier frequency and bandwidth,'' J. Acoust. Soc. Am. **93**, 470–479.

Eddins, D. A. (**1993b**). ''Auditory temporal acuity and across-frequency temporal processing,'' doctoral thesis, University of Florida, Gainesville, FL.

Eddins, D. A. (**1999**). ''Amplitude modulation detection at low and high audio frequencies,'' J. Acoust. Soc. Am. **105**, 829–837.

Eddins, D. A., and Wright, B. A. (**1994**). ''Comodulation masking release for single and multiple rates of envelope fluctuation,'' J. Acoust. Soc. Am. **96**, 3432–3442.

Eddins, D. A., Hall III, J. W., and Grose, J. H. (**1992**). ''The detection of temporal gaps as a function of frequency region and absolute noise bandwidth,'' J. Acoust. Soc. Am. **91**, 1069–1077.

Fletcher, H. (**1940**). ''Auditory patterns,'' Rev. Mod. Phys. **12**, 47–65.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.

Grose, J. H., Hall III, J. W., and Gibbs, C. (**1993**). ''Temporal analysis in children,'' J. Speech Hear. Res. **36**, 351–356.

Haggard, M. P., Hall III, J. W., and Grose, J. H. (**1990**). ''Comodulation masking release as a function of bandwidth and test frequency,'' J. Acoust. Soc. Am. **88**, 113–118.

Hall III, J. W., Cokely, J. A., and Grose, J. H. (**1988**). ''Combined monaural and binaural masking release,'' J. Acoust. Soc. Am. **83**, 1639–1645.

Hall III, J. W., Grose, J. H., and Haggard, M. P. (**1990**). ''Effects of flanking band proximity, number, and modulation pattern on comodulation masking release,'' J. Acoust. Soc. Am. **87**, 269–283.

Hall III, J. W., Grose, J. H., and Moore, B. C. J. (**1993**). ''Influence of frequency selectivity on comodulation masking release in normal-hearing listeners,'' J. Speech Hear. Res. **36**, 410–423.

Hall III, J. W., Haggard, M. P., and Fernandes, M. A. (**1984**). ''Detection in noise by spectro-temporal analysis,'' J. Acoust. Soc. Am. **76**, 50–56.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Rodenberg, M. (**1972**). ''Sensitivity of the auditory system to differences in intensity,'' doctoral thesis, Medical Faculty of Rotterdam.

Rosen, S., and Baker, R. J. (**1994**). ''Characterising auditory filter nonlinearity,'' Hear. Res. **73**, 231–243.

Schooneveldt, G. P., and Moore, B. C. J. (**1989**). ''Comodulation masking release (CMR) as a function of masker bandwidth, modulator bandwidth, and signal duration,'' J. Acoust. Soc. Am. **85**, 273–281.

Scott, D. M., and Humes, L. E. (**1990**). ''Modulation transfer functions: A comparison of the results of three methods,'' J. Speech Hear. Res. **33**, 390–397.

Strickland, E. A. (**2000**). ''The effects of frequency region and level on the temporal modulation transfer function,'' J. Acoust. Soc. Am. **107**, 942–952.

Strickland, E. A., and Viemeister, N. F. (**1997**). ''The effects of frequency region and bandwidth on the temporal modulation transfer function,'' J. Acoust. Soc. Am. **102**, 1799–1810.

Verhey, J. L., Dau, T., and Kollmeier, B. (**1999**). ''Within-channel cues in comodulation masking release (CMR): Experiments and model predictions using a modulation filterbank model,'' J. Acoust. Soc. Am. **106**, 2733–2745.

Weber, D. L. (**1977**). ''Spectral and temporal integration in auditory detection as a function of masker intensity,'' doctoral thesis, Harvard University, Cambridge, MA.

Zwicker, E. (**1980**). ''A device for measuring the temporal resolution of the ear,'' Audiological Acoust. **19**, 2–12.

Functional magnetic resonance imaging measurements of sound-level encoding in the absence of background scanner noise

Deborah A. Hall,[a] Mark P. Haggard, A. Quentin Summerfield, Michael A. Akeroyd, and Alan R. Palmer
MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, United Kingdom

Richard W. Bowtell
Magnetic Resonance Centre, School of Physics and Astronomy, University of Nottingham, Nottingham NG7 2RD, United Kingdom

(Received 31 August 2000; accepted for publication 30 November 2000)

Effects of sound level on auditory cortical activation are seen in neuroimaging data. However, factors such as the cortical response to the intense ambient scanner noise and to the bandwidth of the acoustic stimuli will both confound precise quantification and interpretation of such sound-level effects. The present study used temporally ''sparse'' imaging to reduce effects of scanner noise. To achieve control for stimulus bandwidth, three schemes were compared for sound-level matching across bandwidth: component level, root-mean-square power and loudness. The calculation of the loudness match was based on the model reported by Moore and Glasberg [Acta Acust. **82**, 335–345 (1996)]. Ten normally hearing volunteers were scanned using functional magnetic resonance imaging (fMRI) while listening to a 300-Hz tone presented at six different sound levels between 66 and 91 dB SPL and a harmonic-complex tone ($F_0 = 186$ Hz) presented at 65 and 85 dB SPL. This range of sound levels encompassed all three bases of sound-level matching. Activation in the superior temporal gyrus, induced by each of the eight tone conditions relative to a quiet baseline condition, was quantified as to extent and magnitude. Sound level had a small, but significant, effect on the extent of activation for the pure tone, but not for the harmonic-complex tone, while it had a significant effect on the response magnitude for both types of stimulus. Response magnitude increased linearly as a function of sound level for the full range of levels for the pure tone. The harmonic-complex tone produced greater activation than the pure tone, irrespective of the matching scheme for sound level, indicating that bandwidth had a greater effect on the pattern of auditory activation than sound level. Nevertheless, when the data were collapsed across stimulus class, extent and magnitude were significantly correlated with the loudness scale (measured in phons), but not with the intensity scale (measured in SPL). We therefore recommend the loudness formula as the most appropriate basis of matching sound level to control for loudness effects when cortical responses to other stimulus attributes, such as stimulus class, are the principal concern. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1345697]

PACS numbers: 43.66.Cb, 43.66.Ba, 43.64.Ri [LHC]

I. INTRODUCTION

Functional magnetic resonance imaging (fMRI) is a valuable tool for the study of human auditory cortical processing as it can noninvasively measure the topographical organization of cortical responses to an acoustic signal. fMRI indirectly measures the neuronal population responses through the metabolic consequences of neuronal activity on changes in the ratio of oxygenated to deoxygenated hemoglobin in the blood (the blood-oxygen-level-dependent, BOLD, response). Here, we sought to quantify the effects of sound level on both the extent and magnitude of activation in the auditory cortex and to determine the form of any relationship between them. By using a sparse imaging protocol with silent intervals between image acquisition, responses to sound level were measured in the absence of background scanner noise.

A. Neuroimaging studies of sound intensity

Both fMRI and positron emission tomography (PET) studies have measured systematic changes in auditory activation with sound level, particularly on the supratemporal plane and superior temporal gyrus (e.g., Jäncke *et al.*, 1998; Lockwood *et al.*, 1999; Millen *et al.*, 1995; Mohr *et al.*, 1999; Strainer *et al.*, 1997). Results have generally indicated a growth in auditory activation with sound level. Understanding brain responses to sound level contributes not only to knowledge of sensory coding, but also to methodological awareness of stimulus control. However, the organizing principles of the cortical response to sound level are not yet clear. The lack of a clear consensus on sound-level effects may be a consequence of differences in the range of sound levels presented, particularly if the response saturates at higher levels. Studies have also used different criteria for the specification of sound-level; e.g., sound-pressure level (Jäncke *et al.*, 1998), sensation level (Lockwood *et al.*, 1999), or dB relative to a measure of speech intelligibility for each individual subject (Mohr *et al.*, 1999). The ability to

[a]Electronic mail: debbie@ihr.mrc.ac.uk

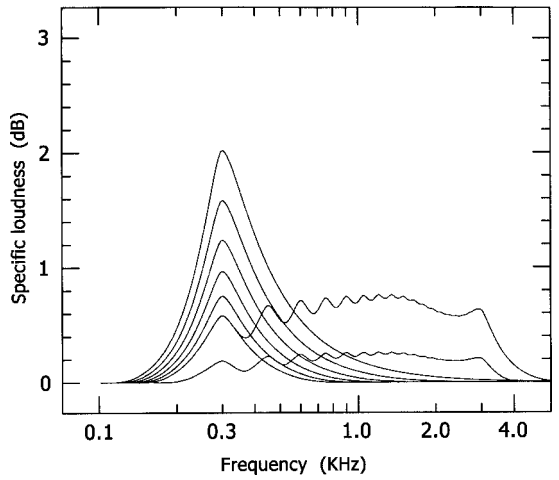


FIG. 1. Specific loudness patterns induced by the six 300-Hz pure tones and the two harmonic complex tones according to the model of Moore *et al.* (1997). The specific loudness pattern is calculated from the physical spectrum after correction for the effects of transmission through the outer and middle ear. The two complex tones generate broader and flatter loudness patterns than those for the pure-tone stimuli. The overall loudness of a sound is obtained by summing the specific loudness [i.e., the loudness per equivalent rectangular bandwidth (ERB)] across the frequency axis, multiplied by 2 for diotic presentation.

exercise control or draw conclusions has also been limited by the coarse quantization or narrow range of levels. Two or three levels are insufficient to determine accurately the shape of the response curve; e.g., to establish whether or not it is linear in dB. The response to sound level has generally been measured across only a few tens of dB relative to the overall system dynamic range of about 120 dB (Viemeister and Bacon, 1988).

B. Effect of the intense background noise in fMRI

fMRI measures of the auditory response to sound intensity, in particular, risk contamination by the intense background scanner noise. The principal source of noise is the mechanical deformation of the gradient coils caused by the electrical currents flowing in them within the high static magnetic field. For fMRI, the level almost invariably exceeds 100 dB SPL and here was about 127 dB SPL (Foster *et al.*, 2000). For the present study, the acoustic spectrum of the scanner noise was dominated by a harmonic-complex tone ($F_0 = 1.9$ kHz) rising above a background of broadband noise (see Hall *et al.*, 2000, Fig. 1).

The scanner noise is a problem for three main reasons. First, it produces its own auditory activation which can overlap with the stimulus-driven auditory response, particularly with conventional imaging protocols where images are acquired at a continuous rate (e.g., Bandettini *et al.*, 1998; Hall *et al.*, 1999, 2000; Shah *et al.*, 1999, 2000; Talavage *et al.*, 1999). This increased baseline level of activation can reduce the size of the experimental effect, and hence reduce the statistical power for detecting stimulus-induced auditory activation. In addition, the suprathreshold magnitude information in the measured response in part reflects an interaction between the responses to the stimulus and to the background noise. Second, the scanner noise can mask the auditory stimulus whenever the stimulus and image acquisition occur close in time. Masking increases the cognitive load for low stimulus levels relative to higher levels, changing the nature of the task by making it harder to detect target stimuli, particularly at low sound levels. To overcome masking, presentation levels have tended towards the upper end of the comfortable range of human hearing, and here the cortical response is most likely to saturate. Third, even with hearing protection, the intense background noise can induce a stapedial muscle reflex that alters the sound level transmitted to the inner ear. The level of the sound that triggers this reflex varies across individuals. Although differences in the acoustic reflex threshold do not necessarily reflect differences in loudness perception in normally hearing listeners (Olsen *et al.*, 1999; Olsen, 1999), imaging studies that use EPI and/or present stimuli towards the upper end of the dynamic range need to consider its effect on sound-level perception.

Temporally sparse imaging can reduce the problems caused by the scanner noise as it intersperses a relatively long period of silence (about 10 s) between bursts of scanner noise and separates the noise in time from stimulus delivery. Sparse imaging minimizes perceptual masking and the residual auditory activation effects of the scanner sound. Known effects of intense noise on middle-ear reflexes which may contribute to the masking effects are also thereby minimized. The ability of sparse imaging to detect activation is *not* measurably compromised by the reduction in the number of data averages because sensitivity is maximized by contrasting the peak response with the poststimulus negative phase of the response (which maximizes the percentage change in the BOLD signal between activation and baseline conditions) and, by achieving greater MR signal recovery between image acquisitions (which enhances the BOLD signal-to-noise ratio) (Hall *et al.*, 1998).

C. Requirement for stimulus control

Changes in response magnitude have been used successfully to investigate neural activity patterns associated with different acoustical stimuli including changes in frequency (Talavage *et al.*, 1997), bandwidth (Rauschecker *et al.*, 1997; Wessinger *et al.*, 1997), and pitch strength (Griffiths *et al.*, 1998). When explicitly defining stimulus-evoked activation, it is a potential concern that the observed pattern of activation may partly reflect a response to a separate acoustical property of the stimulus that may not have been tightly controlled across conditions. Evidence to date suggests that both bandwidth (Wessinger *et al.*, 1997) and sound level (Jäncke *et al.*, 1998) have similar effects on the spread of activation on the supratemporal plane and superior temporal gyrus. Thus, subject to what has been controlled, the two effects may not be easily separable.

There may be four possible interpretations for the bandwidth effect as observed. The first explanation, preferred by Wessinger *et al.* (1997), is that broadband stimuli are intrinsically more effective than pure tones in evoking a neuronal response. This may be especially true in regions of nonprimary auditory cortex where neurons respond more strongly to broadband stimuli than to pure tones (Redies *et al.*, 1989). Second, the effect of bandwidth could be straightforwardly attributed to the basic spread of auditory excitation across

frequency channels, particularly in the multiple tonotopically organized auditory cortical fields (Merzenich and Brugge, 1973; Thomas *et al.*, 1993). This second explanation is difficult to distinguish operationally from the first. The third explanation draws attention to sound level, an additional acoustical feature that can differ between a pure tone and a broadband stimulus and may therefore be a potential confound. Where details are reported, imaging studies that manipulate bandwidth have sought to control for sound level by equating sound-pressure level (e.g., Wessinger *et al.*, 1997). However, if the bandwidth of a broadband signal is increased while its SPL is held constant, then loudness nevertheless increases. When an increase in bandwidth adds frequency components to separate frequency channels, the loudness model of Moore, Glasberg, and Baer (1997) predicts that the excitation pattern widens, increasing the total area and hence the loudness. This result holds true for signals whose bandwidth exceeds the critical bandwidth (e.g., 160 Hz at a center frequency of 1 kHz) and for signals above about 30 dB SPL (Zwicker *et al.*, 1957; Moore and Glasberg, 1996). As a consequence, the bandwidth effect shown in neuroimaging data may be artifactually overestimated. It is possible that bases for matching other than SPL, such as via a loudness model (e.g., Moore *et al.*, 1997), would have greater physiological validity at the cortical level. The fourth explanation invokes the consequences of spatial smoothing on voxels that have differing signal-to-noise ratios, where magnitude partly determines extent. If broadband stimuli produce a stronger BOLD response than pure tones, then this is likely to be associated with a greater extent of activation when the statistical map is thresholded at a common probability level.

Further studies are required therefore which control these variables in order to eliminate alternative explanations for the difference in the patterns of activation for pure tones and broadband stimuli. Investigation of the third explanation (the effect of a loudness mismatch) requires control for sound level when measuring other features of the auditory stimulus that drive a particular pattern of cortical response.

D. The present study

The present study quantified effects of sound level on both the extent and magnitude of auditory activation using 300-Hz tones at six sound levels ranging from 66 to 91 dB SPL, in 5-dB increments. The relatively small step size in sound level allowed the determination of the shape of the BOLD response across this intensity range. By using temporally sparse imaging to overcome masking by the background noise, we were also able to present the tones across a range of naturally occurring sound intensities, the lower end of the range being limited only by the broadband background ambient noise (63 to 68 dB SPL). Ear defenders attenuated these background levels by about 10 dB at 300 Hz and by about 30 dB between 500 Hz and 10 kHz. There is currently no neuroimaging evidence to favor one sound-level matching scheme over another, and so the present study also addressed whether matching based on equal loudness is appropriate for studies of cortical coding. Across normally hearing listeners, there are individual differences in judgements of loudness measured experimentally (Stevens and Guirao, 1962; Buus *et al.*, 1998). Thus, we used a computational method to define loudness (measured in units of phons) that makes no assumptions about individual decision criteria (Moore and Glasberg, 1996; Moore *et al.*, 1997). We investigated the conditions under which activity due to a complex tone approached the activity for a pure tone presented at a higher intensity, but matched for loudness. The design also allowed us to investigate whether bandwidth influences magnitude, as well as extent, of activation, an issue not previously addressed.

II. METHODS

A. Subjects

Ten right-handed subjects, aged 20–46, participated in the study. Subjects had no history of neurological impairment. All subjects gave informed written consent and the study was approved by the Nottingham University Medical School ethical committee. Prior to the imaging session, the hearing sensitivity of subjects was measured using pure-tone audiometry. The hearing thresholds of all subjects fell within the normal range ($<$20 dB HL) at octave frequencies between 500 and 8000 Hz inclusive.

B. Stimuli and calibration

A harmonic-complex signal was synthesized with 16-bit amplitude quantization at a sampling rate of 48 kHz. This stimulus had a fundamental frequency of 150 Hz and was composed of harmonics extending from 300 Hz to 3 kHz inclusive. The harmonics were synthesized with equal amplitude but, as a consequence of the frequency response of the headphones, ranged by 12.6 dB when measured at the ear. The phases of the harmonics were chosen randomly and so stimuli were not specifically selected for their lowest peak factors. This stimulus was then resynthesized to exploit the full 16-bit digital dynamic range. The harmonic complex was presented at overall levels of 65 and 85 dB SPL (see below for details of the calibration procedure). Next, six 300-Hz pure tones were synthesized. The first had the same amplitude as the 300-Hz component of the 85-dB SPL complex tone, measured at the ear; its level was 66-dB SPL. The remaining five 300-Hz tones had amplitudes that were greater than the 66-dB SPL tone by 5, 10, 15, 20, and 25 dB, i.e., they were presented at levels of 71, 76, 81, 86, and 91 dB SPL, respectively (see Table I). All stimuli were 800 ms in duration, including 10-ms inverted-cosine ramps at onset and offset.

Stimuli were presented diotically through a specially engineered, MR-compatible sound system that delivers acoustic stimuli using electrostatic drivers built into industrial ear defenders (Palmer *et al.*, 1998). The presentation levels reported above were calibrated by mounting the system on KEMAR (Burkhard and Sachs, 1975), equipped with a Brüel & Kjær microphone (type 4134) connected to a Brüel & Kjær measuring amplifier (type 2636), and by measuring the overall power of the stimulus.

The choice of sound levels meant that the 85-dB SPL complex tone was matched in two ways to two different members of the set of pure tones. First, by design, it had the

TABLE I. Presentation levels for the pure and harmonic complex tones. Intensity (dB SPL) was calibrated using KEMAR (Burkhard and Sachs, 1975). The loudness for the same stimuli was calculated using the computational model of Moore *et al.* (1996).

Intensity (dB SPL)	Loudness (phons)
Pure tone	
66.0	56.8
71.0	61.9
76.0	67.0
81.0	72.0
86.0	77.3
91.0	82.4
Harmonic complex tone	
65.0	69.4
85.0	86.2

same component level as the 66-dB SPL pure tone. Second, it had approximately the same rms level (average rms power =79 dB, 19 harmonic components) as the 76-dB SPL pure tone (average rms power=76 dB). The loudness of all stimuli was also calculated using the computational model of Moore and colleagues (Moore and Glasberg, 1996; Moore *et al.*, 1997). In this model, loudness is a function of the auditory excitation induced by a sound, integrated across frequency (see Fig. 1). The background noise level in the scanner room was sufficiently attenuated by the ear defenders to cause little partial masking and thus had an insignificant effect on the loudness estimates of the stimuli. The loudness estimates of the stimuli are reported in Table I. The model predicted that none of the pure tones was sufficiently intense to match the loudness of the 85-dB complex tone. Instead, the model predicted that the loudness of the 76- and 81-dB pure tones straddled that of the less intense 65-dB complex tone. This model prediction provides a third basis for matching the pure and complex tones for sound level and one that can be evaluated by the stimulus range used in the present study.[1]

C. Stimulus presentation

The six levels of the pure tone and the two levels of the complex tone defined the eight stimulus conditions of the experiment. Within each condition, ten 800-ms tone bursts were presented in succession separated by 200 ms of silence. The task required target discrimination based on a change in the sound level of the tones. Targets had an intensity that was 10 dB lower than the nontargets. For example, targets in the 66-dB SPL pure-tone condition were presented at 56 dB SPL. To ensure attention to the entire 10-s tone sequence, targets occurred randomly in a ratio of 1:4 and hence could occur more than once in each 10-s sequence. Targets at 56 dB were at the lowest intensity (in 5-dB steps) that was clearly audible in the MR environment and thus defined the lower end of the intensity range that could be used in the present study. So the subjects could determine the reference sound level of the nontargets in each epoch, the first two tone bursts in a sequence were always nontargets. Subjects were instructed to press a button with the index finger of the right hand to the occurrence of each target stimulus. A PC logged the times of occurrence of targets and button presses for off-line analysis of detection performance.

One cycle in the experiment consisted of eight tone conditions followed by a baseline condition and there were 24 such cycles. Tone conditions were ordered so that each condition occurred equally at each of the eight possible positions within a cycle (according to a Latin square). Each of the 9 conditions occurred 24 times, giving a total of 216 stimulus epochs and a total experimental time of 39 min.

D. fMRI scanning

The study was performed on a 3-Tesla MR scanner with head gradient coils and a birdcage radio-frequency coil (Bowtell *et al.*, 1994; Bowtell and Peters, 1999). An MBEST echo-planar sequence was used to acquire sets of 16 contiguous coronal images covering the auditory cortex. Each image measured 128×128 voxels and the voxel resolution was 3×3×8 mm. Through-plane voxel resolution was sacrificed in order to achieve better in-plane resolution for the same voxel volume, and hence an adequate signal-to-noise ratio. A set of images took 1072 ms to acquire and these were acquired every 10.75 s at the transitions between stimulus conditions (see Fig. 2). The intense burst of scanner noise produced during image acquisition occurred mostly during the 750-ms silent intervals between stimulus epochs and consequently did not mask the stimuli presented.

E. Image analysis

Image analysis was performed using statistical parametric mapping software (Friston *et al.*, 1995a; http:www.fil.ion.ucl.ac.uk/spm) following a standard protocol for each subject. Images were corrected for three-dimensional head movement using a computational algorithm that minimized the sum of squared differences between the mean image and each image in the time series (Friston *et al.*, 1995b, 1996). For each subject, the amount of motion

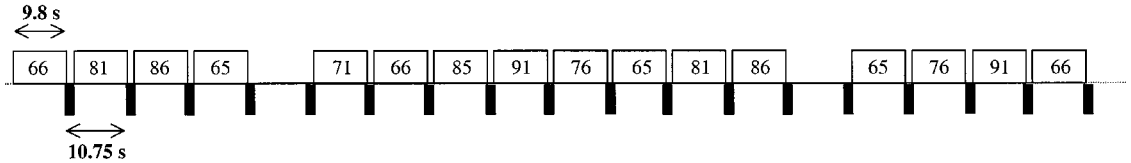


FIG. 2. Schematic diagram representing the sparse-temporal-sampling imaging protocol used in the study. In the experimental sequence shown, stimulus conditions are represented by the blocks labeled by intensity (in dB SPL), with a period of silence occurring at every eighth epoch. The solid black lines denote the occurrence of the 1072-ms imaging sequence which occurs mostly in the 950-ms interval between epochs. Image acquisition is repeated every 10.75 s at the offset of each epoch.

correction required was generally less than 1 mm in each plane and less than 1° rotation about each axis. Realigned images were spatially smoothed using an isotropic Gaussian kernel of 5 mm full-width-at-half-maximum to enhance the signal-to-noise ratio. For the optimal detection of activated regions, a general smoothness heuristic of twice the voxel size was applied in-plane, while to reduce artifactual bleeding of activation across adjacent images, lighter smoothing was applied through-plane. Image data were not temporally smoothed, since the interscan interval exceeded the width of the smoothing kernel. Low-frequency artifacts, corresponding to aliased respiratory and cardiac effects and other cyclical variations in signal intensity, were removed by high-pass filtering the time series at 0.3 cycles/min.

Image analysis was conducted for each subject using the general linear model. For each time course, 8 *t* contrasts were performed between each tone condition and the baseline condition. These *t* statistics were computed for all voxels in the brain. Maps of *t* values were transformed to the unit normal distribution to give maps of *Z* values (referred to as SPM{*Z*}). The SPM{*Z*} for each stimulus condition relative to the baseline was thresholded at a probability level of $P<0.001$ (Z value$=3.09$).

For the purposes of descriptive anatomical localization, the maps of functional activation were overlaid onto brain images for each subject. The location of the auditory cortex can be identified from these images by the position of the Sylvian sulcus, a deep fissure which follows the supratemporal plane. Viewed coronally, the primary auditory cortex is situated medially on the lower bank of the Sylvian fissure and occupies the transverse temporal gyrus of Heschl (e.g., Penhune *et al.*, 1996). Secondary auditory fields include the lateral convexity of the superior temporal gyrus (e.g., Rivier and Clarke, 1997; Westbury *et al.*, 1999).

F. Extent and magnitude analysis

To investigate systematic changes in the pattern of auditory activation as a function of stimulus intensity, the extent of the activated region and magnitude of the response were calculated separately for each tone contrast and for left and right hemispheres. Extent of activation was defined as the number of activated contiguous voxels in the auditory cortex whose probability of activation exceeded $P<0.001$. Response magnitude was represented by the percentage change in the MR signal for each stimulus condition relative to the baseline. Image analysis using statistical parametric mapping requires that the data are spatially smoothed. As a consequence, the extent and magnitude of the BOLD response are partly codependent such that regions with a particularly high response magnitude will also tend to have the greatest extent. To partly separate the codependence of response magnitude and extent, magnitude was computed for a region of fixed size (50 voxels). The 50 voxels were specified by functional, rather than by anatomical, criterion since BOLD contrast images do not specify precise anatomical subdivisions of the auditory cortex. The region was defined using an overall *F* test, which identified regions where there was significant tone activation relative to the baseline. The probability threshold of the *F* test was adjusted for each subject to identify a cluster of 50 voxels around the superior temporal gyrus with the greatest *F* values. Using this criterion, some of those selected voxels did not reach the $P<0.001$ probability threshold required for the extent measurement. However, this ''region of interest'' approach ensured that differences in the extent of activation did not inadvertently affect the measure of response magnitude. For example, a positive signal change was not contingent upon an extent that was greater than zero.

III. RESULTS

A. Psychophysical data

Subjects were able to detect the intensity decreases accurately (mean percent correct$=86.9\%$, s.d. across listeners $=19.9\%$). An analysis of variance (ANOVA) was calculated for the detection of targets in the pure and complex tone conditions. There was no difference in accuracy of detection across the eight stimulus conditions [$F(7,77)=0.83$, $p=0.57$]. Ability to discriminate the targets was also measured using the d' detection index, taking into account the number of false-positive responses, i.e., button presses to the ''nontarget'' sounds. The resulting d' values ranged from 3.9 (for the 66-dB single tone) to 4.3 (for the 85-dB complex tone), with a mean d' of 4.1. The d' values were consistently high across subjects (ranging from 3.1 for subject 10 to 5.42 for subject 4), indicating that subjects were actively detecting intensity changes for all tones.

B. General pattern of functional activation

All ten subjects showed auditory activation in at least six of the listening conditions relative to the silent baseline. Auditory activation was generally located in both medial and lateral regions of the superior temporal gyrus. This region includes the primary and secondary areas of the auditory cortex as anatomically defined in Sec. II E. (Examples of the activation pattern for two subjects are shown in Fig. 6.) Changes in the activation pattern induced either by increasing sound level or by changing the stimulus bandwidth occurred within a co-occurring region of the auditory cortex, suggesting that the activation effects for the manipulations of stimulus level and bandwidth are not easily separable.

Although the experimental design and analysis were not maximally sensitive to the detection of the motor response, in five subjects, task-specific activation most likely to be associated with the finger press was also observed in several cortical regions: the pre- and postcentral gyri along the fronto–parietal border containing the primary and secondary motor and somatosensory cortices, and the medial part of the superior frontal gyrus which contains the supplementary motor area and supplementary eye fields (Dejardin *et al.*, 1998; Pickard and Strick, 1996). Three subjects without movement-elicited activation also had the lowest performance scores, perhaps reflecting the greater number of response misses (i.e., fewer button presses).

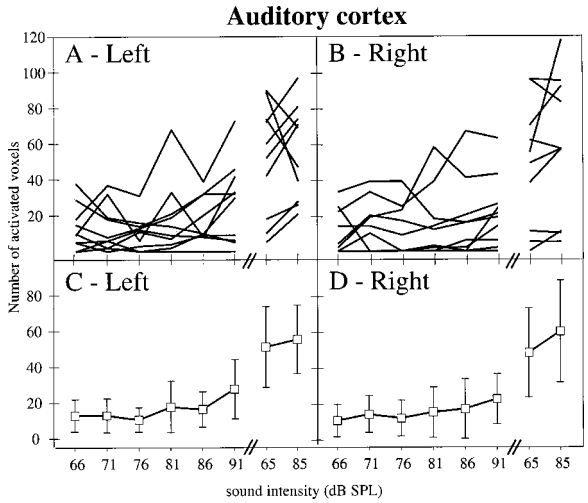


FIG. 3. Extent of activation across stimulus intensity for pure and complex tones. The responses to the two classes of tones can be represented within the same range of sound intensity, but for clarity are represented side-by-side within each panel, with pure tones on the left and complex tones on the right. Panel (A) shows the extent of activation for all ten subjects in the left hemisphere and panel (B) shows the results of the same analysis in the right hemisphere. Panels (C) and (D) show the mean data. Error bars indicate the 95%-confidence limits of the means.

C. Effect of sound level on the extent of auditory activation

In general, increasing sound intensity generated a spread of activation medially and laterally along the supratemporal plane. The data for all ten subjects for the extent of the activated region in each tone contrast are shown in Fig. 3, with the mean and (95%) confidence intervals for each condition (where extent is defined in Sec. II F). Four subjects showed bilateral activation in all conditions. The remaining six subjects showed bilateral activation for the pure tones presented at the upper end of the range (91 dB SPL) and for the two harmonic-complex conditions and unilateral activation for the pure tones presented at the lower end of the range (66 to 86 SPL) (in five subjects, this was in the left hemisphere).

Data for the pure tones were subjected to an analysis of variance, with intensity and hemisphere as within-subject factors. There was a significant overall effect of intensity on extent of activation [$F(5,40)=2.46$, $P=0.05$], with the mean activation/intensity function showing a growth in the extent of activation with increasing level. However, the linear component of the response did not reach significance [$F(1,8)=3.83$, $P=0.09$]. There was no main effect of hemisphere [$F(1,8)=0.31$, $P=0.59$]. For the complex tones, a paired t test showed that extent of activation over both hemispheres did not change with sound intensity [$t(19)=1.35$, $P=0.20$], nor with hemisphere [$t(19)=0.02$, $P=0.98$]. Indeed, while some subjects showed an increase in extent for the complex tone as a function of sound level, others showed a decrease. Thus, the data provide (at most) weak evidence for a general increase in the extent of activation for pure tones presented over 66 to 91 dB SPL, but no greater extent for complex tones presented at 85- than at 65-dB SPL.

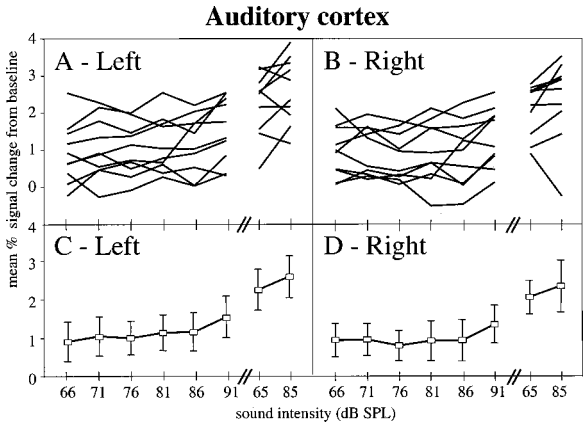


FIG. 4. Mean percentage response change from baseline for all stimulus conditions within each defined region of auditory activation. Within each panel, pure-tone responses are shown on the left-hand side and complex-tone responses on the right-hand side. Results for each hemisphere are plotted for all ten subjects in panels (A) and (B). Panels (C) and (D) show the mean response change. Error bars plot the 95%-confidence intervals of the means.

D. Effect of sound level on the response magnitude

Within the activated auditory area defined in Sec. II F, the mean percentage change in the BOLD response ranged from 0.9% in the 66-dB pure-tone condition to 2.5% in the 85-dB harmonic-complex condition. The magnitudes of the auditory response are in the range commonly observed (Hall *et al.*, 1999; Hall *et al.*, 2000). Overall, the mean percentage change in the response to pure tones increased with intensity. The data are plotted in Fig. 4. Subjects generally showed a positive response to all of the stimuli relative to the quiet baseline condition, although a relative decrease in the response was seen for two subjects in some tone conditions.

Within-subjects analysis of variance for the pure tones showed a significant effect of intensity [$F(5,45)=4.64$, $P<0.01$]. Moreover, within-subject contrasts indicated a significant linear change [$F(1,9)=7.48$, $P=0.02$], showing that the overall growth in magnitude as a function of intensity in dB was linear. For the complex tones, the response magnitude was greater at the higher intensity [paired $t(19)=2.30$, $P=0.03$] (see Fig. 4). There was no hemispheric effect on the magnitude of activation for either pure [$F(1,9)=0.91$, $P=0.37$] or complex tones [$t(19)=1.76$, $P=0.10$].

E. Correlating measures of extent and response magnitude

The correlation between extent of activation and response magnitude was examined on the data collapsed across hemisphere (Fig. 5). The expected positive correlation was high ($r=0.82$, $P<0.001$, $N=80$). As the magnitude of the response increased, so did extent.

Although the analyses in Secs. III C and III D determined the effect of sound level on the extent and magnitude of the response, they did so separately for each stimulus class. If the pattern of auditory activation predominantly reflects a general response to sound level, then the extent and/or magnitude of activation should rise with sound level irrespective of stimulus class. It is therefore powerful to ex-

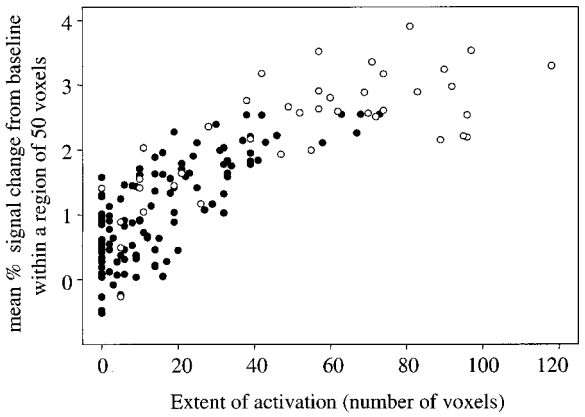


FIG. 5. The relationship between extent and magnitude of the auditory response for the same data as are plotted in Figs. 3 and 4. The extent of activation is defined by the number of voxels whose probability of activation exceeds $P<0.001$ in the SPM{Z}. The percent signal change from baseline is the averaged signal change across 50 voxels in the auditory cortex, whose probability of activation may not exceed $P<0.001$. Thus, a positive signal change is not contingent upon an extent that is greater than zero as it is measured using a different procedure. Data are plotted for the eight stimulus conditions, for all subjects, and are shown for the pure tone as filled circles and for the complex tone as open circles.

amine, pooled over stimulus class, the relationships between the measures of activation (extent and magnitude) and the two sound-level scales (intensity and the measure of loudness from the model of Moore *et al.*, 1997) as a means of distinguishing between the two bases for sound-level matching. Neither extent nor magnitude of activation correlated with intensity as measured in SPL [$r=0.04$, $P=0.59$ and $r=0.06$, $P=0.48$ ($N=80$), respectively]. However, both measures of activation correlated significantly with loudness as measured in phons ($r=0.36$, $P<0.001$ and $r=0.35$, $P<0.001$). The greater correlation arose from the placement of the two complex tones at higher points of both the activation and the sound-level ranges when stimuli were measured in phons rather than in SPL. This result does not mean that the cortical response reflects loudness coding directly, since the correlations explain barely 10% of the variance in the data. However, it does establish the need to control for loudness when quantifying auditory activation to other stimulus attributes, such as bandwidth.

F. Evaluating the three schemes for sound-level matching across bandwidth

If the observed auditory activation reflected a response to sound level alone, then it should be possible to identify a matching scheme which equates the pattern of activation across both classes of stimulus. Three planned *t*-test comparisons were conducted for those *a priori* matched conditions. The 85-dB SPL complex tone produced a significantly greater extent of activation and response magnitude ($P<0.01$) than the pure tone matched for component level (the 66-dB SPL pure tone) and root-mean-square level (the 76-dB SPL pure tone). Matching based on the calculated loudness identified the 76-dB and 81-dB pure tones (67.0 and 72.2 phons respectively) as straddling the 65-dB complex tone (69.4 phons) (see Table I). *t*-test comparisons showed that the complex tone produced significantly more activation ($P<0.01$) than either of these two pure tones in terms of both extent and magnitude of auditory activation. Thus, the complex tone produced a greater response than any of the pure tones matched on the basis of the three schemes. The principal acoustical property that distinguished the pure and the complex tones was that of stimulus bandwidth and we therefore suggest that stimulus bandwidth probably contributed more strongly to the pattern of auditory cortical activation than did the dB SPL or loudness of that stimulus.

The clearest demonstration of the bandwidth effect is seen by evaluating the degree to which activation is sensitive to changes in bandwidth and insensitive to changes in loudness. Activations (i.e., the SPM{Z} maps) for pairs of tone conditions were superimposed onto brain images for each individual. Two pairs of conditions were contrasted; first, the pure tone at 56.8 and 72 phons and second, the pure tone at 56.8 phons and the complex tone at 69.4 phons. Visual inspection of these contrast maps permitted the identification of voxels that were activated in one condition, but not in another, and vice versa. An increase in loudness of 15.2 phons activated small regions of auditory cortex. However, an increase in bandwidth, with a similar increase in loudness (12.6 phons), produced a two- to threefold greater increase in auditory activation (Fig. 6). The greater spread of activation with bandwidth than with loudness, illustrated for subjects 2 and 3, is representative of the rest of the group.

IV. DISCUSSION

This experiment demonstrates that acoustic intensity influenced the pattern of auditory cortical activation when an imaging protocol was used that restricted the contaminating effects of scanner noise. Extent and magnitude of activation were positively correlated, as expected, but not perfectly. This result indicates that, for the mapping of cortical responses to stimulus class and sound level, both measures should be extracted. Sound level influenced the extent of activation for pure tones presented at levels of 66 to 91 dB SPL, but not for the complex tones presented at levels of 65 and 85 dB SPL. On the other hand, an increase in sound level was associated with a significant growth in response magnitude for both classes of stimulus. The six increments in dB SPL for the pure tones permitted the magnitude/intensity function to be determined and, for these stimuli, magnitude was found to increase linearly when responses were averaged over all subjects. Using the three sound-level-matching schemes, statistical comparisons between the auditory activation for pairs of stimuli showed that the response was always greater for the complex tone than for the pure tone. Thus, the activated region in the superior temporal gyrus responded differentially both to changes in sound level and in bandwidth, but the effect of bandwidth was the greater.

A. Neurophysiological bases for the auditory activation changes as a function of sound level

Independent physiological knowledge of the neural coding of intensity/loudness enables speculation about the relation between this and the observed changes in the BOLD signal because it drives the neurovascular effects. Sound

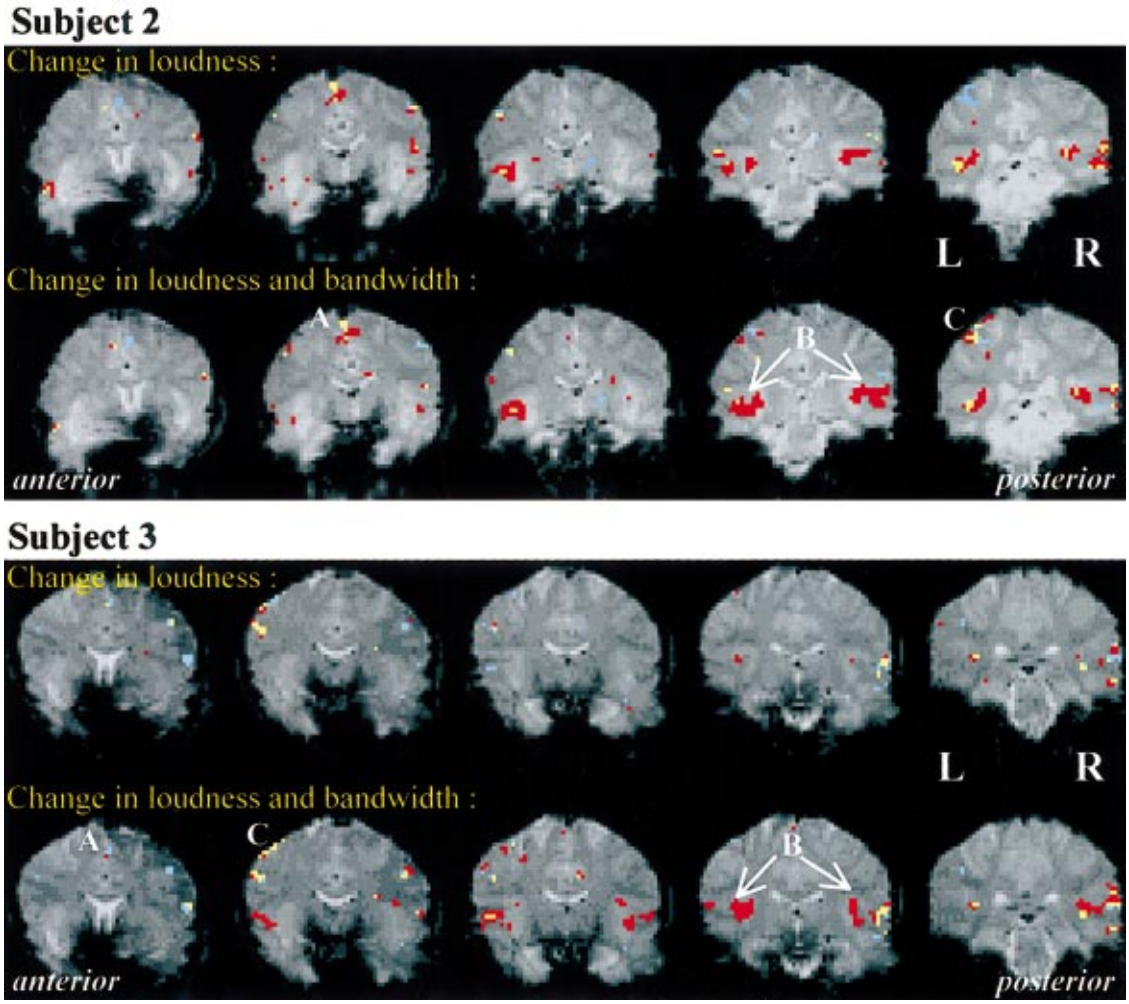


FIG. 6. Activation maps overlaid onto five functional images showing greater response to bandwidth than to loudness for two subjects. The effect of a change in loudness is represented by superimposing activations by the pure tone presented at 56.8 and 72 phons onto the same image. In the top row for each subject, red voxels denote significant activation at 72 phons and blue voxels denote significant activation at 56.8 phons. Images in the row below demonstrate the additional change in bandwidth. There is much more activation by the complex tone (shown in red) than by the pure tone (shown in blue) when the two tones have a similar loudness separation as above (56.8 and 69.4 phons). In all images, voxels in yellow are conjointly activated by both tones. Auditory activation formed a mediolateral band in the superior temporal gyrus on the lower bank of the Sylvian fissure (regions B). Activation possibly reflecting the motor component of the task can be seen in region A, the medial part of the superior frontal gyrus containing the supplementary motor area, and region C, the left pre/post-central gyri.

level may be represented by the activation of units which are distributed within volumes containing units subserving other functions (e.g., Heil *et al.*, 1994; Taniguchi and Nasu, 1993), including the sharpness of frequency tuning to pure tones (Recanzone *et al.*, 1999). The imaging data are at least consistent with these neurophysiological data, as effects of both sound level and bandwidth were found in overlapping regions of auditory cortex. Within the mammalian primary auditory cortex, an orderly spatial organization of a number of parameters related to the encoding of sound level has been demonstrated, including minimum threshold, dynamic range, best SPL, and nonmonotonicity of intensity functions (e.g., Heil *et al.*, 1994), although the present spatial and temporal resolution of the fMRI technique is insufficient to detect these dimensions of cortical representation.

At peripheral levels of the ascending auditory pathway, sound level is represented by the firing rates of neurons at the center of the excitation pattern (e.g., Liberman, 1978); by the spread of the excitation pattern (e.g., Chatterjee and Zwislocki, 1998) and by the patterns of temporal synchrony of the neuronal firing (e.g,. Brosch and Schreiner, 1999; Carlyon and Moore, 1984). However, at the level of the auditory cortex, the neuronal population response to sound level becomes quite complicated: temporal coding has largely disappeared and rate coding is a mixture of both monotonic and nonmonotonic neuronal responses to increasing sound level (e.g., Heil *et al.*, 1994; Phillips and Orman, 1984). The combined contribution from neurons that have different rate-intensity functions results in a changing topographical distribution of activity as the sound level of a pure tone is increased. Interestingly, those same authors demonstrated that the cumulative activity across a population increases in a manner similar to the growth of loudness. This is despite (or perhaps because of) the widespread inhibitory effects that are also brought into play at higher sound levels. If the cortical responses entirely reflect processing prior to the auditory cortex, the BOLD response should reflect the increasing cumulative spike count.

Interplay between excitatory and inhibitory inputs can result in nonmonotonic rate-intensity functions (Greenwood and Maruyama, 1965; Brugge and Merzenich, 1973). Thus, the physiological response to increasing sound level also involves inhibition of the activity of narrow-band, low intensity-sensitive neurones. Such inhibition is in evidence from the dorsal cochlear nucleus all the way up to the cortex, and there are certainly local inhibitory contributions to nonmonotonicity at least as high as the inferior colliculus (Yang *et al.*, 1992). At the cortical level, there are profuse local inhibitory influences (Cox *et al.*, 1992; Manunta and Edeline, 1998; Wang *et al.*, 2000), although a direct local contribution to the observed nonmonotonicity of rate-level functions has yet to be demonstrated. Neural inhibition is likely to be metabolically costly and may not be distinguishable from the metabolic cost of excitation. Consequently, if nonmonotonicity in the cortex receives a local inhibitory contribution, spiking and nonspiking regions would both be costly in terms of their oxygen requirement. Thus, measures based on local hemodynamic changes, such as the BOLD response in fMRI and the rCBF measure in PET, are unable to distinguish neural excitation from local inhibition. Neuroimaging data therefore require careful interpretation where inhibitory, as well as excitatory, responses are likely to be involved. However, despite this fact, an overall increase in oxygen requirement would be predicted along the iso-frequency strip for pure tones with an increase in sound level. This combined increase in both neural excitation and inhibition with increasing sound level may provide the basis for the function observed in the present study.

B. Imaging central auditory responses to sound level

Although intensity encoding occurs throughout the ascending auditory pathway, the small size of subcortical nuclei places them on the limits of detection using current imaging techniques. Using PET, sound-level effects have been observed in a subcortical site putatively near the medial geniculate nucleus (MGN) of the thalamus (Lockwood *et al.*, 1999). This result has not been replicated using the better spatial resolution of fMRI, possibly due to (i) increased physiological motion in the brainstem region, and (ii) masking by the background scanner noise. Synchronization of the image acquisition to a fixed point in the cardiac cycle may be required in order to detect brainstem activation reliably (Guimares *et al.*, 1998) by reducing confounding signal variance from cardiac-related pulsatile brainstem motion. Temporally sparse imaging should also reduce the baseline level of activation to the scanner noise at all levels of the auditory pathway. Thus, it may be profitable for future studies to image subcortical structures using cardiac gating in conjunction with a low-noise imaging sequence.

At a higher cortical level, activation has also been reported in a discrete region of the *posterior* superior temporal gyrus in the right hemisphere (Belin *et al.*, 1998)—a region probably located beyond secondary auditory fields. Activation in this brain region was correlated with the difficulty of an intensity-discrimination judgment rather than with sound level *per se*. Belin *et al.* attributed the function of this region to the computation of sound-intensity differences. Although our task involved only a simple intensity discrimination, it nonetheless required a sound-level difference judgment and hence one might predict activation in the right posterior superior temporal gyrus. The region of auditory activation did include posterior sections of superior temporal gyrus, but due to the lack of fine spatial localization of activation, a direct comparison of the localization of activation between the two studies is not appropriate.

C. Consistency between extent and magnitude measurements of the response to sound level

For increases in sound level, previous studies have generally reported significant increases in the spatial extent of activation and/or the response magnitude (e.g., Jäncke *et al.*, 1998; Lockwood *et al.*, 1999; Mohr *et al.*, 1999). Growth in both the extent and magnitude of the response with sound level is physiologically consistent with a regional increase in the metabolic demands of the underlying neuronal population. Growth in the magnitude of the BOLD response may reflect a greater change in blood oxygenation as a result of increased neuronal metabolism, while growth in the extent may reflect either recruitment of a larger neuronal population or increased response contribution from draining veins relative to that from the capillary bed.

In general, the extent of activation and the response magnitude both seem to increase as a function of sound level, but this has rarely been demonstrated in the same study. Surprisingly, studies that do report data for both activation measures have failed to demonstrate a clear and consistent effect of stimulus intensity on either one or the other measure. For example, using fMRI, Jäncke *et al.* (1998) found a significant increase in extent for syllables and pure tones presented at levels of 75, 85, and 95 dB SPL, but no effect on response magnitude. On the other hand, Mohr *et al.* (1999) found a reliable effect of sound level of monosyllabic words on response magnitude, but not on extent: response magnitude increased significantly as a function of intensity (ranging from 65 to 110 dB C-weighted), and this increase was fairly consistent across subjects, but the extent of activation varied more widely, and less systematically, across subjects. In a PET study however, significant sound-level effects on *both* the extent and magnitude of auditory activation were demonstrated for pure tones (Lockwood *et al.*, 1999). Unlike fMRI, PET produces no intense acoustic noise that may confound the stimulus-driven response. However, the interpretation of the relationship between response magnitude and sound level may not be straightforward because the data suggest an interaction between frequency and level. For example, for a 4-kHz tone, the number of activated voxels in the auditory system and the response magnitude increased strongly across the full 30–90-dB SL range, but, for a 500-Hz tone, there was no change in the extent and magnitude of activation between 70 and 90 dB HL, suggesting an approach towards response saturation. There is also some evidence for saturation of the response measured using fMRI, as Mohr *et al.*'s (1999) data suggest that the response magnitude reached a ceiling level at 90 to 95 dB.

In the present study, despite general intersubject variation, sound level exerted a significant influence across both

stimulus types on magnitude, but not on extent of response. Evidence from a reproducibility study concurs with the implication that extent is a less reliable measure of the underlying pattern of activation than is magnitude (Mohr *et al.*, 1999). Up to fourfold variations occurred in the extent of activation between experiments for the same individual, but the response magnitude within activated voxels varied little across sessions. We propose the following explanation for these apparent differences in reliability between measures. ''Extent'' of activation is defined as the number of voxels whose response exceeds a given probability threshold, and so extent can encompass one or more auditory fields. An increase in extent will change the relative contributions of activations arising from these different auditory fields. For example, a large region of activation encompasses a greater proportion of secondary auditory fields compared with the primary auditory field than does a small region. As a consequence, across sound levels for example, increased extent of activation implies increasingly heterogeneous underlying neuronal and neurovascular responses. In contrast, response magnitude is calculated for voxels within a region that is uniquely specified by the omnibus *F* test and is fixed across sound-level conditions. Even if multiple auditory fields are embraced, the greater reliability in the effect of sound level on magnitude can be attributed to the contribution of a response that is at least based in the same neurovascular region for a given subject. Hence, less intrasubject variability would be expected in response magnitude than in response extent.

D. Cortical responses to sound level and bandwidth—Which is the more potent acoustical feature for fMRI?

When characterizing sound-level effects, an increase in activation with intensity will often be hard to dissociate from an increase in activation with loudness because; for pure tones presented in quiet, loudness is a simple monotonic function of intensity (Stevens, 1975). The direct relationship between intensity and loudness can sometimes break down because, when intensity of a sound is held constant, the loudness of that sound changes (i) in background noise versus in quiet and (ii) as a function of bandwidth.

First, the loudness of a pure tone is reduced by presenting it in a broadband noise (partial masking). This point is of specific relevance to auditory fMRI studies because the intense background noise generated by the scanner would effectively mask a tone stimulus if the signals overlapped in time. Partial masking may arise by suppression of the tone by the noise (Moore *et al.*, 1985), where suppression reduces the excitation level evoked by the tone. Partial masking would encumber interpretation of the effect of stimulus loudness in any fMRI study that uses a conventional imaging protocol in which sets of images are acquired at a rapid rate and image acquisition coincides with stimulus presentation. However, it is unlikely to occur in the present study because the temporally sparse imaging separated the stimulus delivery from the image acquisition (and hence the scanner noise).

With respect to point (ii), the sound levels presented in this study achieve some dissociation between intensity and loudness because the set of stimuli embraces such a bandwidth difference. Extent and magnitude significantly correlated with the sound level measured in phons but not in SPL. Since the cortical response to sound level appears to reflect loudness more closely than SPL, loudness may be an important aspect of cortical encoding. However, clearer demonstrations of a dissociation between the cortical response to intensity and loudness are required before a firm conclusion can be drawn. The data also suggest that bandwidth contributes more strongly to the pattern of activation than sound level. For pairs of tones matched for the same increase in loudness, a greater disparity in the extent of activation was observed when the bandwidth was also increased between tones. Additional activation induced by the complex tone relative to the pure tone was observed spreading mediolaterally along the superior temporal gyrus. Thus, even when sounds of a different bandwidth are matched on the most stringent measure of sound level (i.e., for the loudness match, a complex tone was matched to a pure tone presented at a greater SPL), the bandwidth itself also contributes to the observed activation pattern. This result indicates that potential mismatches in sound level are unlikely to account for much of the increase in the spread of cortical activation for broadband signals relative to pure tones, since the pattern of cortical activation measured was mostly driven by the bandwidth itself. This finding rules out one of the four possible interpretations of the bandwidth effect proposed in the Introduction. It is more likely that the greater spread of activation for the complex tone relative to the pure tone is due either to (i) stimuli of a greater bandwidth being more effective than pure tones in driving a response in nonprimary auditory fields that lie medially and laterally along the superior temporal gyrus, or (ii) stimuli of a greater bandwidth leading to a spread of activity across frequency channels in these tonotopically organized brain regions.

ACKNOWLEDGMENTS

This work was partly supported by a Special Project Grant, No. G9302591, from the Medical Research Council to the Magnetic Resonance Centre, University of Nottingham. We are grateful to Mr. Miguel Gonçalves for assistance with the data analysis, and to Dr. Brian Glasberg for advice on the implementation of the loudness model.

[1]As part of the initial stimulus calibration, three of the authors performed an experimental loudness match by adjusting the level of the pure tone until it most closely matched that of the 85-dB SPL complex tone. The authors judged the 86-dB SPL pure tone to match the loudness of the 85-dB complex tone. This experimental match should have yielded approximately the same result as that computed by the loudness model. However, the model assigned the 86-dB pure tone a loudness of 77.3 phons and the 85-dB SPL complex tone a loudness of 86.2 phons (a difference of 8.9 phons). The underestimation of the experimental loudness match may reflect the particular decision criteria used by the three listeners. Because of the uncertainty of the reliability of the experimental match, this is not reported.

Bandettini, P. A., Jesmanowicz, A., Van Kylen, J., Birn, R. M., and Hyde, J. S. (**1998**). ''Functional MRI of brain activation induced by scanner acoustic noise,'' Magn. Reson. Med. **39**, 410–416.

Belin, P., McAdams, S., Smith, B., Savel, S., Thivard, L., Samson, S., and Samson, Y. (**1998**). ''The functional anatomy of sound intensity discrimination,'' J. Neurosci. **18**, 6388–6394.

Bowtell, R., Mansfield, P., Coxon, R.J., Harvey, P. R., and Glover, P. M. (**1994**). ''High-resolution echo-planar imaging at 3.0 T,'' MAGMA (N.Y.) **2**, 241–245.

Bowtell, R. W., and Peters, A. (**1999**). ''Analytic approach to the design of transverse gradient coils with co-axial return paths,'' Magn. Reson. Med. **41**, 600–608.

Broch, M., and Schreiner, C. E. (**1999**). ''Correlations between neural discharges are related to receptive field properties in cat primary auditory cortex,'' Eur. J. Neurosci. **11**, 3517–3530.

Brugge, J. F., and Merzenich, M. M. (**1973**). ''Responses of neutrons in auditory cortex of the macaque monkey to monaural and binaural stimulation,'' J. Neurophysiol. **36**, 1138–1158.

Burkhard, M. D., and Sachs, R. M. (**1975**). ''Anthropometric manikin for acoustic research,'' J. Acoust. Soc. Am. **58**, 214–222.

Buus, S., Musch, H., and Florentine, M. (**1998**). ''On loudness at threshold,'' J. Acoust. Soc. Am. **104**, 399–410.

Carlyon, R. P., and Moore, B. C. J. (**1975**). ''Intensity discrimination: A severe departure from Weber's law,'' J. Acoust. Soc. Am. **76**, 1369–1376.

Chatterjee, M., and Zwislocki, J. J. (**1998**). ''Cochlear mechanisms of frequency and intensity coding. II. Dynamic range and the code for loudness,'' Hear. Res. **124**, 170–181.

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.

Dejardin, S., Dubois, S., Bodart, J. M., Schiltz, C., Delinte, A., Michel, C., Roucoux, A., and Crommelinck, M. (**1998**). ''PET study of human voluntary saccadic eye movements in darkness: Effect of task repetition on the activation pattern,'' Eur. J. Neurosci. **10**, 2328–2336.

Foster, J. R., Hall, D. A., Summerfield, A. Q., Palmer, A. R., and Bowtell, R. W. (**2000**). ''Sound-level measurements and calculations of safe noise dosage during fMRI at 3T,'' J. Magn. Reson. Imaging **12**, 157–163.

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C. R., Frackowiak, R. S. J., and Turner, R. (**1995a**). ''Analysis of fMRI time-series revisited,'' NeuroImage **2**, 45–53.

Friston, K. J., Ashburner, J., Frith, C. D., Poline, J. B., Heather, J. D., and Frackowiak, R. S. J. (**1995b**). ''Spatial registration and normalization of images,'' Hum. Brain Mapping **3**, 165–189.

Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. J., and Turner, R. (**1996**). ''Movement-related effects in fMRI time-series,'' Magn. Reson. Med. **35**, 346–355.

Greenwood, D. D., and Maruyama, N. (**1965**). ''Excitatory and inhibitory response areas of auditory neurons in the cochlear nucleus,'' J. Neurophysiol. **28**, 863–892.

Griffiths, T. D., Büchel, C., Frackowiak, R. S. J., and Patterson, R. D. (**1998**) ''Analysis of temporal structure in sound by the human brain,'' Nat. Neurosci. **1**, 422–427.

Guimaraes, A. R., Melcher, J. R., Talavage, T. M., Baker, J. R., Ledden, P., Rosen, B. R., Kiang, N. Y. S., Fullerton, B. C., and Weisskoff, R. M. (**1998**). ''Imaging subcortical auditory activity in humans,'' Hum. Brain Mapping **6**, 33–41.

Hall, D. A., Elliott, M. R., Bowtell, R. W., Gurney, E., and Haggard, M. P. (**1998**). ''Sparse temporal sampling in fMRI enhances detection of activation by sound for both magnetic and acoustical reasons,'' NeuroImage **7**, S551.

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., Gurney, E., and Bowtell, R. W. (**1999**). ''Sparse temporal sampling in auditory fMRI,'' Hum. Brain Mapping. **7**, 213–223.

Hall, D. A., Summerfield, A. Q., Goncalves, M. S., Foster, J. R., Palmer, A. R., and Bowtell, R. W. (**2000**) ''Time-course of the auditory BOLD response to scanner noise,'' Magn. Reson. Med. **43**, 601–606.

Heil, P., Rajan, R., and Irvine, D. R. F. (**1994**). ''Topographic representation of tone intensity along the isofrequency axis of cat primary auditory cortex,'' Hear. Res. **76**, 188–202.

Jäncke, L., Shah, N. J., Posse, S., Grosse-Ryuken, M., and Müller-Gärtner, H. W. (**1998**). ''Intensity coding of auditory stimuli: An fMRI study,'' Neuropsychologia **36**, 875–883.

Liberman, M. C. (**1978**) ''Auditory-nerve responses from cats raised in a low-noise chamber,'' J. Acoust. Soc. Am. **63**, 442–455.

Lockwood, A. H., Salvi, R. J., Coad, M. L., Arnold, S. A., Wack, D. S., Murphy, B. W., and Burkard, R. F. (**1999**). ''The functional anatomy of the normal human auditory system: responses to 0.5 and 4.0 kHz tones at varied intensities,'' Cereb. Cortex **9**, 65–76.

Manunta, Y., and Edeline, J. M. (**1998**). ''Effects of noradrenaline on rate-level function of auditory cortex neurons: Is there a ''gating'' effect of noradrenaline?,'' Exp. Brain Res. **118**, 361–372.

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.

Millen, S. J., Haughton, V. M., and Yetkin, Z. (**1995**). ''Functional magnetic resonance imaging of the central auditory pathway following speech and pure-tone stimuli,'' Laryngoscope **105**, 1305–1310.

Mohr, C. M., King, W. M., Freeman, A. J., Briggs, R. W., and Leonard, C. M. (**1999**). ''Influence of speech stimuli intensity on the activation of auditory cortex investigated with functional magnetic resonance imaging,'' J. Acoust. Soc. Am. **105**, 2738–2745.

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.

Moore, B. C. J., and Glasberg, B. R. (**1996**). ''A revision of Zwicker's loudness model,'' Acta Acust. **82**, 335–345.

Moore, B. C. J., Glasberg, B. R., Hess, R. F., and Birchall, J. P. (**1985**). ''Effects of flanking noise bands on the rate of growth of loudness of tones in normal and recruiting ears,'' J. Acoust. Soc. Am. **77**, 1505–1515.

Olsen, S. O. (**1999**) ''The relationship between the uncomfortable loudness level and the acoustic reflex threshold for pure tones in normally hearing and impaired listeners—A meta-analysis,'' Audiology **38**, 61–68.

Olsen, S. O., Rasmussen, A. N., Nielsen, L. H., and Borgkvist, B. V. (**1999**). ''The acoustic reflex threshold: Not predictive for loudness perception in normally hearing listeners,'' Audiology **38**, 303–307.

Palmer, A. R., Bullock, D. C., and Chambers, J. D. (**1998**). ''A high-output, high-quality sound system for use in auditory fMRI,'' NeuroImage **7**, S359.

Penhune, V. B., Zatorre, R. J., MacDonald, J. D., and Evans, A. C. (**1996**). ''Interhemispheric anatomical differences in human primary auditory cortex: Probabilistic mapping and volume measurement from magnetic resonance scans,'' Cereb. Cortex **6**, 661–672.

Phillips, D. P., and Orman, S. S. (**1984**). ''Responses of single neurons in posterior field of cat auditory cortex to tonal stimulation,'' J. Neurophysiol. **51**, 147–163.

Picard, N., and Strick, P. L. (**1996**). ''Motor areas of the medial wall: A review of their location and functional activation,'' Cereb. Cortex **6**, 342–353.

Rauschecker, J. P., Tian, B., Pons, T., and Mishkin, M. (**1997**). ''Serial and parallel processing in Rhesus monkey auditory cortex,'' J. Comp. Neurol. **382**, 89–103.

Recanzone, G. H., Schreiner, C. E., Sutter, M. L., Beitel, R. E., and Merzenich, M. M. (**1999**). ''Functional organization of spectral receptive fields in the primary auditory cortex of the owl monkey,'' J. Comp. Neurol. **415**, 460–481.

Redies, H., Sieben, U., and Creutzfeldt, O. D. (**1989**). ''Functional subdivisions in the auditory cortex of the guinea pig,'' J. Comp. Neurol. **282**, 473–488.

Rivier, F., and Clarke, S. (**1997**). ''Cytochrome oxidase, acetylcholinesterase, and NADPH diaphorase staining in human supratemporal and insular cortex: Evidence for multiple auditory areas,'' NeuroImage **6**, 288–304.

Shah, N. J., Jäncke, L., Grosse-Ruyken, M-L., Müller-Gärtner, H. W. (**1999**). ''Influence of acoustic masking noise in fMRI of the auditory cortex during phoentic discrimination,'' J. Magn. Reson. Imaging **9**, 19–25.

Shah, N. J., Steinhoff, S., Mirzazade, S., Zafiris, O., Grosse-Ruyken, M-L., Jäncke, L., and Zilles, K. (**2000**). ''The effect of sequence repeat time on auditory cortex stimulation during phonetic discrimination,'' NeuroImage **12**, 100–108.

Stevens, S. S. (**1975**). *Psychophysics* (Wiley, New York), p. 28.

Stevens, S. S., and Guirao, M. (**1962**). ''Loudness, reciprocality, and partition scales,'' J. Acoust. Soc. Am. **34**, 1466–1471.

Strainer, J. C., Ulmer, J. L., Yetkin, F. Z., Daniels, D. L., and Millen, S. J. (**1997**). ''Functional MR of the primary auditory cortex: An analysis of pure tone activation and tone discrimination,'' AJNR **18**, 601–610.

Talavage, T. M., Ledden, P. J., Sereno, M. I., Rosen, B. R., and Dale, A. M. (**1997**). ''Multiple phase-encoded tonotopic maps in human auditory cortex,'' NeuroImage **5**, S8.

Talavage, T. M., Edmister, W. B., Ledden, P. J., and Weiskoff, R. M. (**1999**). ''Quantitative assessment of auditory cortex responses induced by

imager acoustic noise,'' Hum. Brain Mapping **7**, 79–88.
Taniguchi, I., and Nasu, M. (**1993**). ''Spatio-temporal representation of sound intensity in the guinea pig auditory cortex observed by optical recording,'' Neurosci. Lett. **151**, 178–181.
Thomas, H., Tillein, J., Heil, P., and Scheich, H. (**1993**). ''Functional organization of auditory cortex in the Mongolian gerbil *(Meriones unguiculatus)* I. Electrophysiological mapping of frequency representation and distinction of fields,'' Eur. J. Neurosci. **5**, 882–897.
Viemeister, N. F., and Bacon, S. P. (**1988**). ''Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones,'' J. Acoust. Soc. Am. **84**, 172–178.
Wang, J., Caspary, D., and Salvi, R. J. (**2000**). ''GABA-A antagonist causes dramatic expansion of tuning in primary auditory cortex,'' NeuroReport **11**, 1137–1140.
Wessinger, C. M., Tian, B., Japikse, K. C., Ghosh, S., Van Meter, J. W., Platenberg, R. C., Pekar, J., and Rauschecker, J. P. (**1997**). ''Processing of complex sounds in human auditory cortex,'' Proceedings of the Annual Mid-winter Meeting of the Association for Research in Otolaryngology, #105, p. 27.
Westbury, C. F., Zatorre, R. J., and Evans, A. C. (**1999**). ''Quantifying variability in the planum temporale: A probability map,'' Cere. Cortex **9**, 392–405.
Yang, L. C., Pollak, G. D., and Resler, C. (**1992**). ''GABAergic circuits sharpen tuning curves and modify response properties in the moustache bat inferior colliculus,'' J. Neurophysiol. **68**, 1760–1774.
Zwicker, E., Flottorp, G., and Stevens, S. S. (**1957**). ''Critical bandwidth in loudness summation,'' J. Acoust. Soc. Am. **29**, 548–557.

The effect of basilar-membrane nonlinearity on the shapes of masking period patterns in normal and impaired hearing[a)]

Magdalena Wojtczak,[b)] Anna C. Schroder, Ying-Yee Kong, and David A. Nelson
Clinical Psychoacoustics Laboratory, University of Minnesota, 420 Delaware Street SE, Minneapolis, Minnesota 55455

(Received 29 August 2000; revised 28 November 2000; accepted 23 January 2001)

Masking period patterns (MPPs) were measured in listeners with normal and impaired hearing using amplitude-modulated tonal maskers and short tonal probes. The frequency of the masker was either the same as the frequency of the probe (on-frequency masking) or was one octave below the frequency of the probe (off-frequency masking). In experiment 1, MPPs were measured for listeners with normal hearing using different masker levels. Carrier frequencies of 3 and 6 kHz were used for the masker. The probe had a frequency of 6 kHz. For all masker levels, the off-frequency MPPs exhibited deeper and longer valleys compared with the on-frequency MPPs. Hearing-impaired listeners were tested in experiment 2. For some hearing-impaired subjects, masker frequencies of 1.5 kHz and 3 kHz were paired with a probe frequency of 3 kHz. MPPs measured for listeners with hearing loss had similar shapes for on- and off-frequency maskers. It was hypothesized that the shapes of MPPs reflect nonlinear processing at the level of the basilar membrane in normal hearing and more linear processing in impaired hearing. A model assuming different cochlear gains for normal versus impaired hearing and similar parameters of the temporal integrator for both groups of listeners successfully predicted the MPPs.
[DOI: 10.1121/1.1356702]

PACS numbers: 43.66.Ba, 43.66.Dc, 43.66.Mk, 43.66.Sr [SPB]

I. INTRODUCTION

Masking produced by stimuli with fluctuating envelopes has been shown to depend on the rate of fluctuations, the depth of fluctuations, and on the relative frequencies of the masker and the signal (e.g., Zwicker, 1976a, b; Zwicker and Schorn, 1982; Nelson and Swain, 1996; Bacon *et al.*, 1997). Different experimental paradigms have been used to study various aspects of masking by these stimuli. Zwicker (1976a, b, 1986) measured masking period patterns (MPPs) in order to assess temporal resolution in auditory processing and to relate it to the understanding of speech. An MPP is obtained when detection threshold for a short stimulus, or probe, is measured as a function of the temporal position of the probe within a modulated masker. Zwicker and Schorn (1982) suggested that there is no need to ''track'' the whole modulation cycle to assess temporal limitations of envelope processing. They postulated that temporal resolution could be described based on the difference between the masked threshold observed with the probe presented at a peak of a modulated masker and that observed with the probe placed in a valley. Their data showed that this difference decreased with increasing modulation rate of the masker suggesting limited temporal resolution of auditory processing.

It is questionable whether MPPs or simply the difference between thresholds measured at a peak and in a valley of a modulated masker provide information about temporal processing that is not confounded by other factors. Zwicker (1976c) measured MPPs using tonal maskers and probes having either the same or different frequencies. These two masking conditions will be referred to as on-frequency masking (masker frequency equals probe frequency) and off-frequency masking (masker frequency below probe frequency) throughout this paper. Although Zwicker's data were most likely affected by energy splatter produced by the short probe, due to the relatively low frequencies used in this study, they clearly suggest that the shapes of MPPs depend on the relative frequencies of the masker and the probe. In particular, his data show that the dynamic range of MPPs is substantially larger for off-frequency masking than it is for probe frequencies near the masker frequency. Such a result suggests that temporal resolution of fluctuating envelopes is better within the upper accessory excitation produced by the masker (the place above that corresponding to the masker frequency), than it is in the region of the main excitation (the place corresponding to the masker frequency). A similar result was observed by Nelson and Swain (1996) who used probes that did not produce detectable splatter. The probe in their study spanned one cycle of the 100%-modulated masker (starting and ending at zero amplitude) and was temporally positioned so that the peak of the probe coincided either with a peak or with a valley of the masker. Nelson and Swain found that the difference between the thresholds measured with the probe at the peak and in the valley was relatively large in the off-frequency masking condition and was much smaller in the on-frequency masking condition. From the masked thresholds measured at the peak and in the valley for different masker levels, Nelson and Swain constructed

[a)]Portions of these data were presented at the 138th meeting of the Acoustical Society of America [M. Wojtczak, A. C. Schroder, Y.-Y. Kong, and D. A. Nelson, J. Acoust. Soc. Am. **106**, 2147(A) (1999)] and the 139th meeting of the Acoustical Society of America [M. Wojtczak, A. C. Schroder, and D. A. Nelson, J. Acoust. Soc. Am. **107**, 2881(A) (2001)].

[b)]Electronic mail: wojtc00l@tc.umn.edu

growth-of-masking functions for each masker-probe combination. For the off-frequency case, the rate of masking growth was faster than for the on-frequency conditions. This was true both at the peak and in the valley, though the difference between the slopes observed in the valley was much smaller than the difference observed at the peak. The steeper growth of masking observed with an off-frequency masker is in agreement with the larger dynamic range of the MPP produced by such maskers (Zwicker, 1976c).

The greater peak-valley difference measured with off-frequency maskers was originally explained in terms of nonlinear growth of simultaneous masking occurring in the region of a peak. It was suggested that the nonlinear growth of masking was mediated by nonlinear neural responses and suppression (Javel *et al.*, 1983; Delgutte, 1990a, b). However, recent physiological and psychophysical studies shed new light onto the observed differences in the rate of masking growth. Ruggero (1992) and Ruggero *et al.* (1997) measured basilar-membrane (BM) responses to tonal stimulation in the chinchilla cochlea. Their data show that a characteristic-frequency (CF) tone produces a BM response that grows in a compressive manner with the tone's intensity. In contrast, a below-CF tone produces a response that grows linearly with intensity. Nonlinear responses of the BM were described earlier by Rhode (1971), Sellick *et al.* (1982), and Robles *et al.* (1986), but their data were obtained using a much less accurate technique, and thus prevented construction of a detailed BM input–output function. The compressive growth of response is believed to reflect the operation of the so-called active mechanism in the cochlea, which was first discovered by Kemp (1978). In a healthy cochlea, this mechanism applies a level-dependent gain to the BM response resulting in strong amplification of the softest stimuli and progressively less amplification of increasingly intense stimuli. As a result of this level-dependent gain, a function exhibiting compressive response growth is observed at the output of the cochlea. When there is damage to the cochlea, to outer hair cells in particular, the gain is reduced and can even be completely eliminated. The reduced gain results in a linearization of BM-response growth (Ruggero and Rich, 1991; Ruggero *et al.*, 1993, 1995, 1996).

The description of the BM mechanical response triggered a series of psychophysical experiments designed to test the consequences of the nonlinear BM behavior. Oxenham and Plack (1997) used a forward-masking paradigm to measure growth of masking produced by an on- and off-frequency masker in listeners with normal and impaired hearing in the presence of a high-pass noise. The growth-of-masking function (described by Oxenham and Plack as the masker level required to just mask the probe, plotted as a function of the probe level) observed in normal-hearing listeners with the off-frequency masker had a substantially shallower slope than the function observed for on-frequency masking. This is in agreement with physiological data, which demonstrate that the BM response to below-CF tones grows faster than the BM response to CF tones (e.g., Ruggero *et al.*, 1997). A different result was observed for the hearing-impaired listeners tested by Oxenham and Plack. For those listeners, the slopes of the masking-growth functions measured with on- and off-frequency maskers were similar. This effect presumably reflects the lack of the active process in those listeners. Based on their data, Oxenham and Plack estimated the rate of response growth for the probe in normal hearing. They accomplished this by computing a ratio between the slopes measured with the off- and on-frequency maskers. This ratio turned out to be very similar to the slope of the function describing BM-response growth in the study by Ruggero *et al.* (1997), supporting the idea that the nonlinear growth of masking observed psychophysically reflects nonlinear response growth at the BM. Further support for the connection between the nonlinear BM behavior and psychophysically observed differences in growth of masking produced by on- and off-frequency maskers was provided by the studies of Nelson and Schroder (1996, 1997), Gregan *et al.* (1998), Hicks and Bacon (1999a, b), and Moore *et al.* (1999).

In the present study, MPPs were measured using on- and off-frequency maskers. The study was primarily motivated by the finding that for maskers with fluctuating envelopes, temporal resolution of the envelopes is better in the upper accessory excitation produced by the masker than it is in the region of the main excitation. The goal of the study was to demonstrate that the seemingly better temporal resolution exhibited for off-frequency maskers reflects nonlinear BM behavior. It is generally believed that temporal processing takes place at a higher stage of the auditory system, whereas the compressive transformation of the stimulus occurs at the level of the BM. It is hypothesized that the differences observed in the measurement of on- and off-frequency MPPs reflect BM nonlinearity rather than an actual difference in temporal processing. In experiment 1, normal-hearing listeners were tested using different masker levels. The MPPs were measured in great detail to establish differences in the shapes of the masking patterns, as measured with on-frequency versus off-frequency maskers (masker frequency an octave below the probe frequency). A model was used to predict the shapes of the measured MPPs. While the model assumed different rates of growth of the response to stimulation by on- and off-frequency maskers at the probe-frequency place, it assumed that the same temporal integrator was operating on the internal representation of the input stimulus in the two masking conditions. In experiment 2, MPPs were measured in listeners with hearing loss of cochlear origin. This was done in order to test whether the differences in the shapes of the on- and off-frequency MPPs are much smaller or practically nonexistent in the case of significant hearing loss, where BM response is more linear.

There is another important aspect to this study. An MPP reflects a combination of simultaneous and nonsimultaneous (forward and backward) masking. The model presented here provides satisfactory predictions of the effects of this combined masking. Although additivity of masking has been modeled in past studies (e.g., Robinson and Pollack, 1973; Penner, 1980; Penner and Shiffrin, 1980; Lutfi, 1983; Moore *et al.*, 1988; Plack and Moore, 1990; Oxenham and Moore, 1994, 1995), to our knowledge none of the models allowed the prediction of masked thresholds in the case where the masker and the signal differed in frequency and overlapped

in time. The present study proposes a way to deal with such a complex masking situation.

II. EXPERIMENT 1

To establish differences in the shapes of MPPs observed with on- and off-frequency maskers in normal-hearing listeners, detection threshold for a short probe was measured as a function of the temporal position of the probe within one period of a modulated masker. The experiment was similar to that described by Zwicker (1976c), but higher frequencies and a longer probe duration were used to avoid detection of energy splatter in channels tuned to frequencies different from the probe frequency.

A. Subjects

MPPs were measured in four listeners with normal hearing, but not every subject was available for testing in each condition. Absolute thresholds in these listeners conformed to ANSI standards (1996) for normal hearing at all audiometric frequencies. All subjects had previous experience in psychoacoustic tasks. Subject S1 was the third author and subject S2 was the first author. The other two subjects were paid for their services.

B. Conditions and stimuli

MPPs were measured using a short tonal probe presented at different times during one envelope cycle of a longer-duration sinusoidally amplitude-modulated masker. The probe was gated on for 8 ms, which included a 4-ms steady-state portion and 2-ms raised-cosine ramps. A probe frequency of 6 kHz was used to avoid detection of splatter outside the critical band, due to the relatively short rise and decay times. The masker had a total duration of 512 ms and was 100% modulated at a 3.91-Hz rate, resulting in two full 256-ms modulation cycles during masker presentation. The modulating tone was a cosine starting at the 180-deg phase, thus the masker started and ended at zero amplitude and comprised two envelope peaks and one centered valley, as shown in Fig. 1.

At the beginning of each session, thresholds in quiet were measured separately for the modulated masker and for the probe. Subsequently, masked thresholds for the probe were measured in the presence of the masker with the probe presented at different modulator phases between the two peaks in the masker envelope. Two masker frequencies were used in the experiment: a 3-kHz masker was used to measure off-frequency MPPs and a 6-kHz masker was used to measure on-frequency MPPs. The probe always started with a phase of 90 deg relative to the masker phase at the probe's onset. On- and off-frequency MPPs were measured at masker levels of 65, 75, 85, and 90 dB SPL.

The stimuli were generated digitally on a computer and played out through separate channels of the Tucker-Davis Technologies (TDT) digital-to-analog converter (DD1) at a sampling rate of 50 kHz. The masker and probe were low-pass filtered at 20 kHz (TDT FT6-2) to prevent aliasing and routed to separate attenuators (TDT PA4s). After attenuation, the stimuli were added electrically (TDT SM3) and passed (via a TDT HB6) to a TDH-49 earphone for monaural presentation.

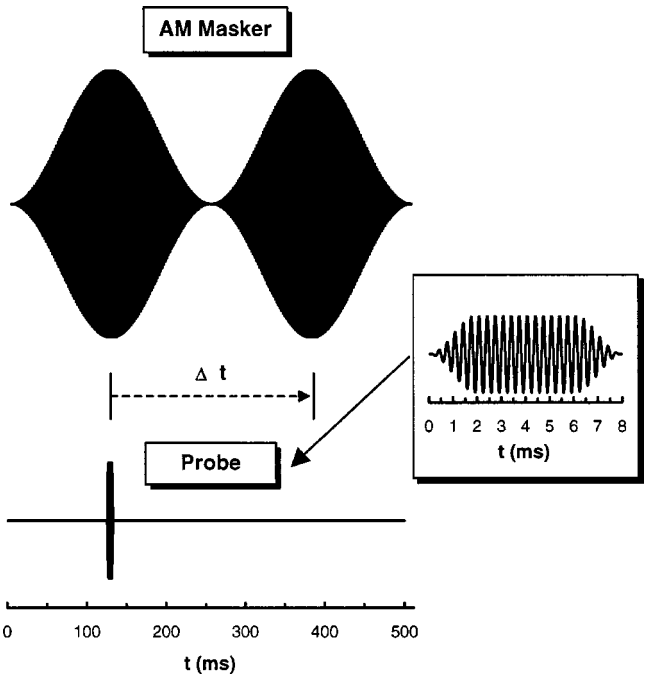


FIG. 1. Schematic illustration of the stimuli used to measure MPPs. A modulated masker is shown in the top section and the probe is shown at the bottom (also magnified in the insert). Time interval Δt marks the cycle of masker envelope, within which masked thresholds for the probe were measured.

C. Procedure

All thresholds were measured using a three-interval forced-choice procedure. The three observation intervals were marked by lights and were separated from one another by 250 ms of silence. For the thresholds measured in quiet, the signal appeared randomly in one of the observation intervals while the other two were silent. When the MPPs were measured, the signal interval contained the masker and the probe, while the other two intervals contained the masker alone. At the beginning of each track, the signal was presented at a highly detectable level. Initially, the signal level was decreased by 8 dB after each correct response and increased by the same amount after each incorrect response until four reversals were obtained. The next two reversals were also obtained using a 1-up/1-down paradigm, but the step size was reduced to 2 dB. This was done in order to quickly reach the range of levels around threshold. Subsequently, another six reversals were obtained using 2-dB steps and a 2-down, 1-up technique, which estimates the 70.7% point on the psychometric function (Levitt, 1971). The threshold was computed as the mean of the last six reversals. After each trial, up to 4 s were allowed for the subject's response. If no response was registered during this period, the trial was repeated with the signal presented at the same level in a randomly selected interval. Feedback indicating the correct interval was provided after each response. Final data points were computed based on at least three threshold esti-

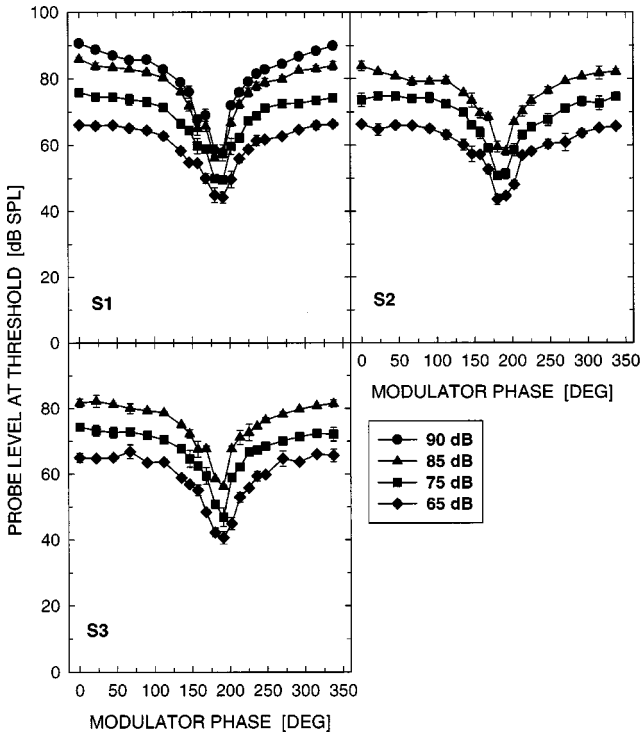


FIG. 2. Individual MPPs measured with a 6-kHz masker and a 6-kHz probe (on-frequency masking) in listeners with normal hearing. Different symbols indicate data obtained at different masker levels. Each panel shows data for a different listener.

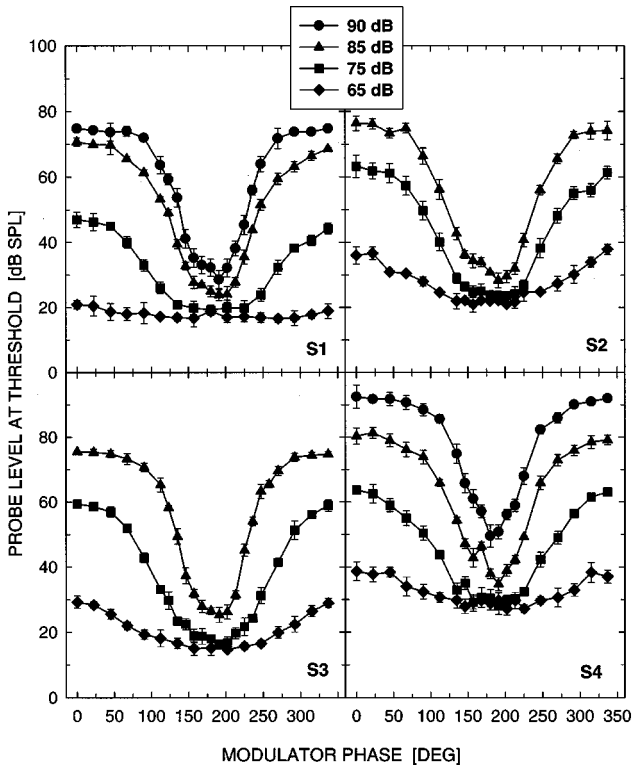


FIG. 3. Individual MPPs measured with a 3-kHz masker and a 6-kHz probe (off-frequency masking) in listeners with normal hearing. Different symbols indicate data obtained at different masker levels. Each panel shows data for a different listener.

mates. More threshold estimates were obtained when the standard deviation for the averaged threshold was larger than 3 dB. It took 35–45 min to obtain an MPP with a single threshold estimate for each temporal position of the probe within the modulation cycle of the masker. Two such masking patterns were typically obtained during one experimental session and the listeners were given a 10-min break between the two runs.

III. RESULTS AND DISCUSSION

Individual masked thresholds measured with the on-frequency masker are shown in Fig. 2. They are plotted as a function of the modulator phase within one modulation cycle of the masker, assuming that the cycle of interest begins at the phase corresponding to the first peak of the masker (0 deg) and ends at the phase corresponding to the second peak of the masker [(360 deg) see Fig. 1]. Separate panels show data obtained from three normal-hearing listeners. Different symbols represent MPPs observed for different masker levels, noted within the insert. These on-frequency MPPs have very short, sharp valleys at all masker levels used in this experiment. The average peak-valley differences were 23 dB, 25.5 dB, 26 dB, and 32.9 dB for masker levels of 65, 75, 85, and 90 dB, respectively. The observed increase in peak-valley difference with increasing masker level reflects slower growth of masking in the valley compared with growth of masking at the peak. This is consistent with the notion that thresholds measured at the peak of a modulated masker are governed by simultaneous masking whereas thresholds measured in a valley are governed mainly by forward masking from the preceding peak. It has been shown that the growth-of-masking function measured in forward masking has a slope that is less than one (Jesteadt *et al.*, 1982; Moore and Glasberg, 1983) and the growth-of-masking slope for simultaneous masking is close to one (Hawkins and Stevens, 1950; Viemeister, 1972). Gregan *et al.* (1998) specifically compared growth-of-masking functions measured at a peak and in a valley of a sinusoidally-modulated masker with growth-of-masking functions measured in simultaneous and forward masking using an unmodulated masker. Their data strongly support the idea that the threshold measured in the valley of a modulated masker reflects forward masking by the preceding peak, whereas the threshold measured at a peak results from simultaneous masking. It has been demonstrated that nonlinear growth of forward masking may be attributed to the nonlinear growth of BM response (Oxenham and Moore, 1995; Plack and Oxenham, 1998). There is evidence based on physiological study that compression on the BM is fast acting, i.e., stimuli presented immediately one after another are compressed independently (Yates, 1995). When the probe is presented during the valley of a modulated masker, the amount of compression applied to the probe most likely differs from the amount of compression applied to the masker during its peak. In effect, nonlinear growth of masking is observed in the valley.

Figure 3 shows MPPs obtained for the off-frequency masker. The peak-valley differences are substantially larger than those observed with the on-frequency masker, except when the threshold in the valley is limited by absolute threshold for the probe (e.g., at the 65 and 75 dB masker

levels). For masker levels of 85 and 90 dB SPL, the average peak-valley differences were 47.7 dB and 44.6 dB, respectively. This result is in agreement with earlier reports that, for a given masker frequency, the peak-valley difference is especially large when the masker frequency is much lower than the signal frequency (Egan and Hake, 1950; Mott and Feth, 1986; Nelson and Swain, 1996; Gregan *et al.*, 1998). Gregan *et al.* (1998) demonstrated that the larger peak-valley difference observed with an off-frequency masker compared to that difference measured with an on-frequency masker is determined by the relative frequencies of the masker and the signal rather than the frequency of the signal itself.

Studies that were confined to the measurements of masked thresholds at a peak and in a valley of a fluctuating masker do not provide exhaustive information about the effects of relative masker and signal frequencies on MPPs. In particular, they do not show that MPPs measured with an off-frequency masker have valleys that are not only deeper but also much longer than those observed with an on-frequency masker. As a result, there are substantial differences in the shapes of MPPs measured with on- and off-frequency maskers, as shown in Figs. 2 and 3. These differences could be explained in terms of nonlinear BM behavior. When a sinusoidally varying envelope is subjected to a compressive transformation, the envelope peaks of the compressed masker become broader and the valleys become narrower. However, a short on-frequency probe added to the masker during its presentation would be subjected to the same compression as the portion of the masker that coincides with the probe. Therefore, if the temporal processing of the auditory system had unlimited temporal resolution, the shape of MPPs measured in the on-frequency case would precisely reflect the shape of the original uncompressed masker envelope. That is because any given change in masker amplitude would require a proportional change in probe amplitude to reach the threshold (assuming a constant detection criterion for threshold).

A different situation would be observed with an off-frequency sinusoidally amplitude-modulated masker, whose carrier frequency is well below the probe frequency. Such a masker would produce a linear response at the probe-frequency place, whereas the response to the probe would grow in a compressive way. When moving along a cycle of the masker envelope, any change in the instantaneous masker amplitude would require a much greater change in the probe amplitude to reach the criterion for threshold. The change in the probe amplitude would have to be large enough to overcome the compression. Thus under the assumptions of unlimited temporal resolution and a constant detection criterion for threshold, MPPs with narrower peaks and extended valleys would be observed in the off-frequency masking situations. As demonstrated by results of numerous masking and gap detection tasks, auditory processing appears to have a limited temporal resolution. Assuming the same limited resolution of temporal processing in on- and off-frequency masking situation, more extended valleys should still be expected for the off-frequency MPPs due to the nonlinear BM response.

Sluggishness of auditory temporal processing has often been modeled by applying a temporal integrator (Robinson and Pollack, 1973; Viemeister, 1979; Penner and Shiffrin, 1980; Moore *et al.*, 1988; Plack and Moore, 1990; Oxenham and Moore, 1994), which is assumed to be a linear device. At least to a first approximation, the parameters of the temporal integrator are independent of the frequencies of the stimuli and are also independent of their levels.[1] Therefore, the narrow valleys observed with the on-frequency masker in Fig. 2, and the much longer valleys observed with the off-frequency masker in Fig. 3, most likely reflect the combined effects of the BM nonlinearity *and* the integration due to a limited temporal resolution.

IV. EXPERIMENT 2

Studies of masking using on- and off-frequency maskers in hearing-impaired listeners have consistently demonstrated that the amount of off-frequency masking grows more gradually with masker level than it does in listeners with normal hearing (Nelson and Bilger, 1974; Smits and Duifhuis, 1982; Stelmachowicz *et al.*, 1987; Buus and Florentine, 1989; Murnane and Turner, 1991; Nelson and Schroder, 1996, 1997; Oxenham and Plack, 1997). This is consistent with changes observed in the BM mechanical response of a damaged cochlea. Damage to outer hair cells results in a reduced sensitivity to incoming sounds and a more linear growth of response once the stimuli are presented at levels above their detection thresholds (Ruggero and Rich, 1990, 1991; Ruggero *et al.*, 1993, 1996). As a result, in more severely damaged cochleae, BM responses to CF tones and below-CF tones grow at a similar rate. Oxenham and Plack (1997) measured growth-of-masking functions in hearing-impaired listeners and found that the slopes were similar when the functions were measured with on- and off-frequency maskers, and that they were also similar to the slopes observed in listeners with normal hearing for the on-frequency masking condition. All those results indicate that, in hearing-impaired listeners with a sufficient amount of hearing loss, masking measured with an on-frequency masker grows in a similar way to masking measured with an off-frequency masker. In the discussion of the results obtained from normal-hearing listeners in the present study, it was hypothesized that the different shapes of MPPs observed with the on- and off-frequency maskers result from different rates of masking growth. If our reasoning was correct, then for listeners with cochlear hearing loss, similar MPP shapes should be observed in the two masking conditions. Experiment 2 was designed to test this hypothesis.

A. Subjects

Four listeners with sensorineural hearing loss participated in the study. Their absolute thresholds, expressed in dB SPL, are presented by open symbols in Fig. 4. They are plotted along with thresholds for normal hearing represented by the solid curve (ANSI, 1996). Two listeners (S5, S6) had relatively flat hearing losses across all audiometric frequencies. Listener S7 had normal thresholds for the two lowest frequencies, 125 Hz and 250 Hz, and hearing losses at all higher frequencies. The fourth listener (S8) had a high-frequency hearing loss with thresholds that were within the

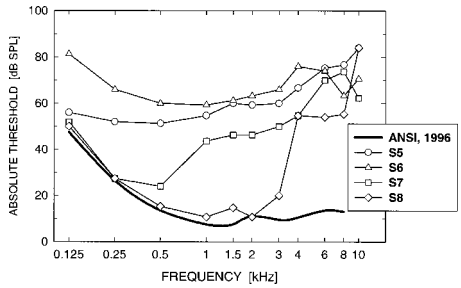


FIG. 4. Absolute thresholds (expressed in dB SPL) measured with 200-ms tones for four listeners with sensorineural hearing loss. The solid line shows the normal threshold curve based on ANSI (1996).

normal range up to about 3 kHz. All four listeners had earlier experience with psychoacoustic tasks and were given practice on the present task until their performance became stable.

B. Stimuli and procedure

MPPs were measured using on- and off-frequency maskers. The duration and gating of the stimuli used to test listeners with impaired hearing were the same as those used in experiment 1, in which only listeners with normal hearing were tested. However, it was not possible to use the same levels of the masker and the same frequencies of the masker and probe in both experiments. To produce a sufficient amount of masking in hearing-impaired listeners, a level of 95 dB SPL was used for the on- and off-frequency maskers. Three out of four listeners (S5, S6, and S7) exhibited substantial hearing loss at a frequency of 6 kHz. Limitations of our equipment prevented us from using probe levels high enough to measure masked thresholds around the peaks of the modulated masker at that frequency for the on-frequency masking condition. Therefore for listeners S5 and S7, a 3-kHz tone was used for the probe and for the carrier of the on-frequency masker. A 1.5-kHz tone was then used as the carrier for the off-frequency masker. Listener S6 was not available for testing with a masker frequency of 1.5 kHz. For this listener, the on-frequency MPP was measured using a 3-kHz probe and the off-frequency MPP was measured using a 6-kHz probe. Only listener S8 was tested using a 6-kHz tone for the probe and the carrier of the on-frequency masker and a 3-kHz tone for the carrier of the off-frequency masker. Similar to experiment 1, a modulation rate of 3.91 Hz was used to measure the MPPs. Because of the reduced dynamic range of hearing in the hearing-impaired listeners, only one masker level was tested. The procedure used to measure MPPs in the hearing-impaired listeners was the same as that used in experiment 1.

V. RESULTS AND DISCUSSION

MPPs measured in the four listeners with cochlear hearing loss are plotted in Fig. 5. The left column shows MPPs obtained with an on-frequency masker and the right column shows MPPs obtained with an off-frequency masker. Data for individual subjects are presented in separate rows of panels.

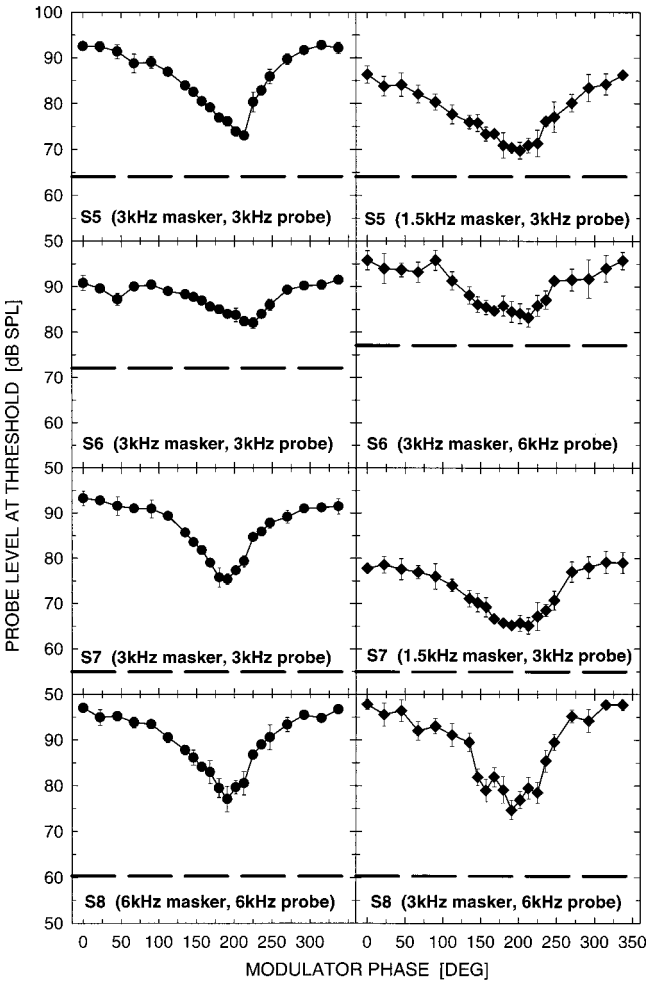


FIG. 5. Individual MPPs for hearing-impaired listeners. The left column of panels shows MPPs measured with an on-frequency masker, and the right column shows MPPs for an off-frequency masker. Masker and probe frequencies are shown in each panel. The dashed lines represent absolute threshold for the probe.

In general, the MPPs measured with the on-frequency masker (left column of panels) are similar in shape across hearing-impaired listeners. They exhibit V-shaped forms and they cover a much smaller range of probe levels compared to the normal-hearing data (see Fig. 2). The peak-valley differences in the MPPs of the hearing-impaired listeners were much smaller than the peak-valley differences observed at any masker level for the normal-hearing listeners. It might be argued that, in hearing-impaired listeners, the reduced dynamic range of the on-frequency MPP results from elevated absolute thresholds for the probe, which might limit performance in the valley. That does not appear to be the case, since the absolute thresholds, as shown by the dashed lines in each panel, were still well below the minimum of the MPP for each listener. For listener S6, different probe frequencies were used in the on- versus off-frequency masking conditions, thus the different thresholds for the probe in quiet in the right and left columns. As will be discussed later, the reduced peak-valley differences seen in the MPPs of hearing-impaired subjects reflect reduced or absent compression.

The off-frequency masking results in the right column of Fig. 5 differ substantially from those observed with the off-

frequency masker in normal-hearing listeners (see Fig. 3). In particular, they do not exhibit deep extended valleys. Instead, they have valleys that are more similar to those observed for the on-frequency masker (left column of panels in Fig. 5). This result would be expected assuming that, for normal-hearing listeners, the difference in the MPP shapes for on- versus off-frequency maskers is due to the different rates of growth of BM response to stimulation by those maskers at the probe-frequency place. Similar MPP shapes observed for the hearing-impaired listeners presumably reflect similar rates of growth of BM responses for on- and off-frequency maskers. It is worth noting that the MPPs measured for listeners S7 and S8 with the off-frequency masker exhibit slightly different patterns. They appear to have more bowl-shaped valleys, similar to those observed with the off-frequency masker in normal hearing, but they are much shallower. This result might suggest that there is some residual compression in those two listeners. This seems likely given that the absolute threshold for the probe, for listener S7 at 3 kHz and that for listener S8 at 6 kHz were lower than the thresholds observed at 3 kHz for listener S5, and at 6 kHz for listener S6 (even though the difference between absolute thresholds for S8 at 6 kHz and S5 at 3 kHz was only about 4 dB).

The differences observed between the MPPs for the two groups of listeners may partly result from the fact that the listeners with normal hearing were tested using a 6-kHz probe, whereas some listeners with impaired hearing were tested using a 3-kHz probe. For the listeners with normal hearing, a high probe frequency was used to prevent detection of spectral splatter. For the listeners with hearing loss, the effect of energy splatter on the data was probably quite small. It has been shown that auditory filters are often broader than normal in regions of cochlear hearing loss (e.g., Glasberg and Moore, 1986). Furthermore, because of the elevated absolute thresholds, most of the splattered energy was not detectable to the listeners tested with the 3-kHz probe. Another factor that could potentially lead to differences between MPP shapes observed for the 3- and 6-kHz probes is the dependence of the amount of compression on stimulus frequency. Although the amount of compression appears to be frequency dependent, it has been shown that for frequencies above 1 kHz, the changes in estimates of compression across frequency appear to be relatively small (Bacon *et al.*, 1999; Plack and Oxenham, 2000; Glasberg and Moore, 2000).[2]

The MPPs for the hearing-impaired listeners were measured using a masker level that was higher than the levels used to test subjects with normal hearing. Data of Oxenham and Plack (1997) and Gregan *et al.* (1998) provide evidence that at high levels of stimulation, a more linear growth of BM response is observed in normal hearing. Thus the higher level of testing could be responsible for some of the differences between the data obtained from listeners with normal and impaired hearing.

An interesting aspect observed in the data presented in Fig. 5 is that, for the listeners with the largest amount of hearing loss (S5, S6), there was a substantial shift of the minimum of the MPP away from 180 deg and toward larger modulation phases. For listener S5 the minima occurred at a phase of 213 and 202 deg, and for listener S6 the minima occurred at 225 and 213 deg, for the on- and off-frequency maskers, respectively. This shift of the minimum was much smaller in listeners S7 and S8, whose hearing loss at the probe frequency was a little milder and whose data, gathered with the off-frequency maskers, suggested that some compression might be preserved in their auditory systems.

VI. DESCRIPTION OF THE MODEL

The data shown above appear to support the idea that the difference in the shapes of MPPs observed for on- versus off-frequency maskers is due to a different BM-response growth at the probe-frequency place for the two masking conditions. Existing models that have been used to make predictions about the effect of combined forward and backward masking have not considered the case where the stimuli (the masker and the probe) were overlapping in time and had different frequencies. This case must be taken into consideration to make predictions about the shape of an off-frequency MPP. Therefore, the present study adapts an existing model to make it usable for predicting the MPPs measured in our experiments. The main frame of the model has been used in the past to account for different aspects of temporal processing (e.g., Viemeister, 1979; Buus and Florentine, 1985; Moore *et al.*, 1988; Plack and Moore, 1990; Oxenham and Moore, 1994). The models used in those studies involved some or all of the following stages of processing: peripheral filtering, half-wave rectification, compression, a temporal integrator, and a decision device.

Because highly distorted waveforms are not observed in basal regions of the BM in response to tonal stimulation (Ruggero *et al.*, 1997), in the model, compression was applied to the envelopes of the waveforms rather than to the waveforms themselves. Therefore, half-wave rectification was replaced by computation of the absolute value of the analytic signal. This operation led to extraction of the input signal's envelope. Each point of the envelope was then subjected to compression corresponding to the instantaneous level.

The model assumes that, when the carrier frequency f_m of the modulated masker and the probe frequency f_p are equal ($f_m = f_p$), the two stimuli are subjected to the same compression as described by compression exponent α. We will denote the envelope of waveform x by $\{x\}$. Following the finding of Ruggero (1992) and Ruggero *et al.* (1997), it can be assumed that the BM response R, to the simultaneous stimulation by the masker x_m and the probe x_p, is proportional to the compressed amplitude of the sum of the two stimuli:

$$R_{m+p} = a \cdot \{x_m + x_p\}^{\alpha}, \tag{1}$$

where a is a proportionality constant. The response to the masker alone can be represented by

$$R_m = a \cdot \{x_m\}^{\alpha}. \tag{2}$$

Because the probe used to measure the MPPs was very short, it was assumed that for each position of the probe during the

modulated masker, the temporal integrator $W(t)$ was centered at the temporal center of the probe t_c. The integrator was assumed to be operating on a squared representation of the input stimulus. This square-law nonlinearity has been applied in past studies that used a temporal window (e.g., Oxenham *et al.*, 1997; Plack and Oxenham, 1998) because it correctly predicts additivity of masking in the case of hearing loss (Oxenham and Moore, 1995).[3] The squared representation of the stimulus was then corrupted by internal noise before it was input to the temporal integrator. Limiting the integration time to the duration of the masker, the output of the temporal integrator for the masker and the probe combined can be described by

$$y_{m+p}(t_c) = a \cdot \int_0^{T_m} [\{x_m(\tau) + x_p(\tau)\}^{2\alpha} + N_{\mathrm{INT}}] \cdot W(\tau - t_c)\, d\tau. \tag{3}$$

The output of the temporal integrator for the masker alone would be

$$y_m(t_c) = a \cdot \int_0^{T_m} [\{x_m(\tau)\}^{2\alpha} + N_{\mathrm{INT}}] \cdot W(\tau - t_c)\, d\tau, \tag{4}$$

where T_m represents the total duration of the masker, and N_{INT} represents an additive internal noise, the purpose of which is to account for absolute threshold.

For the detection of the probe, the ratio of the output of the temporal window for the masker and probe presented together and the output of the temporal window for the masker alone has to reach some constant criterion denoted by *K*. Thus the masked threshold of the probe is obtained when

$$\frac{y_{m+p}(t_c)}{y_m(t_c)} = K. \tag{5}$$

The situation becomes more complicated when the frequency of the masker is lower than the frequency of the probe ($f_m < f_p$). In this case, the rate of response growth at the probe-frequency place is different for the masker than it is for the probe. Similar to the on-frequency masking condition, when the two stimuli overlap in time, the BM response is not simply the sum of the responses produced by the masker and the probe when each is presented alone. The present model takes into consideration the fact that in the case of simultaneous masking the BM response results from a nonlinearly combined effect of the masker and the probe. At the same time it assumes, based on physiological and psychophysical studies mentioned earlier, that the response to the masker at the probe-frequency place grows at a different rate than does the response to the probe. We assume that the BM response to the masker at the probe-frequency place (R_{mp}) grows at a rate described by a compression exponent denoted by β. In this case

$$R_{mp} = b \cdot \{x_m\}^{\beta}, \tag{6}$$

where *b* is a proportionality constant for the masker representation at the probe-frequency place. When the frequency of the masker is an octave lower than the frequency of the probe (as was the case in the present study), the masker produces a linear response at the probe-frequency place and the exponent β can be assumed equal to one. However, for greater generalization, we will keep the exponent β in the following equations. This exponent becomes smaller as the masker frequency approaches the probe frequency, finally reaching the value of the compression exponent applied to the probe when the two stimuli have the same frequency.

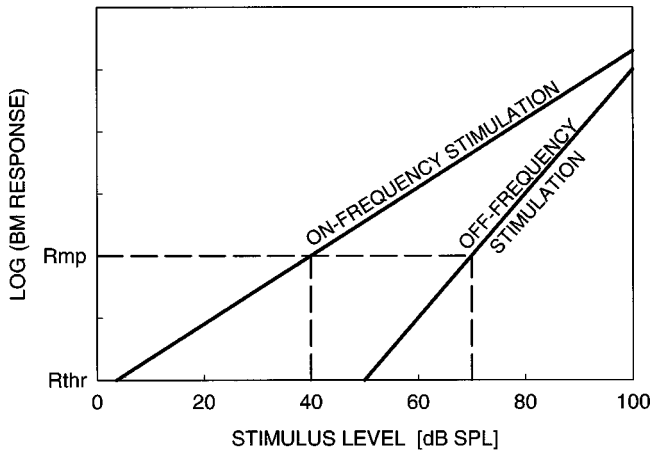


FIG. 6. Illustration of BM-response growth for on- and off-frequency stimulation, presented in a log–log coordinate system. The rate of response growth is slower for the on-frequency stimulus due to compressive BM nonlinearity, and based on Eq. (7) is described by α. The rate of response growth for the off-frequency stimulus is faster (close or equal to one), but the stimulus does not produce any response until its level reaches the threshold value. Based on Eq. (6) the slope of the steeper line is represented by β. Equation (8) is derived under the assumption that the two stimuli produce the same response R_{mp}.

The model further assumes that the same response (R_{mp}) could be produced by a signal x_s whose frequency (f_s) is equal to the frequency of the probe ($f_s = f_p$), as long as the amplitude of that signal is appropriately adjusted. This is schematically illustrated in Fig. 6. Such a signal would be compressed according to the same BM input–output function as the probe. Therefore, analogous to Eq. (1), the response to such a signal could be expressed as

$$R_{mp} = a \cdot \{x_s\}^{\alpha}. \tag{7}$$

For each magnitude of the response to the masker, we can find the amplitude of the on-frequency signal by expressing it in terms of the amplitude of the masker based on Eqs. (6) and (7):

$$\{x_s\} = \left(\frac{b}{a}\right)^{1/\alpha} \cdot \{x_m\}^{\beta/\alpha}. \tag{8}$$

If signal x_s is always chosen so that it produces the same response as the masker, it can be used to compute the combined response to the masker and the probe at the probe-frequency place. Such a signal can be thought of as the representation of the masker at the probe-frequency place. The signal can be added directly to the probe and it has to produce the same growth-of-response function as the probe. Similar to the on-frequency case, an intensity-like (squared) representation was input to the temporal integrator. Thus the output of the temporal integrator for the masker and the probe presented together is described by this formula, which is similar to Eq. (3):[4]

$$y_{m+p}(t_c) = a\int_0^{T_m} [\{C^{1/\alpha}\cdot x_m(\tau)^{\beta/\alpha} + x_p(\tau)\}^{2\alpha} + N_{\mathrm{INT}}] \cdot W(\tau - t_c)\,d\tau, \quad (9)$$

where $C = b/a$, and represents attenuation of the masker by the auditory filter centered at the probe frequency. The output of the temporal integrator for the masker alone at the probe-frequency place would be

$$y_m(t_c) = a\int_0^{T_m} [\{C^{1/\alpha} x_m(\tau)^{\beta/\alpha}\}^{2\alpha} + N_{\mathrm{INT}}] \cdot W(\tau - t_c)\,d\tau. \quad (10)$$

The masked threshold for the probe is obtained when the criterion K described by Eq. (5) is fulfilled. The constant a does not need to be known because it appears in the numerator and the denominator of the ratio describing the criterion in the on- and off-frequency masking case, and thus will cancel out. The approach applied in the present study to the case where the masker frequency is lower than the signal frequency is similar to the approach used by Goldstein (1990), which is depicted by the diagram in Fig. 1 of his paper.

Several types of temporal integrator have been proposed in the past, but the most frequently used have been an exponential temporal window and a ROEX (rounded exponential) temporal window (Moore *et al.*, 1988; Plack and Moore, 1990; Oxenham and Moore, 1994; Moore *et al.*, 1996; Plack and Oxenham, 1998). For the model described here, a ROEX-shaped window was chosen as the temporal integrator. Following Moore *et al.* (1988), it was assumed that each side of the temporal window is described by two exponential functions added together: one representing the central part of the window and the other representing the skirt of the window. Thus each side of the temporal integrator was described by

$$W(t) = (1-w)\cdot\left(1+\frac{2t}{T_p}\right)\cdot\exp\left(-\frac{2t}{T_p}\right) + w\cdot\left(1+\frac{2t}{T_s}\right)\cdot\exp\left(-\frac{2t}{T_s}\right),$$

where time t is measured relative to the center of the window, w is a weighting function describing the relative contributions of the steeper (central part) and shallower (the skirt) exponential functions, T_p is the time constant determining the sharpness of the central part of the window, and T_s is the time constant determining the sharpness of the skirt. It was also assumed that the temporal window is asymmetric with the longer time constants describing the side preceding the center of the temporal window and the shorter time constants describing the side following the center. The ratio of the time constants describing both sides of the center of the window was assumed to be the same as the ratio of the time constants describing the skirts. This assumption reduced the number of the parameters of the ROEX-shaped integrator to four.

The compression exponent α was computed based on the gain function described by a formula proposed by Glasberg and Moore (2000). The function is based on the assumption that in a normal auditory system, strong gain is applied by the cochlear amplifier to very faint stimuli, and that the gain decreases as stimulus level increases. The shape of the gain function depends on the maximum gain, which is applied to the softest sounds. The maximum gain, in turn, depends on the functional status of the outer hair cells. When there is some damage to outer hair cells, the maximum gain decreases so that faint stimuli are amplified less and therefore are rendered inaudible. Consequently, absolute threshold increases and the BM response becomes more linear. The function proposed by Glasberg and Moore is given by the following formula

$$\mathrm{Gain} = -0.1\cdot L + A + B\cdot(1 - 1/(1+\exp(0.05\cdot(50-L)))), \quad (11)$$

where L is the input level in dB SPL. A and B are defined as

$$A = -0.0894\cdot G_{\max} + 10.894, \quad B = 1.1789\cdot G_{\max} - 11.789,$$

where $G_{\max}$ represents the maximum gain.

The left panel of Fig. 7 shows a set of the gain functions plotted against the input level for different values of the maximum gain. The intercept with the ordinate for each curve represents the maximum gain. The right panel of Fig. 7 shows the BM input–output functions based on the gain functions computed for the corresponding maximum gain. The functions resemble the BM input–output functions measured by Robles *et al.* (1986), Ruggero (1992), and Ruggero *et al.* (1997). They clearly show that as the maximum gain decreases, the threshold is shifted toward higher levels and the BM input–output function becomes more linear. The figure also shows that at high levels, the outputs corresponding to the larger maximum gains and the output corresponding to the maximum gain of zero converge (an effect observed in psychophysical study as loudness recruitment).

Compression exponents α were computed based on the gain function using the following formula

$$\alpha = (L + \mathrm{Gain} - G_{\max})/L. \quad (12)$$

The resulting compression exponent is plotted in Fig. 8 as a function of the input level for different values of the maximum gain.

It has to be noted that the compression exponent observed with the largest maximum gain of 60 dB does not reach a value lower than 0.34, even though the shallowest slope of the BM-response growth was only 0.17 for the 60-dB maximum gain.[5] As the maximum gain decreases, the compression exponent increases in the mid-to high-level region reaching a value of one (linear response) for the maximum gain of zero.

To obtain predictions for the MPPs, the instantaneous level L corresponding to each point of the stimulus envelope was computed. Based on the instantaneous level, the exponent α was obtained for each point of the envelope using Eq. (12). This quasi-instantaneous compression (point-by-point envelope compression) was equivalent to the level-dependent nearly instantaneous gain and thus it did not distort the compressed waveforms. For the off-frequency masker, the exponent β was set to be equal to one assuming

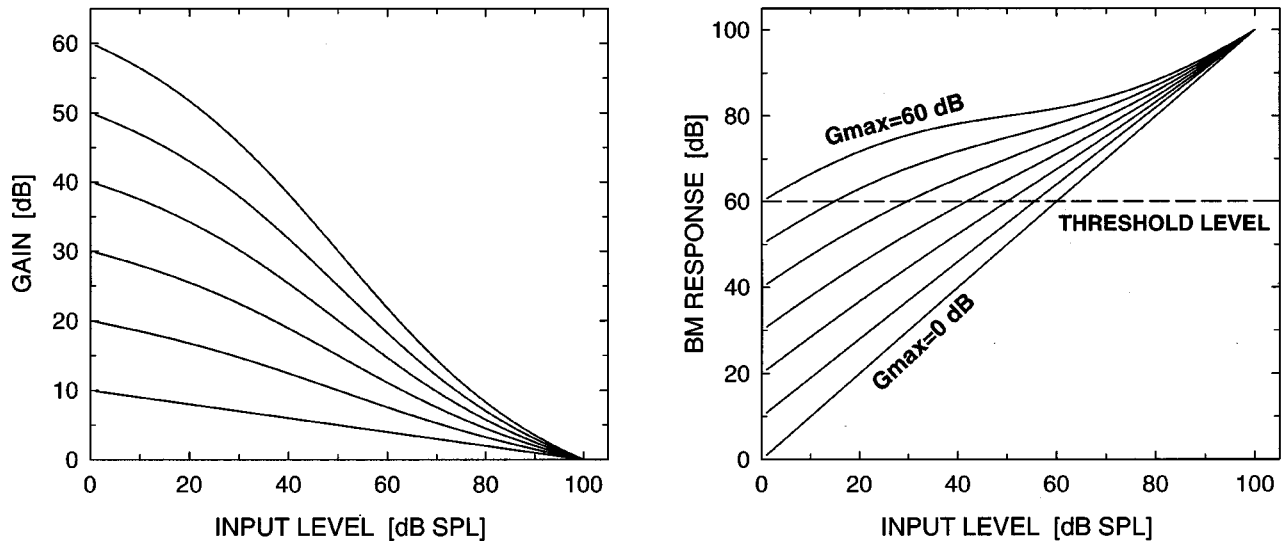


FIG. 7. Cochlear gain in dB (left panel) and BM response (right panel) plotted as a function of stimulus level. Different curves were obtained for different values of maximum gain using a formula proposed by Glasberg and Moore (2000). In the left panel, the maximum gain corresponds to the intercept of each curve with the *y*-axis. In the right panel, the maximum gain of 60 dB corresponds to the most compressive function and the maximum gain of 0 dB corresponds to the straight line with a slope of one. The dashed line in the right panel shows the level of the BM response that produces threshold sensation (detection threshold).

a linear response to the masker. This assumption is reasonable, since the frequency of the masker was one octave lower than the frequency of the probe.

VII. PREDICTIONS BY THE MODEL

Predictions for the measured MPPs were obtained using the model described above. Each point for the model predictions was obtained by iteratively adjusting the amplitude of the probe, x_p, until the absolute value of the difference between a fixed criterion (estimated in pilot computations) and the value of K computed from Eq. (5) reached the minimum. The parameter C in Eq. (10) was originally estimated to be equal to 0.0022 based on the rate at which attenuation by the ROEX-shaped auditory filter changes with masker level (Moore and Glasberg, 1987a). However, this parameter had to be adjusted slightly for each individual listener to produce a good fit. The adjustments were justifiable since the auditory filters derived by Moore and Glasberg are based on data averaged across the normal-hearing listeners used in their study. In studies measuring auditory filters, the inter-subject variability has often been reported to be quite high. That was likely to be the reason why the adjustments of the parameter representing attenuation of the masker by the filter tuned to the probe frequency were necessary in our computations for individual listeners. In the present model, values of C between 0.002 and 0.006 were used, depending on the listener.

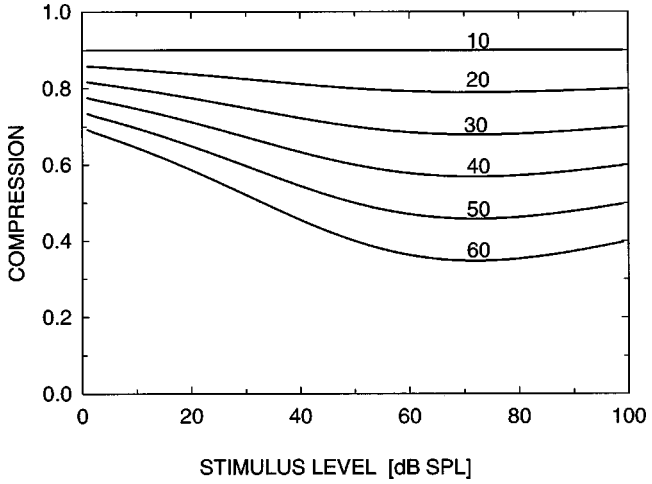


FIG. 8. Compression exponents plotted as a function of stimulus level. The functions were computed based on Eq. (12) using different values of maximum gain (shown on top of the respective curves).

For each listener, only one MPP (for the highest level of the on-frequency masker) was simulated using four free parameters. The free parameters were the following parameters of the ROEX-shaped temporal window: weight w, the time constant describing the central part of the temporal window before the center (T_p), the time constant describing the skirt before the center (T_s), and a ratio of the time constant T_p (or T_s) to the corresponding time constant describing the side after the center of the temporal window (denoted by Asym in Table I). These four parameters were changed iteratively, during a computation of the MPP, until the squared deviations between the predicted and the measured thresholds, summed across the modulation cycle of the masker envelope, produced a minimal value. Once the parameters of the temporal window producing the best prediction for that selected case were established, they were used in computation of all the other MPPs for the listener.

The predicted MPPs observed with the on-frequency masker in listeners with normal hearing are shown by solid lines in Fig. 9. The symbols represent the data replotted from Fig. 2. Separate panels show the data and the predictions for individual listeners. In each panel, different symbols represent the data obtained for different masker levels.

Overall, there is good agreement between the data and the predictions. For the on-frequency masker the model pro-

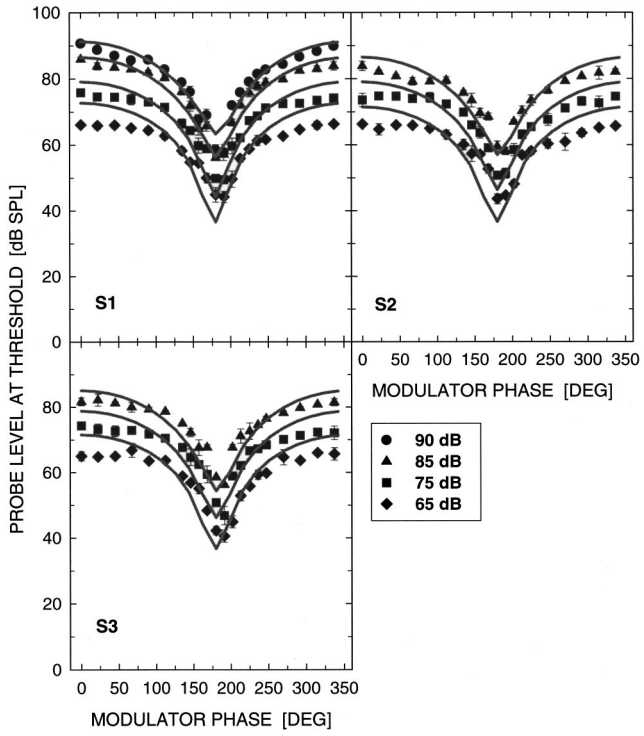


FIG. 9. On-frequency MPPs for normal-hearing listeners, replotted from Fig. 2 (symbols) along with predictions by the model (solid lines).

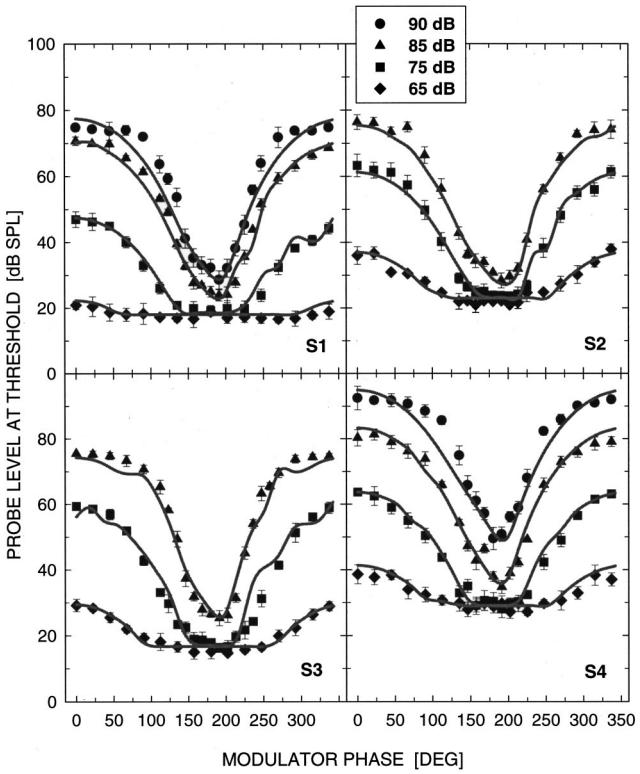


FIG. 10. Off-frequency MPPs for normal-hearing listeners, replotted from Fig. 3 (symbols) along with predictions by the model (solid lines).

duces MPPs with sharp narrow valleys. However, in many cases the model overpredicts the measured thresholds near the peaks of the modulated masker by about 2–8 dB. It also sometimes underpredicts the thresholds in the region of the valley.

One possible reason for the discrepancy between the predictions and the data may be related to the gain function used for the computation of compression exponents. The function was derived based on both physiological and psychophysical data and represents some average shape for a given maximum gain in the auditory system. In the model, the compression exponent could only vary according to the strictly defined function based on that ''average'' gain function. This might have contributed to the error in the predictions produced for each individual listener.

Figure 10 shows the predictions and the data observed for the off-frequency masker in listeners with normal hearing. The model correctly predicts the longer and deeper valleys observed in the data. The predictions for the off-frequency masking condition are more accurate than those produced for the on-frequency masker. The solid lines fall very close to the data for each condition and each listener.[6]

Because of a relatively small dynamic range of hearing at the probe frequency, the hearing-impaired listeners were tested using only one masker level for the on- and off-frequency masking conditions. For each listener, the MPP for the on-frequency masker was computed using the four parameters of the temporal window as free parameters of the simulation. The values of those parameters producing the best predictions ''on-frequency'' were then used to compute the MPP observed with the off-frequency masker. The compression exponent α was computed for each point of the on-frequency masker envelope and each point of the envelope of the probe and masker combined. It was done using Eq. (12) for the value of maximum gain selected based on pilot computations. For the off-frequency masker, the compression exponent β was set equal to one. Figure 11 shows the predictions (solid lines) and the data (symbols) for the hearing-impaired listeners. The left column presents the results for the on-frequency masker and the right column presents the results for the off-frequency masker. For all four listeners, the predictions fall reasonably close to the data.

It should be noted that for the hearing-impaired listeners, the model produces a minimum of the MPP that is shifted away from the 180-deg phase toward larger phases, in agreement with what is observed in the data. Within the model, this shift results from combined effects of the temporal-window asymmetry and reduced maximum gain. The shift is not produced by the model when the maximum gain is large or when both sides of the temporal window are described by the same time constants.

Table I contains the parameters resulting from the computations of the MPPs. The maximum gain producing the predictions for listeners with normal hearing presented in Figs. 9 and 10 had to be set to values larger than 40 dB. For two out of four listeners with hearing loss, the maximum gain of 0 dB produced the best predictions. This presumably suggests that there is no compressive nonlinearity in the auditory systems of those hearing-impaired listeners in the region corresponding to the probe frequency. For listeners S7 and S8, the maximum gains of 15 and 8 dB were used to produce the predictions. A maximum gain greater than 0 dB

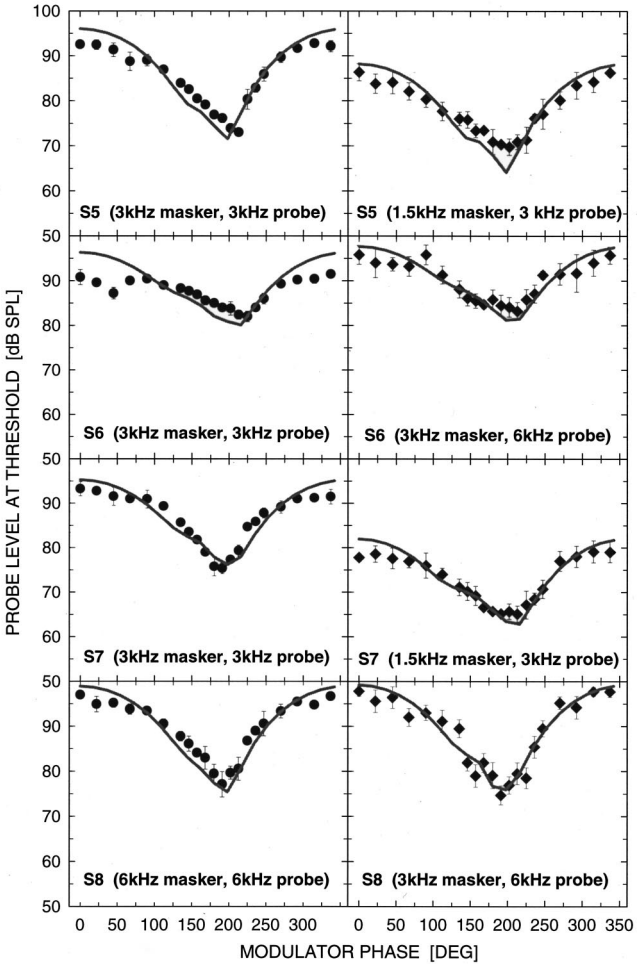


FIG. 11. MPPs measured in hearing-impaired listeners, replotted from Fig. 5 (symbols) along with predictions by the model (solid lines). The data and the predictions for the on- and off-frequency maskers are shown in the left and right column of panels, respectively.

suggests some residual compression in those listeners' auditory systems at the probe frequency.

The parameters of the temporal window resulting from the simulations were similar across all listeners. In all cases, the differences between values of any given parameter across the listeners (with normal and impaired hearing) were not greater than those differences seen across the individual normal-hearing listeners in studies by Moore *et al.* (1988), Plack and Moore (1990), and Oxenham and Moore (1994). Therefore, the results of the present simulations appear to indicate that the temporal integrator in the hearing-impaired subjects participating in our study does not differ from the temporal integrator operating in the listeners with normal hearing. Consequently, it can be concluded that the shallower MPPs observed in the hearing-impaired listeners with the on-frequency masker (shown in Fig. 5) result from reduced or absent compression (due to dysfunction of the active mechanism and possibly due to the higher level of masker presentation) rather than reduced temporal resolution.

TABLE I. The maximum gain and the four parameters of the ROEX-shaped temporal window that produced the predictions shown in Figs. 9–11.

Subject	G_{max} [dB]	W	Asym	T_p [ms]	T_s [ms]
S1	48	0.09	1.5	2.0	36
S2	47	0.09	1.7	2.0	38
S3	45	0.09	1.6	2.0	34
S4	42	0.07	1.5	2.0	30
S5	0	0.06	1.7	2.0	25
S6	0	0.09	1.9	1.2	34
S7	15	0.09	1.8	2.5	37
S8	8	0.09	1.7	2.0	27

VIII. GENERAL DISCUSSION AND CONCLUSIONS

The purpose of the present study was to account for the differences in the shapes of MPPs measured with a masker whose frequency is the same as the frequency of the probe and MPPs measured with a masker whose frequency is much lower than the frequency of the probe. In experiment 1, MPPs were measured in detail over one modulation cycle of the masker in listeners with normal hearing. It was found that when the frequency of the masker is well below the frequency of the probe, MPPs exhibit deep and extended valleys. In contrast, the valleys of MPPs measured with the on-frequency masker are sharp and narrow, and the dynamic range of those patterns is not as large as the dynamic range observed for off-frequency masking. The results are consistent with the often-reported release from masking within the upper accessory excitation observed with fluctuating maskers. It has been shown that masked thresholds, measured using a tonal probe and a masker much lower in frequency than the probe, are substantially reduced when the masker is a narrowband noise compared with the case where the masker was a pure tone of equal intensity (Buus, 1985; Mott and Feth, 1986; Moore and Glasberg, 1987b; Nelson and Schroder, 1996). To explain the difference in the amount of masking, it has been suggested that when an off-frequency masker with a fluctuating envelope is used, listeners can detect the probe by ''listening in the valleys'' of the masker envelope. This explanation seems plausible, since the valleys in the internal representation of an off-frequency masker at the probe-frequency place appear to be longer compared with those in the internal representation of the on-frequency masker, due to nonlinear BM behavior. Results of the present study indicate that listeners with normal hearing can take advantage of those longer and deeper valleys.

In listeners with hearing loss of cochlear origin, the discussed differences in the shape of MPPs are much smaller or are not observed, depending on the degree of hearing loss at the test frequency. The MPPs are very shallow and tend to have V-shaped valleys when measured with both on- and off-frequency maskers. This was shown in experiment 2. The results are consistent with the data of Nelson and Schroder (1996), who demonstrated that the release from masking within the upper accessory excitation occurring in normal hearing for temporally fluctuating maskers is markedly reduced and sometimes absent in the regions of high-frequency hearing loss. Very shallow MPPs indicate more difficulty in detecting the probe.

A model consisting of peripheral filtering, envelope extraction, compression, a temporal window operating on an

intensity-like representation, and a decision device was used to demonstrate that the data can be accounted for by assuming that an off-frequency masker produces a linear response at the probe-frequency place while the probe produces a compressed response. For the on-frequency masking case, the masker and the probe were compressed according to the same compression function. In the computation of the response to the masker and the probe, the model assumed that the internal representations of the two stimuli are combined in a nonlinear manner, i.e., the combined response is not a simple sum of the responses produced by the masker and the probe in their separate presentations. The model correctly predicted the different shapes of the MPPs observed with on- and off-frequency maskers and produced predictions that fall reasonably close to the data.

Results of the simulations indicate that good predictions for hearing-impaired listeners are obtained when the maximum gain in Eq. (11) is set to a value equal or slightly higher than zero, depending on the amount of hearing loss at the probe frequency. As expected, for listeners with normal hearing, much higher values of maximum gain had to be used, as shown in Table I. The four parameters of the temporal window were similar across all listeners used in the study. The obtained time constants describing the central part of the temporal window before its center (T_p) were slightly shorter than those reported by Moore *et al.* (1988), Plack and Moore (1990), and Oxenham and Moore (1994). The time constants describing the skirt (T_s) were comparable to those reported in their study. Computations of equivalent rectangular durations (ERD) produced values between 6.9 and 7.2 ms for the normal-hearing listeners and values between 5.5 and 7.7 ms for the hearing-impaired listeners. Those values are similar to the ERDs reported by Plack and Moore (1990) for high stimulus frequencies. The important finding seems to be that the time constants did not differ across listeners in our study, whether their hearing was normal or impaired. This suggests that the shallower MPPs observed in the hearing-impaired listeners do not result from reduced temporal resolution. Instead, reduced or absent compression appears to be responsible for the impaired processing of temporal envelopes. Oxenham and Moore (1997) suggested, based on their forward-masking data, that the temporal integrator in listeners with impaired hearing may not differ from that in normal hearing. The difference occurs in the processing at more peripheral stages. The processing is linear (or nearly linear) in impaired hearing and it is compressive in normal hearing. Results of the present simulations provide support for their findings.

The difference between patterns of masking produced by off-frequency maskers with fluctuating envelopes in listeners with normal hearing versus those in listeners with impaired hearing may have consequences for the recognition of target speech in the presence of a competing talker. The results presented in this work appear to imply that when speech is presented at a clearly audible level, to listeners with normal hearing and to listeners with cochlear hearing loss, normal-hearing subjects should be much better at detecting and recognizing short high-frequency consonants of the target talker immediately following or preceding lower-frequency vowels uttered by a competing talker. That is because the masker (the low-frequency vowel) will produce a linear response at the signal-frequency place (corresponding to the higher-frequency content of a consonant) in both groups of listeners, but only listeners with normal hearing can take advantage of deeper valleys in the masking pattern when the vowel is fading off. Hearing-impaired listeners would be expected to perform much worse under such listening conditions. Furthermore, our results argue that the worse performance of hearing-impaired listeners should be observed despite apparently normal temporal processing in their auditory system.

In the present paper, the differences in the shapes of MPPs measured with on- and off-frequency maskers and also the differences observed between MPPs measured in listeners with normal and impaired hearing were all attributed to BM compressive nonlinearity. However, another property of nonlinear cochlear mechanics might have played a role. Delgutte (1990a, b) demonstrated that for maskers well below the signal frequency, two masking mechanisms are operating. One of them is related to spread of excitation produced by a masker, and the other is based on suppression of the activity evoked in response to the probe by a simultaneous lower-frequency masker. Sizeable suppression was also shown by psychophysical data of Oxenham and Plack (1998). In our study, suppression could have affected the shapes of MPPs measured with the off-frequency masker in listeners with normal hearing. It is possible that it affected thresholds in the regions of each peak of the modulated masker where simultaneous masking prevails. In the valley, where the masked thresholds are governed by forward masking by the preceding peak, suppression should play little or no role (Arthur *et al.*, 1971). Suppression presumably did not affect the data from the hearing-impaired listeners. It has been demonstrated that suppression is reduced or eliminated as a result of damage to outer hair cells (Schmiedt *et al.*, 1980; Dallos *et al.*, 1980; Hicks and Bacon, 1999b). While the presented model provides satisfactory predictions without accounting for the effects of suppression, a more realistic model of the auditory periphery needs to take this nonlinear process into consideration.

The main conclusions from the present study may be summarized as follows:

(1) MPPs measured with on-frequency maskers in listeners with normal hearing exhibit sharp narrow valleys. For maskers much lower in frequency than the probe, MPPs have deeper and longer valleys. This difference in the shapes of MPPs observed for on- versus off-frequency maskers most probably results from different rates of masking growth in the two masking conditions. Different rates of masking growth are, in turn, the consequence of different BM-response growth rates produced by on- and off-frequency maskers at the probe-frequency place.
(2) The on- and off-frequency MPPs were more similar for hearing-impaired listeners than they were for listeners with normal hearing. This reflects similar rates of BM-response growth produced by the two types of masker in an impaired auditory system.
(3) The shapes of MPPs in listeners with normal hearing can be predicted by a model assuming compressive transfor-

mation of the probe and the on-frequency masker, and a linear BM response to the off-frequency masker at the probe-frequency place. The shapes of MPPs in listeners with impaired hearing can be predicted using the same model under the assumption that the response to the probe and the on-frequency masker grows at the same or a similar rate as the response to the off-frequency masker, which is assumed to grow linearly.

(4) Best predictions by a model assuming linear or nearly linear BM behavior in the case of hearing loss, and nonlinear BM behavior in the case of normal hearing, are obtained when similar parameters of the temporal integrator are used for both normal-hearing and hearing-impaired listeners. This suggests that the apparent poorer temporal processing of envelopes observed in impaired hearing results from reduced or absent compression rather than from a reduced efficacy of the temporal processing mechanism. This result supports a similar observation made by Oxenham and Moore (1997) based on their forward-masking data.

ACKNOWLEDGMENTS

This work was supported by Grant No. R01-DC00149 from NIDCD. It was also supported in part by the Lion's 5M International Hearing Foundation. We also thank Sid Bacon and the reviewers, Andrew Oxenham and Van Summers, for their helpful and constructive comments.

[1]Plack and Moore (1990) used different frequencies and different levels of the stimulus to estimate the parameters of their rounded-exponential-shaped temporal window. Their results suggested that the sharpness of the temporal window increases most when the stimulus frequency changes from 300 Hz to 900 Hz, but only a slight further increase is observed when the frequency increases up to 8100 Hz. It has to be noted that Plack and Moore did not apply compression to the stimulus intensity in their derivation of the parameters of the temporal window. That might have affected their estimates. As temporal integration is believed to be a central process, it seems reasonable to assume that the shape of the integrator does not depend on stimulus frequency or level. However, as the output of the temporal window would be affected by the combined effect of compression and temporal integration, the ''apparent'' temporal resolution might be better at higher frequencies and at moderate levels (where BM response is most compressive).

[2]Data collected using some of the normal-hearing listeners participating in the present study, for masker frequencies of 500 Hz and 1200 Hz paired with a probe frequency of 1200 Hz and for masker frequencies of 1200 Hz and 2400 Hz paired with a probe frequency of 2400 Hz, showed a similar difference in the shapes of MPPs measured with an on- and off-frequency masker as the difference observed in this study for the normal-hearing listeners tested at higher frequencies. Those data are not presented here because they were affected by splatter. However, they provide additional evidence that the similar MPPs observed with on- and off-frequency maskers for the hearing-impaired listeners, tested with a probe frequency of 3 kHz do not result from the effect of frequency on the amount of compression.

[3]The temporal window must be integrating information encoded as some aspect of neural activity. Yates *et al.* (1990) found that, once BM compression is factored out, the auditory-nerve response appears to grow in proportion to stimulus intensity rather than stimulus amplitude. This finding is in agreement with additivity-of-masking data (Oxenham and Moore, 1995). In listeners with cochlear hearing loss, two equally intense maskers produce a 3-dB rather than 6-dB increase in threshold relative to the threshold observed when only one of the maskers is present. Assuming that the BM response is linear (or nearly linear) in the case of cochlear damage, this result suggests linear additivity of the auditory response with respect to intensity rather than with respect to amplitude. Psychophysical data of Oxenham and Plack (1997) suggest that, in normal hearing, the nonlinear BM response is preserved at higher stages of the auditory processing. Based on their results it is assumed in our model that the squared BM response is essentially unaltered by further neural processing. Therefore, only a stochastic process (internal noise) is added to the squared BM output before the stimulus representation is input to the temporal integrator. This assumption allows greater simplification of the model; however, one needs to be aware that it may not be entirely correct.

[4]In Eq. (9), the internal representation of the masker at the probe-frequency place was computed by raising the masker envelope (the modulator) to the power of β/α and then multiplying it by the carrier of the masker. Subsequently, such a ''transformed'' masker multiplied by the constant $C^{1/\alpha}$ was added to the probe, and the envelope of the obtained waveform was extracted. Extracting the envelope of the masker representation and adding it to the envelope of the probe would not be equivalent to adding the stimuli first and extracting the envelope of the resulting stimulus, and thus it would lead to a different (incorrect) result. Physical stimuli add before they reach the ear and undergo the processing at the basilar membrane and at higher stages of the auditory system, which involves envelope extraction.

[5]In a healthy auditory system, gain applied by the cochlear amplifier to an input stimulus decreases progressively with increasing stimulus level. The rate at which gain changes is not constant across levels. The greatest decrease is observed at medium levels leading to the shallowest slope of the growth-of-response function in this range (see Fig. 7). As a result of this level-dependent gain, a compressive function is observed at the output of the cochlear amplifier. To compute the output level for each given input level, the input level needs to be multiplied by the compression exponent corresponding to that level. This compression exponent is *not* equivalent to the rate at which the response grows at the output of the cochlear amplifier. Because gain decreases at a slower rate for low input levels than it does for medium levels, the compression exponents for levels corresponding to the shallowest part of the output function have to be *larger* than the slope for this range of levels. The slope of the growth-of-response function would be equivalent to the compression exponent only if gain decreased at a constant rate across levels, which is not the case for a healthy cochlea. Therefore, it is not justifiable to expect that models applying compression, such as the ones used by Oxenham and Moore (1994) and in the present work, should produce compression exponents equal to the slope of the BM input-output function.

[6]The wiggles in the predicted MPPs are an artifact of the iteration procedure. The model itself does not produce such nonmonotonicities since the BM input–output function is monotonic and the temporal window is linear.

ANSI (**1996**). ANSI S3.6-1996, ''Specifications for audiometers'' (ANSI, New York).

Arthur, R. M., Pfeiffer, R. R., and Suga, N. (**1971**). ''Properties of 'two-tone inhibition' in primary auditory neurons,'' J. Physiol. (London) **212**, 593–609.

Bacon, S. P., Boden, L. N., Lee, J., and Repovsch, J. L. (**1999**). ''Growth of simultaneous masking for $f_m < f_s$: Effects of overall frequency and level,'' J. Acoust. Soc. Am. **106**, 341–350.

Bacon, S. P., Lee, J., Peterson, D. N., and Rainey, D. (**1997**). ''Masking by modulated and unmodulated noise: Effects of bandwidth, modulation rate, signal frequency, and masker level,'' J. Acoust. Soc. Am. **101**, 1600–1610.

Buus, S. (**1985**). ''Masking release caused by envelope fluctuations,'' J. Acoust. Soc. Am. **78**, 1958–1965.

Buus, S., and Florentine, M. (**1985**). ''Gap detection in normal and impaired listeners: The effect of level and frequency,'' in *Time Resolution in Auditory Systems*, edited by A. Michelsen (Springer-Verlag, New York), pp. 159–179.

Buus, S. and Florentine, M. (**1989**). ''Simulated hearing loss as a baseline for the assessment of auditory function in cochlearly impaired listeners,'' in *Proceedings of the Fifteenth Annual Bioengineering Conference*, edited by S. Buus (IEEE, Piscataway, NJ).

Dallos, P., Harris, D. M., Relkin, E., and Cheatham, M. A. (**1980**). ''Two-tone suppression and intermodulation distortion in the cochlea: Effect of outer hair cell lesions,'' in *Psychophysical, and Behavioral Studies in Hearing*, edited by G. van den Brink and F. A. Bilsen (Delft U. P., The Netherlands), pp. 242–249.

Delgutte, B. (**1990a**). ''Physiological mechanisms of psychophysical masking: Observations from auditory nerve fibers,'' J. Acoust. Soc. Am. **87**, 791–809.

Delgutte, B. (**1990b**). ''Two-tone rate suppression in auditory-nerve fibers: Dependence on suppression frequency and level,'' Hear. Res. **49**, 225–246.
Egan, J. P., and Hake, H. W. (**1950**). ''On the masking pattern of a simple auditory stimulus,'' J. Acoust. Soc. Am. **22**, 622–630.
Glasberg, B. R., and Moore, B. C. J. (**1986**). ''Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments,'' J. Acoust. Soc. Am. **79**, 1020–1033.
Glasberg, B. R., and Moore, B. C. J. (**2000**). ''Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise,'' J. Acoust. Soc. Am. **108**, 2318–2328.
Goldstein, J. L. (**1990**). ''Modeling rapid waveform compression on the basilar membrane as multiple-bandpass-nonlinearity filtering,'' Hear. Res. **49**, 39–60.
Gregan, M. J., Bacon, S. P., and Lee, J. (**1998**). ''Masking by sinusoidally amplitude-modulated tonal maskers,'' J. Acoust. Soc. Am. **103**, 1012–1021.
Hawkins, J. E., and Stevens, S. S. (**1950**). ''The masking of pure tones and of speech by white noise,'' J. Acoust. Soc. Am. **22**, 6–13.
Hicks, M. L., and Bacon, S. P. (**1999a**). ''Psychophysical measures of auditory nonlinearities as a function of frequency in individuals with normal hearing,'' J. Acoust. Soc. Am. **105**, 326–338.
Hicks, M. L., and Bacon, S. P. (**1999b**). ''Effects of aspirin on psychophysical measures of frequency selectivity, two-tone suppression, and growth of masking,'' J. Acoust. Soc. Am. **106**, 1436–1451.
Javel, E., McGee, J., Walsh, E., Farley, G., and Gorga, M. (**1983**). ''Suppression of auditory nerve responses. II. Suppression threshold and growth, iso-suppression contours,'' J. Acoust. Soc. Am. **74**, 801–813.
Jesteadt, W., Bacon, S. P., and Lehman, J. R. (**1982**). ''Forward masking as a function of frequency, masker level, and signal delay,'' J. Acoust. Soc. Am. **71**, 950–962.
Kemp, D. T. (**1978**). ''Stimulated acoustic emissions from within the human auditory system,'' J. Acoust. Soc. Am. **64**, 1386–1391.
Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Lutfi, R. A. (**1983**). ''Additivity of simultaneous masking,'' J. Acoust. Soc. Am. **73**, 262–267.
Moore, B. C. J., and Glasberg, B. R. (**1983**). ''Growth of forward masking for sinusoidal and noise maskers as a function of signal delay: Implications for suppression in noise,'' J. Acoust. Soc. Am. **73**, 1249–1259.
Moore, B. C. J., and Glasberg, B. R. (**1987a**). ''Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns,'' Hear. Res. **28**, 209–225.
Moore, B. C. J., and Glasberg, B. R. (**1987b**). ''Factors affecting thresholds for sinusoidal signals in narrow-band maskers with fluctuating envelopes,'' J. Acoust. Soc. Am. **82**, 69–79.
Moore, B. C. J., Glasberg, B. R., Plack, C. J., and Biswas, A. K. (**1988**). ''The shape of the ear's temporal window,'' J. Acoust. Soc. Am. **83**, 1102–1116.
Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1996**). ''Detection of decrements and increments in sinusoids at high overall levels,'' J. Acoust. Soc. Am. **99**, 3669–3677.
Moore, B. C. J., Vickers, D. A., Plack, C. J., and Oxenham, A. J. (**1999**). ''Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism,'' J. Acoust. Soc. Am. **106**, 2761–2778.
Mott, J. B., and Feth, L. L. (**1986**). ''Effects of the temporal properties of a masker upon simultaneous-masking patterns,'' in *Auditory Frequency Selectivity*, edited by B. C. J. Moore and R. D. Patterson (Plenum, New York), pp. 381–386.
Murnane, O., and Turner, C. W. (**1991**). ''Growth of masking in sensorineural hearing loss,'' Audiology **30**, 275–285.
Nelson, D. A., and Bilger, R. C. (**1974**). ''Pure-tone octave masking in listeners with sensorineural hearing loss,'' J. Speech Hear. Res. **17**, 252–269.
Nelson, D. A., and Schroder, A. C. (**1996**). ''Release from upward spread of masking in regions of high-frequency hearing loss,'' J. Acoust. Soc. Am. **100**, 2266–2277.
Nelson, D. A., and Schroder, A. C. (**1997**). ''Linearized response growth inferred from growth-of-masking slopes in ears with cochlear hearing loss,'' J. Acoust. Soc. Am. **101**, 2186–2201.
Nelson, D. A., and Swain, A. C. (**1996**). ''Temporal resolution within the upper accessory excitation of a masker,'' Acoustica **82**, 328–334.
Oxenham, A. J., and Moore, B. C. J. (**1994**). ''Modeling of additivity of nonsimultaneous masking,'' Hear. Res. **80**, 105–118.
Oxenham, A. J., and Moore, B. C. J. (**1995**). ''Additivity of masking in normally hearing and hearing-impaired subjects,'' J. Acoust. Soc. Am. **98**, 1921–1934.
Oxenham, A. J., and Moore, B. C. J. (**1997**). ''Modeling the effects of peripheral nonlinearity in listeners with normal and impaired hearing,'' in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Hillsdale, NJ), pp. 273–288.
Oxenham, A. J., Moore, B. C. J., and Vickers, D. A. (**1997**). ''Short-term temporal integration: Evidence of the influence of peripheral compression,'' J. Acoust. Soc. Am. **101**, 3676–3687.
Oxenham, A. J., and Plack, C. J. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.
Oxenham, A. J., and Plack, C. J. (**1998**). ''Suppression and the upward spread of masking,'' J. Acoust. Soc. Am. **104**, 3500–3510.
Penner, M. J. (**1980**). ''The coding of intensity and the interaction of forward and backward masking,'' J. Acoust. Soc. Am. **67**, 608–6616.
Penner, M. J., and Shiffrin, R. M. (**1980**). ''Nonlinearities in the coding of intensity within the context of a temporal summation model,'' J. Acoust. Soc. Am. **67**, 617–627.
Plack, C. J., and Moore, B. C. J. (**1990**). ''Temporal window shape as a function of frequency and level,'' J. Acoust. Soc. Am. **87**, 2178–2187.
Plack, C. J., and Oxenham, A. J. (**1998**). ''Basilar-membrane nonlinearity and the growth of forward masking,'' J. Acoust. Soc. Am. **103**, 1598–1608.
Plack, C. J., and Oxenham, A. J. (**2000**). ''Basilar-membrane nonlinearity estimated by pulsation threshold,'' J. Acoust. Soc. Am. **107**, 501–507.
Rhode, W. S. (**1971**). ''Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique,'' J. Acoust. Soc. Am. **49**, 1218–1231.
Robinson, C. E., and Pollack, I. (**1973**). ''Interaction between forward and backward masking: a measure of the integrating period of the auditory system,'' J. Acoust. Soc. Am. **53**, 1313–1316.
Robles, L., Ruggero, M. A., and Rich, N. C. (**1986**). ''Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input–output functions, tuning curves, and phase responses,'' J. Acoust. Soc. Am. **80**, 1364–1374.
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. (**1990**). ''Systemic injection of furosemide alters the mechanical response to sound of the basilar membrane,'' in *The Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, and C. R. Steele (Springer Verlag, Berlin), pp. 314–321.
Ruggero, M. A., and Rich, N. C. (**1991**). ''Furosemide alters organ of Corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane,'' J. Neurosci. **11**, 1057–1067.
Ruggero, M. A., Rich, N. C., and Recio, A. (**1993**). ''Alteration of basilar membrane responses to sound by acoustic overstimulation,'' in *Biophysics of Hair Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. v. Dijk, and S. M. v. Netten (World Scientific, Singapore), pp. 258–265.
Ruggero, M. A., Rich, N. C., Robles, L., and Recio, A. (**1995**). ''The effects of acoustic overstimulation, other cochlear injury and death on basilar membrane responses to sound,'' in *Effects of Noise on Hearing: Vth International Symposium*, edited by R. J. Salvi, A. Axelsson, D. Henderson, and R. Hamernik (Thieme Medical, Stockholm).
Ruggero, M. A., Rich, N. C., Robles, L., and Recio, A. (**1996**). ''The effects of acoustic trauma, other cochlear injury and death on basilar-membrane responses to sound,'' in *Proceedings of the Vth International Symposium on the Effects of Noise on Hearing*, edited by A. Axelsson, H. Borchgrevink, D. Henderson, R. P. Hamernik, and R. Salvi (Thieme Medical, Stuttgart), pp. 258–264.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.
Schmiedt, R. A., Zwislocki, J. J., and Hamernik, R. P. (**1980**). ''Effects of hair cell lesions on responses of cochlear nerve fibers. I. Lesions, tuning curves, two-tone inhibition, and responses to trapezoidal-wave patterns,'' J. Neurophysiol. **43**, 1367–1389.
Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (**1982**). ''Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique,'' J. Acoust. Soc. Am. **72**, 131–141.
Smits, J. T. S., and Duifhuis, H. (**1982**). ''Masking and partial masking in

listeners with a high frequency hearing loss,'' Audiology **21**, 310–324.
Stelmachowicz, P., Lewis, D. E., Larson, L., and Jesteadt, W. (**1987**). ''Growth of masking as a measure of response growth in hearing-impaired listeners,'' J. Acoust. Soc. Am. **81**, 1881–1887.
Viemeister, N. F. (**1972**). ''Intensity discrimination of pulsed sinusoids: The effects of filtered noise,'' J. Acoust. Soc. Am. **51**, 1256–1269.
Viemeister, N. F. (**1979**). ''Temporal modulation transfer functions based upon modulation thresholds,'' J. Acoust. Soc. Am. **66**, 1364–1380.
Yates, G. K. (**1990**). ''Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions,'' Hear. Res. **50**, 145–162.
Yates, G. K. (**1995**). ''Cochlear structure and function,'' in *Hearing*, edited by B. C. J. Moore (Academic, San Diego, CA), pp. 41–73.
Yates, G. K., Winter, I. M., and Robertson, D. (**1990**). ''Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range,'' Hear. Res. **45**, 203–220.
Zwicker, E. (**1976a**). ''Psychoacoustic equivalent of period histograms,'' J. Acoust. Soc. Am. **59**, 166–175.
Zwicker, E. (**1976b**). ''Masking-period patterns of harmonic complex tones,'' J. Acoust. Soc. Am. **60**, 429–439.
Zwicker, E. (**1976c**). ''Mithörschwellen-Periodenmuster amplitudenmodulierter Töne (Masking-period patterns of amplitude modulated pure tones),'' Acoustica **36**, 113–120.
Zwicker, E. (**1986**). ''Das Zeitauflösungsvermögen des Gehörs—Eine zweckmäßige Meßmethode im Hinblick auf die Sprachverständlichkeit,'' Audiol. Acoustics **25**, 170–184.
Zwicker, E., and Schorn, K. (**1982**). ''Temporal resolution in hard-of-hearing patients,'' Audiology **21**, 474–492.

Gap detection for similar and dissimilar gap markers

John H. Grose, Joseph W. Hall III, Emily Buss, and Debora Hatch
Division of Otolaryngology/Head & Neck Surgery, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7070

(Received 10 May 2000; revised 2 November 2000; accepted 10 January 2001)

Detection thresholds for temporal gaps between markers of dissimilar frequency are usually elevated with respect to thresholds for gaps between markers of similar frequency. Because gaps between markers of dissimilar frequency represent both a spectrally based perceptual discontinuity as well as a temporal discontinuity, it is not clear what factors underlie the threshold elevation. This study sought to examine the effects of perceptual dissimilarities on gap detection. The first experiment measured gap detection for configurations of narrow-band gap markers comprised of pure tones, frequency-modulated tones, and amplitude-modulated tones. The results showed that gap thresholds for frequency-disparate pure-tone markers were elevated with respect to isofrequency tonal markers, but that perceptual discontinuities between markers restricted to the same frequency region did not uniformly elevate threshold. The second experiment measured gap detection for configurations of markers where the leading and trailing markers could differ along the dimensions of bandwidth, duration, and pitch. The results showed that, in most cases, gap detection deteriorated when the bandwidth of the two markers differed, even when the spectral content of the narrower-band marker was completely subsumed by the spectral content of the wider-band marker. This finding suggests that gap detection is sensitive to spectral dissimilarity between markers in addition to spectral discontinuity. The effects of marker duration depended on the marker bandwidth. Pitch differences across spectrally similar markers had no effect. [DOI: 10.1121/1.1354983]

PACS numbers: 43.66.Mk [SPB]

I. INTRODUCTION

Detection and differential sensitivity to the presence of a silent interval that is marked by two tones, or other narrow-band stimuli, is most acute when the frequencies of the marker tones are similar. Sensitivity to the gap deteriorates as the frequency separation between the marker tones increases (Kinney, 1961; Perrott and Williams, 1971; Collyer, 1974; Fitzgibbons *et al.*, 1974; Divenyi and Danner, 1977; Divenyi and Hirsh, 1978; Neff *et al.*, 1982; Formby and Forrest, 1991; Formby *et al.*, 1996; Forrest and Formby, 1996; Grose and Hall, 1996; Phillips *et al.*, 1997; Formby *et al.*, 1998a, b). Phillips *et al.* (1997) suggested that isofrequency gap detection and across-frequency gap detection rely on different mechanisms: In the former case, the task is one of within-channel discontinuity detection; in the latter case, the task involves the shifting of auditory attention.

Analogous results have been observed in listeners who use multichannel cochlear implants. Hanekom and Shannon (1998), testing listeners with the Nucleus 22 device, found that gap detection was most acute when the two markers of the gap were carried by the same electrode channel (within-electrode channel). Performance deteriorated as the distance between the electrode channels carrying the first and second marker increased (across-electrode channel). For relatively small separations between electrode channels, Hanekom and Shannon attributed performance to the effects of peripheral neural interaction (i.e., performance was determined by within-channel mechanisms). For widely separated electrode channels, they suggested that a central mechanism was involved that was sensitive to the perceptual similarity between markers. Van Wieringen and Wouters (1999), testing listeners with the Laura device, also found that across-electrode-channel gap detection tended to be poorer than within-electrode-channel gap detection; however, this difference could be greatly reduced or eliminated if the listeners received sufficient training in the task. Moreover, performance in the across-electrode-channel conditions was minimally affected by the continuous activation of an intervening electrode channel between the two marker electrode channels, suggesting that neural interaction was not the basis of the effect. These results led van Wieringen and Wouters to contend that the across-electrode-channel task was not a measure of temporal resolution per se, but rather a measure of the perceptual dissimilarity between two markers. To support this contention, they measured gap detection for stimulus configurations that were not ''across-channel,'' but incorporated perceptually distinct stimuli as the two gap markers. The marker distinctions were provided by varying the number of active electrode channels in the pre- and post-gap markers and by varying the pulse rate across the two markers. Both of these manipulations tended to elevate gap thresholds in comparison to the baseline conditions of within-channel, perceptually similar markers. The latter finding of the effect of differential pulse rate corroborates the earlier work of Chatterjee *et al.* (1998), who also found that within-electrode-channel gap detection in Nucleus-22 users was most acute when the two markers employed the same pulse rate, and that thresholds were elevated when rate differences were introduced between the markers. Thresholds were also elevated when perceptual dissimilarities were in-

troduced between the markers by varying marker level while keeping pulse rate constant. Chatterjee *et al.* (1998) interpreted their results as reflecting a mechanism that is sensitive to perceptual discontinuities in sequential sounds.

It is appropriate at this juncture to consider more explicitly some of the terminology used here. In particular, the notion of perceptual dissimilarity, or discontinuity, requires definition. If acoustic differences are introduced between the two markers that result in a sound quality difference along some perceptual dimension, then these acoustic differences are said to underlie some perceptual dissimilarity. Consequently, the pairing of markers that are perceptually dissimilar results in a perceptual discontinuity across the temporal gap. The acoustic differences between markers may be, for example, in terms of their frequency content (spectral discontinuity), bandwidth, modulation pattern, or interaural level/phase. The resulting perceptual dissimilarities may be in terms of pitch, timbre, fluctuation strength, or spatial location. The degree to which markers are differentiated in perceptual space is not explicitly considered here.

The contention that poor across-electrode-channel gap detection in cochlear implant users is due to perceptual dissimilarities and not to limitations of temporal processing, per se, bears on the interpretation of the across-frequency gap detection work in the normal ear. Two findings in particular deserve attention. The first is from the work of Phillips *et al.* (1997) who found that the detection of a gap marked by two noise stimuli was poorer for a configuration where the first marker was a wide-band noise and the second marker was a narrow-band noise than for a configuration where both markers were isofrequency narrow-band noises. In the marker configuration that resulted in the elevated threshold, the spectral content of the first marker completely overlapped that of the second marker, but also contained additional spectral components. Thus the two markers were perceptually distinct but the task did not technically require an across-frequency analysis. A similar finding is evident in the work of Formby *et al.* (1998b) who measured gap detection for different combinations of tonal markers. As expected, gap detection was most acute when the gap was marked by isofrequency tonal markers. However, they included conditions where the second tonal marker was accompanied simultaneously by an additional tone at a different frequency. Thus the second marker consisted of a two-tone complex, where one of the tones was always the same frequency as that of the first marker and the other was at a different frequency. In these conditions, performance declined, even though the gap was always completely represented within a single frequency channel. A parsimonious interpretation of the Phillips *et al.* (1997) and Formby *et al.* (1998b) findings is that the decline in performance was due to the existence of perceptual differences between the two markers and not to frequency differences, per se. In contrast to this interpretation, Oxennam (2000) found that differential lateralization of gap markers had no effect on gap detection, even though the intra-cranial images of the markers were clearly distinct. In addition, pitch differences between markers with similar spectral envelopes was less disruptive to gap detection than spectral differences between markers with similar pitches. Oxenham interpreted these results as suggesting that peripheral encoding of the gap markers is the dominant factor underlying gap detection, and not higher level perceptual representations.

The purpose of this study was to elucidate further the effect of perceptual discontinuities on gap detection in normal-hearing listeners. A specific aim was to determine whether the diminished performance seen for across-frequency gap detection is a specific manifestation of a more general decline in temporal acuity brought about by the existence of perceptual dissimilarities between the two markers. If poor gap detection performance for frequency-disparate markers is due to perceptual dissimilarities between the markers, then poor gap detection performance should also be observed for other marker configurations that incorporate perceptual dissimilarities, even when both markers exist in the same frequency region. This specific aim was examined in an experiment that measured gap detection for configurations of narrow-band gap markers comprised of pure tones, frequency-modulated (FM) tones, and amplitude-modulated (AM) tones. A second specific aim of this study was to determine the importance of spectral constancy (i.e., equivalence of spectral content) versus spectral overlap for gap detection. This latter aim was examined in a second experiment that also included an assessment of marker duration.

II. EXPERIMENT 1. GAP DETECTION FOR PURE-TONE, FM-TONE, AND AM-TONE MARKERS

A. Method

1. Subjects

In the course of this study, a total of eight normal-hearing listeners ranging in age from 21 to 48 years participated. However, because the various conditions of the study were undertaken at widely different times, different subsets of the listeners participated in different phases of the study. To clarify which listeners participated in which conditions, they will be identified as Obs. 1–8. All listeners had audiometric thresholds $\leqslant$20 dB HL (ANSI, 1989) across the frequency range 250–8000 Hz. All were experienced in psychoacoustic listening experiments, including—for seven of the eight listeners—gap duration discrimination tasks, and all received a minimum of 1–2 h of training in the present task before data collection began.

2. Stimuli

The markers that bounded the gap were either pure tones, FM tones, or AM tones. The frequency of the pure tones, and the carrier frequency of the modulated tones, was 4000 Hz. The exception to this was the pure-tone marker used in the across-frequency condition (see below) which had a frequency of 2285 Hz. The rate of modulation for both FM tones and AM tones was either 10 Hz or 20 Hz. The modulation index for the FM tones was 10 which, for the 10-Hz modulation rate, translated to a 2.5% peak frequency deviation. The starting phase for the modulator was 0 rads which meant that the instantaneous frequency at the onset and offset of each FM marker was 4000 Hz. The modulation index for the AM tones was 1, and the amplitude of the

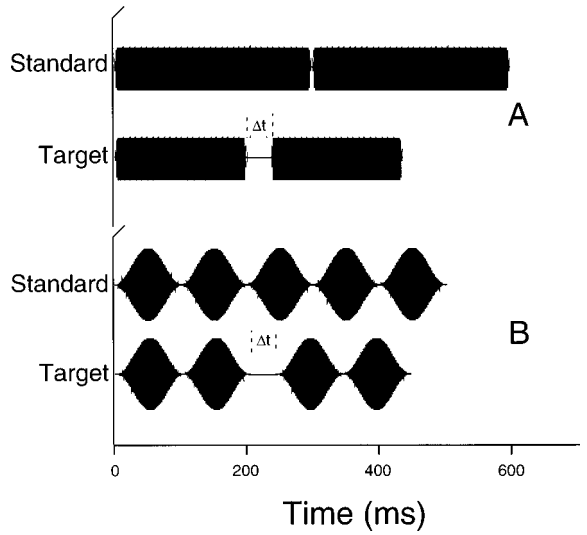


FIG. 1. Stimulus schematic showing a standard stimulus and a target stimulus for two possible combinations of marker type. Upper pair (A) shows stimulus envelope for pure-tone and/or FM-tone markers. Note that overall duration of the standard stimulus could exceed overall duration of target stimulus because marker duration could be either 200 ms or 300 ms, at random. Lower pair (B) shows stimulus envelope for 10-Hz AM-tone markers. Again note randomization of marker duration. Also note that, because of the π radian starting phase of the modulator, the ''gap'' in the standard configuration was more pronounced than in (A).

modulated waveform was adjusted to maintain level equivalence with the pure-tone markers (cf. Viemeister, 1979).[1] For the AM markers, the starting phase of the modulator was π rads which meant that the onset and offset of each marker occurred at a waveform minimum (see Fig. 1). The duration of all markers was randomly set on an interval-by-interval basis to be either 200 or 300 ms. For the modulated markers, this translated to either 2 or 3 periods of 10-Hz modulation. The randomization of duration was implemented to prevent overall stimulus duration (2 markers+gap) from being a viable detection cue. A 5-ms cosine-squared ramp was imposed on the beginning and end of each marker. The stimuli were digitally generated (TDT AP2) at a rate of 20 kHz and were low-pass filtered at 8 kHz (Kemo VBF 10M) prior to presentation. The stimuli were presented at a level of 65 dB SPL to the left ear through Sennheiser 580 headphones.

3. Procedure

Gap detection was measured using a 3-alternative, forced-choice (3AFC) procedure which incorporated a 3-down, 1-up adaptive strategy that converged on the 79.4% correct level. The interval between the 0-voltage point at the offset of marker 1 and the 0-voltage point at the onset of marker 2 was designated the gap, as shown in Fig. 1. In the no-gap standard, the two markers were therefore abutted next to each other; however, the interruption between the two markers was nevertheless detectable because of the 5-ms rise/fall ramps imposed on the two markers. Because a ''gap'' therefore existed even in the standard stimulus, it could be argued that the task should more correctly be termed a gap duration discrimination task. However, given the lack of a consensus regarding terminology for this paradigm, and given the fact that the impetus for this study derived largely from the gap detection literature, the term ''gap detection'' will be used in this report. Following three correct responses in a row, the duration of the gap was decreased by a factor of 1.2; following one incorrect response, the duration of the gap was increased by the same factor. A track was terminated after ten reversals in gap duration, and the geometric mean of the gap durations at the reversal points was taken as an estimation of threshold for that track. For each condition, at least three estimates of threshold were collected and the final threshold value was taken as the geometric mean of all estimates for that condition.

TABLE I. Marker configurations for the two condition sets of experiment 1.

		Marker 1	Marker 2
Condition Set 1	1	4000-Hz pure tone	4000-Hz pure tone
	2	4000-Hz pure tone	2285-Hz pure tone
	3	4000-Hz pure tone	10-Hz FM tone
	4	10-Hz FM tone	4000-Hz pure tone
	5	10-Hz FM tone	10-Hz FM tone
	6	10-Hz FM tone	20-Hz FM tone
Condition Set 2	1	10-Hz AM tone	10-Hz AM tone
	2	10-Hz AM tone	20-Hz AM tone
	3	20-Hz AM tone	20-Hz AM tone

Two condition sets were employed in this experiment. The marker pairs for each set of conditions are listed in Table I. Different combinations of listeners participated in the two sets. The first condition set incorporated the pure-tone markers and the FM-tone markers centered at 4000 Hz, all of which have ''flat'' envelopes. In three of the conditions in this set, marker 1 was a 4000-Hz pure tone, and marker 2 was: (1) a 4000-Hz pure tone; (2) a 2285-Hz pure tone; and (3) a 10-Hz modulated FM tone. The frequency of 2285 Hz was chosen to be greater than half an octave below the test frequency of 4000 Hz, since gap detection becomes an across-frequency process beyond this interval (e.g., Formby *et al.*, 1996). Moreover, the 7/4 ratio between the two frequencies is essentially inharmonic. In the other three conditions in this set, marker 1 was a 10-Hz modulated FM tone, and marker 2 was: (1) a 4000-Hz pure tone; (2) a 10-Hz modulated FM tone; and (3) a 20-Hz modulated FM tone. Four of the listeners (Obs. 1–4) undertook condition set 1. The second condition set utilized the AM-tone markers centered at 4000 Hz, with temporally fluctuating envelopes. In the three conditions of this set, the marker 1–marker 2 AM rate combinations were: (1) 10 Hz–10 Hz; (2) 10 Hz–20 Hz; and (3) 20 Hz–20 Hz. Five listeners (Obs. 2, 4, 5, 6, and 7) undertook all three conditions, while two listeners (Obs. 1 and 3) undertook just the first two conditions at an earlier time. In each of the two condition sets, isotype configurations existed where markers 1 and 2 were identical except for random duration differences. Here, no perceptual differences existed between the two markers. In the remaining configurations, a perceptual difference existed between marker 1 and marker 2, although no assumptions are made as to the equivalence of the marker separations in perceptual space. The purpose of the experiment was to determine whether gap detection was poorer in conditions where a perceptual difference existed between markers, compared to conditions where

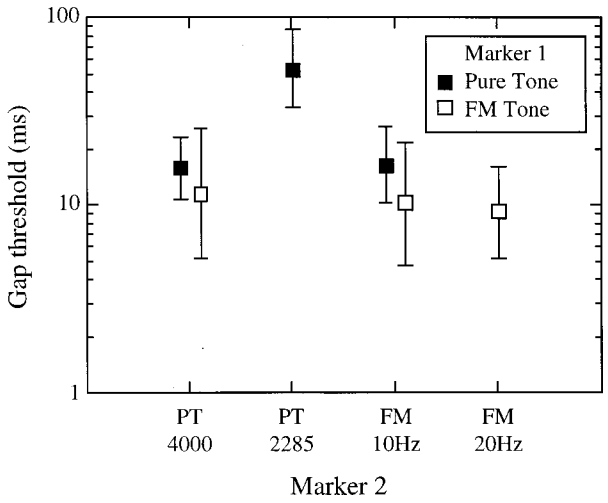


FIG. 2. Group mean gap thresholds for configurations where marker 1 is a pure tone (filled symbols) or an FM tone (unfilled symbols). The abscissa shows the characteristics of marker 2. Error bars indicate ±1 standard deviation.

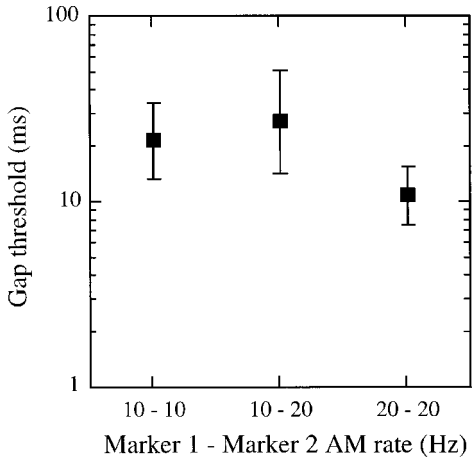


FIG. 3. Group mean gap thresholds for configurations where the gap markers are AM tones. The abscissa shows the configurations of marker 1–marker 2 AM rate. Error bars indicate ±1 standard deviation.

no such perceptual difference existed, and whether the across-frequency condition appeared to be a specific case of a general performance deficit.

B. Results and discussion

The mean results for the four listeners who undertook condition set 1 are shown in Fig. 2. Gap threshold is plotted for each marker 2 type, with marker 1 type being the parameter. Error bars indicate ±1 standard deviation.[2] Thresholds were similar for all conditions (9–16 ms), except for the configuration where the gap markers were pure tones of different frequencies. Here the threshold was elevated to more than 50 ms. This pattern of results was confirmed with a repeated measures analysis of variance (ANOVA) which indicated a significant effect of marker type ($F_{5,15}=18.25$; $p<0.001$). Note that in this, and all subsequent analyses, the log transforms of the raw data were employed. Post hoc contrasts showed that the threshold for the frequency-disparate marker configuration was significantly higher than all the other marker configurations combined ($F_{1,15}=79.6$; $p<0.001$).

The mean results for condition set 2 are shown in Fig. 3. A univariate repeated measures ANOVA indicated a significant effect of marker configuration ($F_{2,10}=16.17$; $p<0.001$). Post hoc contrasts showed that this effect was due to the threshold for the isotype 20-Hz AM tone markers (11 ms) being lower than either configuration where marker 1 was a 10-Hz AM tone (21 and 27 ms). The thresholds for the latter two configurations did not differ.

The key feature to the results from the two condition sets is that gap threshold did not uniformly worsen when perceptual dissimilarities were introduced between the two markers of the gap. When the dissimilarities took the form of either the presence/absence of FM or different rates of FM, thresholds were relatively stable compared to the baseline condition where both markers were isofrequency pure tones. When marker 1 was a 10-Hz AM tone, thresholds were similar across configurations where marker 2 was a 10-Hz AM tone or a 20-Hz AM tone. It should be noted that, in general, thresholds were somewhat higher for configurations where marker 1 was a 10-Hz AM tone, compared to the ''flat'' envelope configurations of condition set 1. It is likely that this is a consequence of the chosen starting phase of the AM waveform. As can be seen in Fig. 1, the choice of a π-radian starting phase for the modulator meant that the abutted markers in the standard (no-gap) stimulus were both at an envelope minimum in the cycle of waveform modulation at the point of abutment. This, in effect, amounted to the presence of a longer gap in the standard stimulus compared to stimuli made up of pure-tone or FM-tone markers, where the duration of the standard gap was dictated by the 5-ms offset/onset ramps. Because detection of increments in the duration of a gap is a function of the base duration of the gap (Abel, 1972; Penner, 1976; Divenyi and Sachs, 1978), it is probable that the somewhat higher thresholds obtained with the 10-Hz modulated AM-tone marker were a consequence of the longer effective base duration of the standard gap. This is supported by the observation that thresholds improved for the 20 Hz–20 Hz AM gap markers, where the higher modulation rate resulted in a shorter ''base'' silent interval in the standard stimulus. Nevertheless, even the threshold for the heterogeneous 10 Hz–20 Hz marker configuration was significantly lower than that of the frequency-disparate pure-tone configuration ($F_{1,6}=11.09$; $p=0.016$). This latter observation underscores the essential finding that the elevation in threshold observed when the two markers of a gap are widely separated in frequency exceeds that seen for the other combinations of dissimilar marker types used here. While it remains an assumption that the given acoustic differences translate to perceptual differences, these results strongly suggest that performance in across-frequency gap detection does not reflect primarily a sensitivity to perceptual discontinuities, per se. That is, the simple presence of perceptual dissimilarities across the two gap markers is not itself a sufficient factor to disrupt gap detection.

Van Wieringen and Wouters (1999), in their cochlear implant study, argued that the decline in gap detection per-

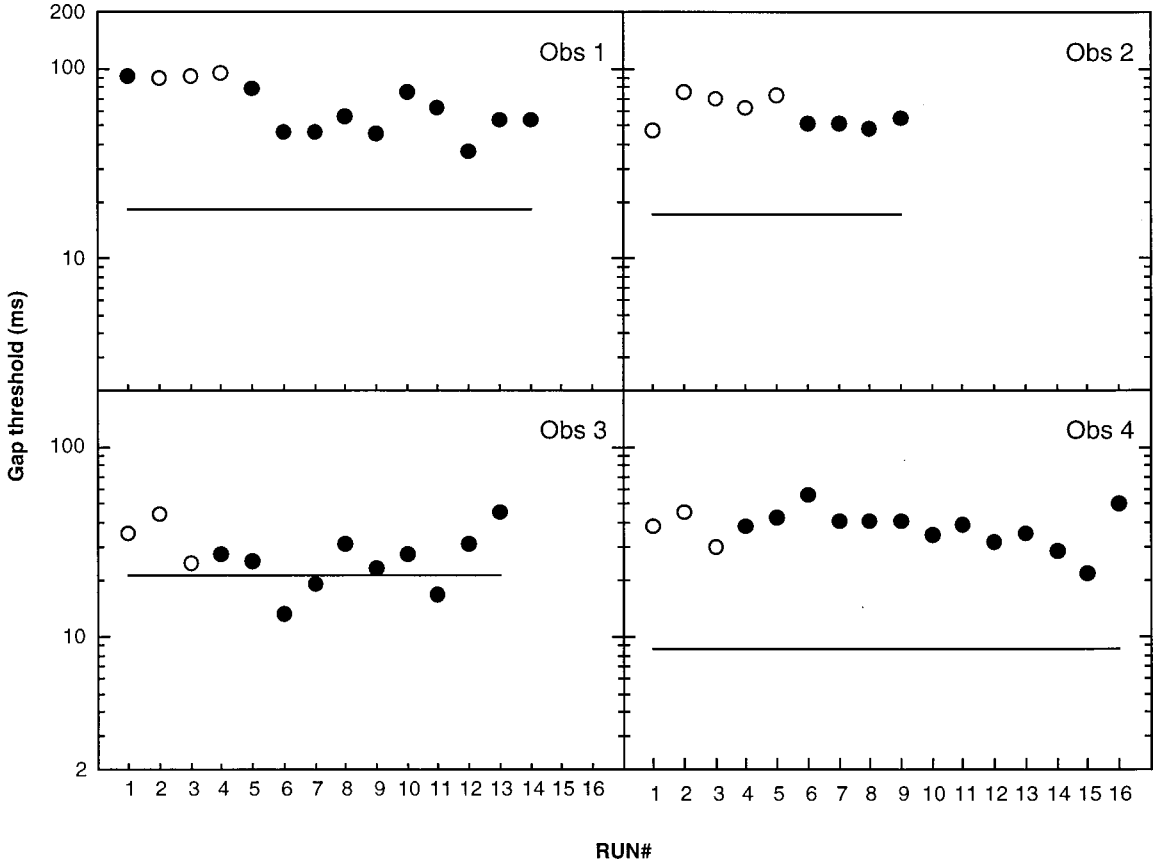


FIG. 4. Individual gap thresholds in the across-frequency pure-tone condition as a function of run number. The open symbols indicate the initial thresholds (obtained after training) on which performance in the main experiment was based. The horizontal line indicates the isofrequency gap threshold for that listener.

formance seen for across-electrode-channel configurations was due to perceptual discontinuities associated with the channel change. To support this argument, they showed that for two of their listeners there was a significant improvement in performance across the multiple repetitions of each across-electrode-channel condition. For one of these listeners, the across-electrode-channel effect was greatly reduced, and in the other it disappeared completely. Van Wieringen and Wouters interpreted this improvement as showing that familiarity with the perceptual characteristics of the stimuli through training could overcome any apparent across-channel effects. To determine whether listeners with normal hearing could also be ''trained out'' of an across-frequency effect, four listeners (Obs. 1–4) in this study underwent serial testing on the across-frequency pure-tone condition. The number of serial runs per listener varied between 9 and 16. The time-series results for the individual listeners are shown separately in the panels of Fig. 4. The open symbols indicate the initial thresholds (obtained after practice) from which the threshold data shown in Fig. 2 were calculated. The filled symbols indicate the additional threshold estimates which, for all listeners, were collected at a later date. For one listener (Obs. 1), the additional data were collected over two sessions; for the remaining observers they were collected in a single session. The horizontal line in each panel represents the baseline isofrequency threshold for that listener. For three of the four listeners (Obs. 1, 2, and 4), continued training did not result in a convergence of the across-frequency threshold with the isofrequency threshold. For one listener (Obs. 3), thresholds in the across-frequency condition were occasionally as good as the isofrequency threshold. This listener had the smallest difference between the across-frequency and isofrequency thresholds of the four listeners to begin with. However, even after repeated runs, there did not appear to be a stable convergence on the isofrequency threshold. Thus the overall pattern of results does not support the notion that performance in the across-frequency gap condition is due to perceptual dissimilarities that can be overcome with continued training.

In summary, this experiment sought to determine whether across-frequency gap detection reflects mechanisms that are sensitive primarily to the perceptual dissimilarity between sequential sounds. The results showed that gap thresholds for frequency-disparate pure-tone markers were elevated with respect to isofrequency tonal markers. However, other acoustic differences between markers that were presumed to also result in some degree of perceptual discontinuity between markers did not uniformly elevate thresholds. This finding is in line with the gap detection study of Formby *et al.* (1993) who found that for markers comprised of speechlike bands of noise, sensitivity to the gap was dependent on the differences in center frequency of the marker bands. Other marker differences in terms of bandwidth, level and duration contributed very little to gap sensitivity.

III. EXPERIMENT 2. GAP DETECTION FOR NOISE MARKERS

The results of experiment 1 leave open two questions: (1) Is the elevation in gap threshold for frequency-disparate markers due to spectral discontinuity, per se, or more generally to spectral dissimilarity between markers?; and (2) Is the poorer performance for frequency-disparate markers related to spectral differences between the markers, or to pitch differences? The question of whether spectral discontinuity or spectral dissimilarity is more relevant to across-frequency gap detection leads back to two studies described in Sec. I. Phillips *et al.* (1997) found that the detection of a gap marked by two noise stimuli was poorer for a configuration where the first marker was a wide-band noise and the second marker was a narrow-band noise than for a configuration where both markers were isofrequency narrow-band noises. Here, the elevated threshold occurred for a configuration where the spectral content of the first marker completely overlapped that of the second marker, making it a case of spectral dissimilarity but not spectral discontinuity in the frequency region of the narrow-band noise. The study of Formby *et al.* (1998b) used tonal markers and made the interesting finding that, relative to isofrequency single tone thresholds, performance declined when the second marker was accompanied by an isochronous second tone, but not when the first marker was accompanied by an isochronous second tone. Here, again, performance appears to be sensitive to spectral dissimilarity (i.e., overlapping but dissimilar spectra), in contrast to spectral discontinuity (i.e., nonoverlapping spectra). However, notice that the direction of the effect in the Formby *et al.* (1998b) study does not align itself with the results of the Phillips *et al.* (1997) study. In the Formby *et al.* (1998b) study, a single-tone marker followed by a widely spaced two-tone marker had a disruptive effect, but a widely spaced two-tone marker followed by a single-tone marker had little effect. The cochlear implant study of van Wieringen and Wouters (1999) also showed a similar asymmetry. In contrast, Phillips *et al.* (1997) observed a marked disruption for the ''wide-band''/''narrow-band'' case. Their study did not include the analogous ''narrow-band''/''wide-band'' configuration. In order to clarify the effect of spectral dissimilarity versus spectral discontinuity, experiment 2 examined gap detection using noise band markers, where the markers overlapped in spectral content but varied in overall bandwidth. In addition, because Phillips *et al.* (1997) found that performance for the spectrally dissimilar markers was sensitive to the duration of the leading marker, the parameter of marker 1 duration was also varied.

The question of the contribution of pitch differences versus spectral differences to across-frequency gap detection has been examined by Oxenham (2000). He measured detection of gaps marked by harmonic complexes that had been bandpass filtered between 4000 Hz and 8000 Hz to remove any resolvable components. When the gap markers had different fundamental frequencies, gap thresholds were elevated relative to the case where both markers had the same fundamental frequency. This finding suggests that pitch differences are sufficient to disrupt gap detection performance. However, he also included configurations where the gap markers had the same fundamental frequency but were bandpass filtered into nonoverlapping spectral regions. Here, performance declined even further, suggesting that spectral discontinuity was a more over-riding factor than pitch discontinuities. In order to address this issue further, experiment 2 measured gap detection for wide-band iterated rippled noises (IRNs) where the perceived pitch of the two gap markers was varied, although the markers stimulated the same frequency region.

A. Method

1. Subjects

Six listeners from the pool of eight participated in this experiment (Obs. 2, 4, 5, 6, 7, and 8). However, Obs. 5 and 6 did not undertake the two IRN conditions.

2. Stimuli

The acoustic markers bounding the gap were noise stimuli of three possible types. One type was a wide-band Gaussian noise (WBN), low-pass filtered at 8000 Hz. The second type of noise was a relatively narrow, digitally filtered, band of noise (NBN) extending from 841 Hz to 1189 Hz; i.e., a quarter octave below, to a quarter octave above, 1000 Hz. The third type of noise was an iterated rippled noise (IRN), bandpass filtered between 2500 and 8000 Hz (cf. Yost *et al.*, 1996). Two IRNs were employed, each with a different delay constant. One had a delay of 4 ms, yielding a pitch of 250 Hz, whereas the other had a delay of 7 ms, yielding a pitch of about 143 Hz. These delays were chosen to result in pitches that did not form an harmonic interval. Regardless of delay constant, each IRN was generated with ten iterations of delay and add. When the first marker was a WBN or NBN stimulus, its nominal duration was either 12.5 ms or 300 ms; the nominal duration of the leading IRN marker was always 300 ms. For all stimuli, the nominal duration of marker 2 was 300 ms. Regardless of nominal duration, the actual duration of each marker on every presentation was randomly varied by $\pm 20\%$. The randomization of duration was implemented to prevent overall stimulus duration (2 markers + gap) from being a viable detection cue. A 5-ms cosine-squared ramp was imposed on the beginning and end of each marker. The stimuli were digitally generated at a rate of 20 kHz (TDT AP2) and were low-pass filtered at 8 kHz (Kemo VBF 10M) prior to presentation. New tokens were generated for every interval of the 3AFC trial procedure.

3. Procedure

To make the presentation levels similar to those of Phillips *et al.* (1997), who presented all markers at 40 dB SL regardless of duration, the first phase of the procedure was to obtain detection thresholds for all markers. A 3AFC task was employed, with the signal occurring in one of the three observation intervals at random.[3] A 3-down, 1-up stepping rule converged upon the 79.4% correct detection threshold. The initial step size for level change was 8 dB and this was reduced to 4 dB after two reversals in level direction. After a further two reversals, the step size was reduced to its final

TABLE II. Marker configurations for experiment 2. The stimuli were either wide-band noises (WBN), narrow-band noises (NBN), or iterated rippled noises (IRN). For the WBN and NBN markers, the leading marker could have a nominal duration of 12.5 ms (short duration) or 300 ms (long duration). For the IRN markers, the add-back delay could be either 4 ms or 7 ms.

	Marker 1	Marker 2
1	WBN (short duration)	WBN
2	WBN (long duration)	WBN
3	WBN (short duration)	NBN
4	WBN (long duration)	NBN
5	NBN (short duration)	NBN
6	NBN (long duration)	NBN
7	NBN (short duration)	WBN
8	NBN (long duration)	WBN
9	IRN (4-ms delay)	IRN (4-ms delay)
10	IRN (4-ms delay)	IRN (7-ms delay)

value of 2 dB. A threshold track was terminated after 12 reversals, and the average level of the final 8 reversals was taken as the threshold estimate for that track. At least three estimates were collected for each stimulus and the arithmetic mean of all estimates was taken as the threshold value for that stimulus. The procedure for measuring gap detection was the same as that used in experiment 1 except that each marker was presented at 40 dB SL for each listener. The marker configurations employed in experiment 2 are listed in Table II. The order of conditions was randomized across listeners and, for each condition, practice was continued until performance was stable; i.e., until thresholds reached asymptote.

B. Results and discussion

The group mean results for the conditions involving Gaussian noisebands are shown in Fig. 5. Gap thresholds are plotted categorically for the four possible combinations of bandwidth and duration for marker 1. The parameter is the bandwidth similarity of marker 2 to marker 1, being either the same (filled symbols) or different (open symbols). There are three main questions that these data address: (1) Does changing the spectral similarity (i.e., bandwidth similarity) of the gap markers affect performance?; (2) Does the duration of the leading marker affect performance?; and (3) Does the temporal order of the marker bandwidth change (wide/narrow, narrow/wide) affect performance? To interpret the data in light of these questions, a repeated measures ANOVA with three within-subject factors (marker bandwidth similarity, marker 1 duration, and marker 1 bandwidth) was undertaken. The analysis revealed a significant three-way interaction between these factors ($F_{1,5}=23.84$; $p=0.005$) which was interpreted by assessing the pairwise means contrasts using Bonferoni correction for the ten relevant comparisons (criterion $p=0.005$). The first four comparisons spoke to the first question of whether changing the bandwidth from marker 1 to marker 2 affected the gap detection threshold. In terms of Fig. 5, this involves determining whether the open symbols (different bandwidths) are reliably above the filled symbols (same bandwidths) for each marker 1 category. The contrasts indicated that, for three of the four cases, varying the marker bandwidth across the gap did disrupt threshold

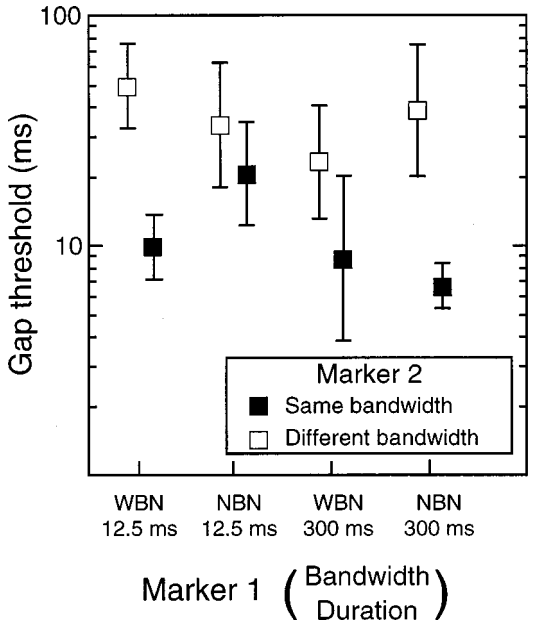


FIG. 5. Group mean gap thresholds plotted categorically for the four combinations of bandwidth and duration that featured in marker 1. The parameter is the bandwidth similarity of marker 2 to marker 1, being either the same (filled symbols) or different (open symbols). Error bars indicate ±1 standard deviation.

relative to that obtained with a constant bandwidth (48 ms vs 10 ms when marker 1 was a short-duration WBN; 23 ms vs 9 ms when marker 1 was a long-duration WBN; and 39 ms vs 7 ms when marker 1 was a long duration NBN). The one exception was for the configuration where marker 1 was a short duration narrow band of noise. Here, thresholds were not sensitive to whether marker 2 was also a narrow band of noise, or was a wide band of noise (20 ms vs 33 ms). However, it is evident from Fig. 5 that this specific outcome appears to result not so much from good performance for the ''different bandwidth'' configuration, but from an elevation in threshold for the ''same bandwidth'' configuration. This finding leads into the second question of whether the duration of marker 1 affected the pattern of results, and four means contrasts are relevant to this issue (i.e., the short versus long marker 1 durations for the marker 1–marker 2 configurations of WBN-WBN, NBN-NBN, WBN-NBN, and NBN-WBN). For the WBN-WBN configuration, the resulting comparison was 10 ms vs 9 ms; for the NBN-NBN configuration, the comparison was 20 ms vs 7 ms; for the WBN-NBN configuration, the comparison was 48 ms vs 23 ms; and for the NBN-WBN configuration, the comparison was 33 ms vs 39 ms. These comparisons revealed that the only case in which marker 1 duration affected performance was for the configuration just described in which both markers 1 and 2 were narrow bands of noise. Here, gap detection thresholds improved significantly from 20 ms to 7 ms as the nominal duration of marker 1 increased from 12.5 ms to 300 ms. Finally, as to the third question, for the configurations where the temporal order of the marker bandwidth changed across the gap, the order of the bandwidth change did not affect performance. The two means contrasts relevant to this ques-

tion are the WBN-to-NBN versus NBN-to-WBN comparisons for the short duration marker 1 (48 ms vs. 33 ms, respectively) and for the long duration marker 1 (23 ms vs. 39 ms, respectively). The two contrasts indicated that the gap detection was not affected by whether markers 1 and 2 changed from wide band to narrow band, or narrow band to wide band.

This pattern of results contains some features that are common to earlier studies, but also highlights some inconsistencies. Like Phillips *et al.* (1997), this experiment found that a wide-band noise marker followed by a narrow-band noise marker resulted in poorer gap thresholds than if both markers had narrow bandwidths, irrespective of the duration of the leading marker. However, unlike Phillips *et al.* (1997), gap detection for isotype narrow-band markers *was* sensitive to the duration of the leading marker. This finding is supported by the gap duration discrimination work of Rammsayer and Leutner (1996) who found that the discrimination of the duration of a silent interval bounded by two isofrequency tones was dependent on the duration of the leading marker. Schneider and Hamstra (1999), using a gap detection paradigm where marker duration was varied equivalently for both markers 1 and 2, also showed that within-channel gap detection was sensitive to marker duration. Snell and Hu (1999) also found an effect of marker 1 duration, where gap detection thresholds were sometimes elevated for short duration leading markers. However, it is evident from their data that this effect was less marked for the most experienced listeners. In line with this, Phillips *et al.* (1998) noted that a within-channel marker duration effect appeared to be present for their two most inexperienced listeners. In considering the contrasting effects of marker 1 duration for isofrequency markers observed in the Phillips *et al.* (1997) study and in this study, differences between the experimental paradigms should be noted. For example, in the Phillips *et al.* (1997) study, the standard gap was dictated by 0.5-ms rise/fall times on the markers (rendering the standard gap ''inaudible''), and several different fixed durations of the leading marker were employed. In the present study, the standard gap was dictated by 5-ms rise/fall times on the markers, and only two nominal durations of marker 1 were used. Further work to determine the conditions under which within-channel gap detection is sensitive to the duration of the leading marker is warranted because this sensitivity has been proposed to distinguish between within-channel and across-channel mechanisms (e.g., Phillips *et al.*, 1997). In summary, the results of the Gaussian noiseband conditions indicated that, in most cases, gap detection was disrupted by introducing spectral dissimilarities (i.e., bandwidth differences in overlapping spectral regions) across the gap, and that the order of the spectral changes made no difference. This suggests that gap detection is sensitive to spectral dissimilarity between markers in addition to spectral discontinuity, or nonoverlapping frequency differences. The exception to the general pattern was for configurations where marker 1 was a narrow band of noise. Here, gap detection was elevated when the duration of marker 1 was short, even when the bandwidth across the markers was held constant.

The results of the IRN conditions showed no effect of perceived pitch. The mean thresholds for the gaps marked by IRNs with the same pitch and different pitches were 8.0 ms and 9.9 ms, respectively. A dependent *t*-test indicated that these thresholds did not differ reliably ($t_3 = 1.41$; $p = 0.25$). Recall that, although the pitch of the stimuli differed, the spectral region occupied by the filtered IRNs was the same. This result contrasts to some degree with that of Oxenham (2000) who found, using filtered harmonic complexes, that changing fundamental frequency while holding spectral region constant did result in a small decline in performance. It is possible that the differences in finding may be due to differences in the degree of temporal envelope fluctuation observed for high-pass filtered harmonic complexes versus IRNs. Gap detection is known to be dependent on the degree of fluctuation in the gap markers (Eddins *et al.*, 1992; Glasberg and Moore, 1992).

IV. GENERAL SUMMARY

The two experiments reported in this study were motivated by the question of whether gap detection for across-frequency markers is a specific instance of poor gap detection for perceptually dissimilar markers. The first experiment verified that thresholds for gaps marked by frequency-disparate markers are poorer than for gaps marked by isofrequency markers. However, by testing gap detection for mixed marker types (pure tones, FM tones, and AM tones), it was shown that performance did not uniformly decline when perceptual dissimilarities existed between the markers. The second experiment also showed that, in most cases, gap detection deteriorates when the bandwidth of the two markers differ, even when there is complete spectral overlap between the two markers. This finding suggests that gap detection is sensitive to spectral dissimilarity between markers in addition to spectral discontinuity. The second experiment also showed that pitch differences across spectrally similar markers are not sufficient to disrupt gap detection performance.

Several findings from earlier studies reviewed in Sec. I, as well as aspects of the experiments reported here, indicate that gap detection is sensitive to the acoustic similarity of the two markers of the gap (or, for listeners with cochlear implants, the electrical similarity of the gap markers). However, it is evident that acoustic dissimilarity in itself is not sufficient to disrupt gap detection. From experiment 1, it would seem that acoustic dissimilarities—with their associated perceptual dissimilarities—that are restricted to the same frequency region are not sufficient to disrupt gap detection. Similarly, from experiment 2, it is evident that pitch differences between stimuli restricted to the same frequency region are not sufficient to disrupt gap detection. In contrast, acoustic differences that involve spectral dissimilarities across the gap do affect gap detection. Further work is required to resolve the question of the interaction of spectral disparity versus spectral overlap in the gap markers, and the symmetry of the effects in terms of the order of the gap markers.

ACKNOWLEDGMENT

This work was supported by the NIH NIDCD (R01-DC01507).

[1]Specifically, the amplitude was adjusted by a factor of $1/\sqrt{1+(m^2/2)}$, where the modulation index, m, was set equal to 1.

[2]Because the individual gap thresholds were averaged in the log duration domain, the standard deviations bracket the mean value symmetrically when plotted on the log duration coordinates of Figs. 2, 3, and 5.

[3]For the detection threshold phase of the experiment, the actual durations of the stimuli equaled the nominal durations (12.5 ms or 300 ms). In other words, there was no randomization of stimulus duration for the measurement of detection thresholds.

Abel, S. M. (**1972**). ''Discrimination of temporal gaps,'' J. Acoust. Soc. Am. **52**, 519–524.

ANSI. (**1989**). ANSI S3.6-1989, ''Specifications for audiometers'' (American National Standards Institute, New York).

Chatterjee, M., Fu, Q.-J., and Shannon, R. V. (**1998**). ''Within-channel gap detection using dissimilar markers in cochlear implant listeners,'' J. Acoust. Soc. Am. **103**, 2515–2519.

Collyer, C. E. (**1974**). ''The detection of a temporal gap between two disparate stimuli,'' Percept. Psychophys. **16**, 96–100.

Divenyi, P. L., and Danner, W. F. (**1977**). ''Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra,'' Percept. Psychophys. **21**, 125–142.

Divenyi, P. L., and Hirsh, I. J. (**1978**). ''Some figural properties of auditory patterns,'' J. Acoust. Soc. Am. **64**, 1369–1385.

Divenyi, P. L., and Sachs, R. M. (**1978**). ''Discrimination of time intervals bounded by tone bursts,'' Percept. Psychophys. **24**, 429–436.

Eddins, D. A., Hall, J. W., and Grose, J. H. (**1992**). ''The detection of temporal gaps as a function of frequency region and absolute noise bandwidth,'' J. Acoust. Soc. Am. **91**, 1069–1077.

Fitzgibbons, P. J., Pollatsek, A., and Thomas, I. (**1974**). ''Detection of temporal gaps within and between perceptual tonal groups,'' Percept. Psychophys. **16**, 522–528.

Formby, C., and Forrest, T. G. (**1991**). ''Detection of silent temporal gaps in sinusoidal markers,'' J. Acoust. Soc. Am. **89**, 830–837.

Formby, C., Sherlock, L. P., and Forrest, T. G. (**1996**). ''An asymmetric roex filter model for describing detection of silent temporal gaps in sinusoidal markers,'' Aud. Neurosci. **3**, 1–20.

Formby, C., Sherlock, L. P., and Li, S. (**1998b**). ''Temporal gap detection measured with multiple sinusoidal markers: Effects of marker number, frequency, and temporal position,'' J. Acoust. Soc. Am. **104**, 984–998.

Formby, C., Barker, C., Abbey, H., and Raney, J. J. (**1993**). ''Detection of silent temporal gaps between narrow-band noise markers having second-formantlike properties of voiceless stop/vowel combinations,'' J. Acoust. Soc. Am. **93**, 1023–1027.

Formby, C., Gerber, M. J., Sherlock, L. P., and Magder, L. S. (**1998a**). ''Evidence for an across-frequency, between-channel process in asymptotic monaural temporal gap detection,'' J. Acoust. Soc. Am. **103**, 3554–3560.

Forrest, T. G., and Formby, C. (**1996**). ''Detection of silent temporal gaps in sinusoidal markers simulated with a single-channel envelope detector model,'' Aud. Neurosci. **3**, 21–33.

Glasberg, B. R., and Moore, B. C. J. (**1992**). ''Effects of envelope fluctuations on gap detection,'' Hear. Res. **64**, 81–92.

Grose, J. H., and Hall, J. W. (**1996**). ''Perceptual organization of sequential stimuli in listeners with cochlear hearing loss,'' J. Speech Hear. Res. **39**, 1149–1158.

Hanekom, J. J., and Shannon, R. V. (**1998**). ''Gap detection as a measure of electrode interaction in cochlear implants,'' J. Acoust. Soc. Am. **104**, 2372–2384.

Kinney, J. A. S. (**1961**). ''Discrimination ability in auditory and visual patterns,'' Am. J. Psychol. **74**, 529–541.

Neff, D. L., Jesteadt, W., and Brown, E. L. (**1982**). ''The relation between gap discrimination and auditory stream segregation.,'' Percept. Psychophys. **31**, 493–501.

Oxenham, A. J. (**2000**). ''Influence of spatial and temporal coding on auditory gap detection,'' J. Acoust. Soc. Am. **107**, 2215–2223.

Penner, M. J. (**1976**). ''The effect of marker variability on the discrimination of temporal intervals,'' Percept. Psychophys. **19**, 466–469.

Perrott, D. R., and Williams, K. L. (**1971**). ''Auditory temporal resolution: Gap detection as a function of interpulse frequency disparity,'' Psychonomic Sci. **25**, 73–74.

Phillips, D. P., Hall, S. E., Harrington, I. A., and Taylor, T. L. (**1998**). ''''Central'' auditory gap detection: a spatial case,'' J. Acoust. Soc. Am. **103**, 2064–2068.

Phillips, D. P., Taylor, T. L., Hall, S. E., Carr, M. M., and Mossop, J. E. (**1997**). ''Detection of silent intervals between noises activating different perceptual channels: Some properties of ''central'' auditory gap detection,'' J. Acoust. Soc. Am. **101**, 3694–3705.

Rammsayer, T. H., and Leutner, D. (**1996**). ''Temporal discrimination as a function of marker duration,'' Percept. Psychophys. **58**, 1213–1223.

Schneider, B. A., and Hamstra, S. J. (**1999**). ''Gap detection thresholds as a function of tonal duration for younger and older listeners,'' J. Acoust. Soc. Am. **106**, 371–380.

Snell, K. B., and Hu, H. (**1999**). ''The effect of temporal placement on gap detectability,'' J. Acoust. Soc. Am. **106**, 3571–3577.

van Wieringen, A., and Wouters, J. (**1999**). ''Gap detection in single- and multi-channel stimuli by LAURA cochlear implantees,'' J. Acoust. Soc. Am. **106**, 1925–1939.

Viemeister, N. F. (**1979**). ''Temporal modulation transfer functions based upon modulation thresholds,'' J. Acoust. Soc. Am. **66**, 1364–1380.

Yost, W. A., Patterson, R., and Sheft, S. (**1996**). ''A time domain description for the pitch strength of iterated rippled noise,'' J. Acoust. Soc. Am. **99**, 1066–1078.

The contribution of two ears to the perception of vertical angle in sagittal planes

Masayuki Morimoto[a)]
Environmental Acoustics Laboratory, Faculty of Engineering, Kobe University, Rokko, Nada, Kobe 657-8501, Japan

(Received 6 December 1999; revised 27 September 2000; accepted 3 January 2001)

Because the input signals to the left and right ears are not identical, it is important to clarify the role of these signals in the perception of the vertical angle of a sound source at any position in the upper hemisphere. To obtain basic findings on upper hemisphere localization, this paper investigates the contribution of each pinna to the perception of vertical angle. Tests measured localization of the vertical angle in five planes parallel to the median plane. In the localization tests, the pinna cavities of one or both ears were occluded. Results showed that pinna cavities of both the near and far ears play a role in determining the perceived vertical angle of a sound source in any plane, including the median plane. As a sound source shifts laterally away from the median plane, the contribution of the near ear increases and, conversely, that of the far ear decreases. For saggital planes at azimuths greater than 60° from midline, the far ear no longer contributes measurably to the determination of vertical angle. [DOI: 10.1121/1.1352084]

PACS numbers: 43.66.Qp, 43.66.Pn [DWG]

I. INTRODUCTION

Most previous studies of sound localization ability have concentrated on localization performance in the horizontal and median planes. It is generally known that binaural disparity cues [i.e., interaural time (phase) differences and interaural level differences] are cues for horizontal-plane localization, while spectral cues (i.e., spectral distortions produced by the pinnae) are cues for median-plane localization. Most previous studies on median-plane localization have treated spectral cues as a monaural auditory phenomenon, since the input signals of two ears from a sound source in the median plane are generally very similar. They have shown that spectral distortions caused by pinnae in the high-frequency range above about 5 kHz acts as cues for median-plane localization (e.g., Roffler and Butler, 1968; Blauert, 1969/1970; Gardner and Gardner, 1973; Hebrank and Wright, 1974a; Butler and Belendiuk, 1977; Watkins, 1978). Only a few studies have speculated that binaural spectral differences contribute to sound localization in the median plane (Butler, 1969; Searle *et al.*, 1975; Duda, 1997). These speculations have been questioned by several experimental findings (Gardner, 1973; Hebrank and Wright, 1974b; Hebrank, 1976; Morimoto and Nomachi, 1982).

If everyday life, however, a sound does not come from only the horizontal and median planes, but may come from any direction around a listener. There is a need to identify localization cues in such a case. Morimoto and Aokata (1984) demonstrated that an interaural-polar-axis coordinate system, as shown in Fig. 1, is more suitable for explaining sound localization in any direction in the upper hemisphere than a geodesic coordinate system defined by the azimuth angle ψ and the elevation angle θ. In an interaural-polar-axis coordinate system, the angle α is the angle between the aural axis and a straight line connecting the sound source with the center of a subject's head, and the angle β is the angle between the horizontal plane and the perpendicular from the sound source to the aural axis, that is, the vertical angle in a plane parallel to the median plane, called the sagittal plane. According to the results of localization tests, Morimoto and Aokata determined that angle α and angle β are independently determined by binaural disparity cues and spectral cues, respectively.

Middlebrooks (1992) constructed a quantitative model of two-dimensional localization in which the azimuth angle and the elevation angle of a subject's response were successfully predicted based on interaural level differences and spectral cues, respectively. In his experiments, the locations of stimuli and responses were specified in a double-pole coordinate system. In this coordinate system, the elevation angle is formed by the sound source (or response location), the center of the subject's head, and the horizontal plane. The azimuth angle is formed by the sound source, the center of the head, and the median plane. This coordinate system is a little different from Morimoto and Aokata's, but the two coordinate systems are similar, in that in both systems, the vertical location is determined in a sagittal plane parallel to the median plane. Accordingly, Middlebrooks' model supports the experimental results of Morimoto and Aokata.

In all geodesic, interaural-polar-axis, and double-pole coordinate systems, there is no doubt that a vertical location of a sound image is determined by spectral cues, though the definition of vertical angle depends on each coordinate system. Spectral cues have been treated as monaural cues in most studies on median-plane localization, as mentioned above. Since input signals to the right and left ears from a sound source in the median plane are generally very similar, it may be appropriate to consider spectral cues to be a monaural phenomenon in median-plane localization. A problem is that the input signals to the left and right ears from a sound

a)Electronic mail: mrmt@kobe-u.ac.jp

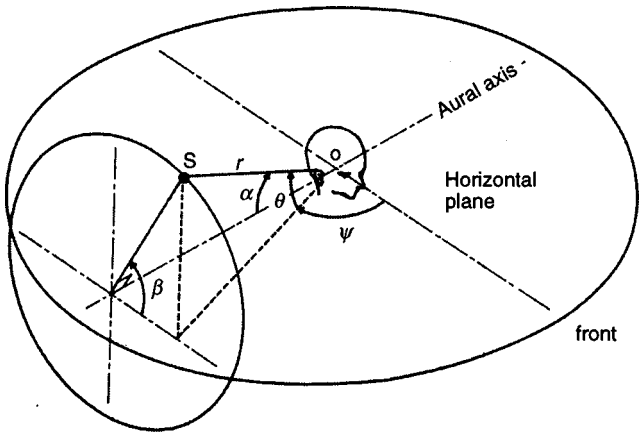


FIG. 1. Definition of an interaural-polar-axis coordinate system. S: sound source; O: center of the head; r: distance between a sound source and the center of head; ψ: azimuth angle; θ: elevation angle; α: the angle between the aural axis and a straight line connecting the sound source with the center of a subject's head; β: the angle between the horizontal plane and the perpendicular from the sound source to the aural axis, that is, the vertical angle in a plane parallel to the median plane, called the sagittal plane.

source outside the median plane are not spectrally identified.

In this case, how do the two ears play a role in the perception of the vertical angle outside the median plane? Musicant and Butler (1984) demonstrated that the ear on the side of the sound source (near ear) contributed powerfully to front–back discrimination, as did the ear on the side opposite the sound source (far ear). Humanski and Butler (1988) showed that the contribution of the near ear was greater than that of the far ear in determining the vertical location of a sound source in the sagittal plane. In both experiments, however, sound sources were arranged only on or near the horizontal plane.

In the present paper, localization tests of sound sources arranged throughout the upper hemisphere were carried out, adopting an interaural-polar-axis coordinate system as shown in Fig. 1. The purpose was to clarify the relationship between the degree of a shift of a sound source from the median plane and the degree of contribution of each ear to the perception of the vertical angle β. In these localization tests, pinna cavities of one or both ears were occluded in the same manner as in the experiments by Gardner (1973) and Humanski and Butler (1988).

II. METHOD

A. Subjects

Subjects were three male students, 22 years of age ± 1 year, with normal hearing sensitivity. All were experienced in this type of localization test.

B. Apparatus and stimulus

The localization tests were conducted in an anechoic chamber. The outer walls of the chamber were constructed of 30 cm thick reinforced concrete. Glass wool 25 cm thick was installed on the walls as sound absorbing material. The dimensions of the working area of the chamber were 3.90 m wide, 4.90 m deep, and 3.90 m high. The background noise level was below 20 dB(A).

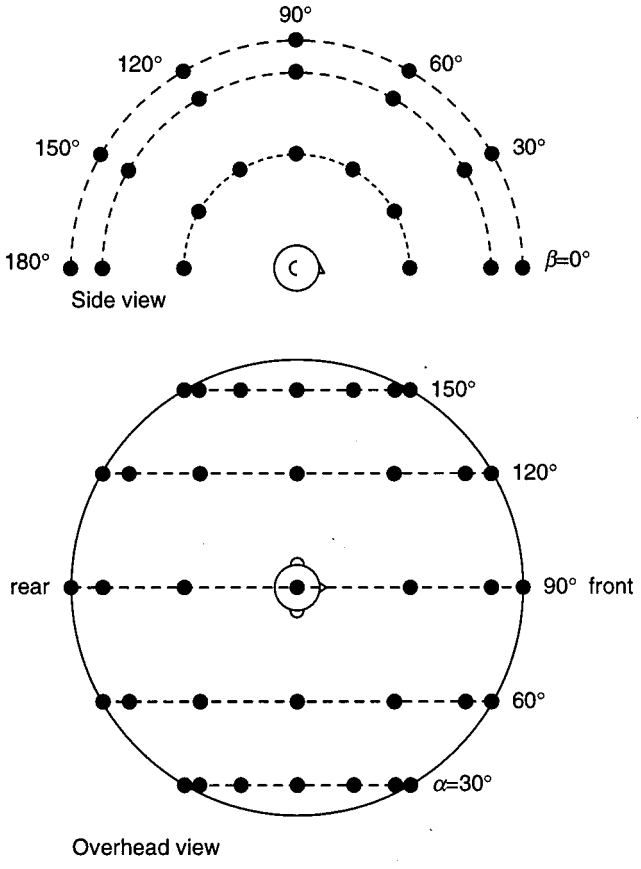


FIG. 2. Arrangement of loudspeakers used for the localization test. In the side view, loudspeakers in sagittal planes defined by angles $\alpha=120°$ and 150° are hidden by those in sagittal planes defined by angles $\alpha=60°$ and 30°, respectively.

Thirty-five small dome-type loudspeakers (Technics EAS-6KH, diameter: 53 mm) were used for the localization tests, as shown in Fig. 2. Their angle α settings were 30°, 60°, 90°, 120°, and 150°, and their angle β settings were 0°, 30°, 60°, 90°, 120°, 150°, and 180°. The speaker radius was 1.6 m relative to the center of the subject's head. The frequency characteristics of the 35 loudspeakers were flattened to within ± 3 dB in the frequency range of the stimulus by a frequency equalizer (Technics SH-9090).

The stimulus was a wide band white noise from 1 kHz to 16 kHz. The wide band white noise was generated from a Hewlett-Packard noise generator (model 3722 A) and was delivered to a custom-built timer gate with a abrupt rise–fall time. To establish the filtered noise condition, the burst was routed through an NF active filter (model FV625A), which resulted in a 48 dB/oct rejection slope. After appropriate filtering (bandpass from 1 kHz to 16 kHz), the stimuli were amplified (Onkyo Integra 713) and delivered one of loudspeakers via a custom-built loudspeaker selector. The stimulus was delivered at 50 dB(A) (SPL for 1 s, followed by an interval of 4 s.

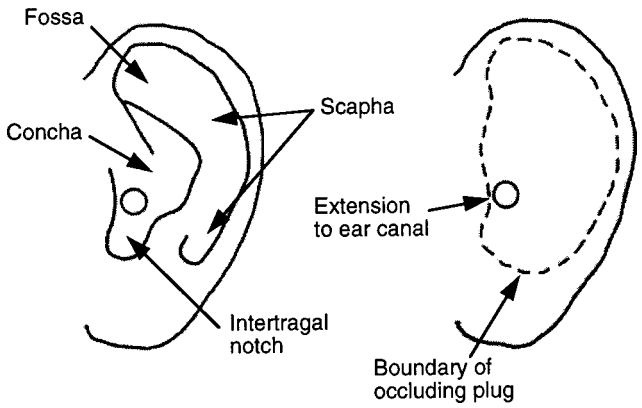


FIG. 3. The occluded part of a pinna.

C. Pinna conditions

The localization tests were done under four pinna conditions: (a) both ears open, i.e., the pinna cavities of both ears were not occluded; (b) right ear open, i.e., the pinna cavities of the right ear were entirely open while those of the left ear were occluded except for a passageway to the ear canal; (c) left ear open, i.e., the pinna cavities of the left ear were entirely open while those of the right ear were occluded except for a passageway to the ear canal; and (d) both ears occluded, as described above. The occluded part of cavities is shown in Fig. 3. In this experiment, the pinna cavities were occluded using a material normally used for dental impressions (Algix), and the passageway to the ear canal was made of a drinking straw. The surface of the occlusion was flattened to be level with the end of the straw and the helix. The length of the straw was almost equal to the distance from the surface to the entrance of the auditory canal.

Head-related transfer functions were measured at the entrance of the passageway of the occluded ear and at the entrance of the auditory canal of the open ear (see the Appendix). Figure 4 shows the head-related transfer functions of one of the three subjects from seven sound sources in the median plane (angle α=90°) to the occluded ear, compared with those to the open ear, as an example. Although the occlusion substantially distorted the spectral shape, it still provided consistent spectral information related to the source angle β. But Gardner and Gardner (1973) indicated that localization performance in the vertical plane was dramatically disrupted when cavities of both ears were occluded. Moreover, Hofman *et al.* (1998) demonstrated that more than several weeks were necessary to adapt subjects to new spectral cues after the occlusion and that when occlusion was removed immediately after adaptation, the subjects' localization accuracy with unoccluded ears was still as high as before the occlusion. In the present localization tests, the occlusion period was too short to allow subjects to adapt to spectral cues, as described in Sec. II D.

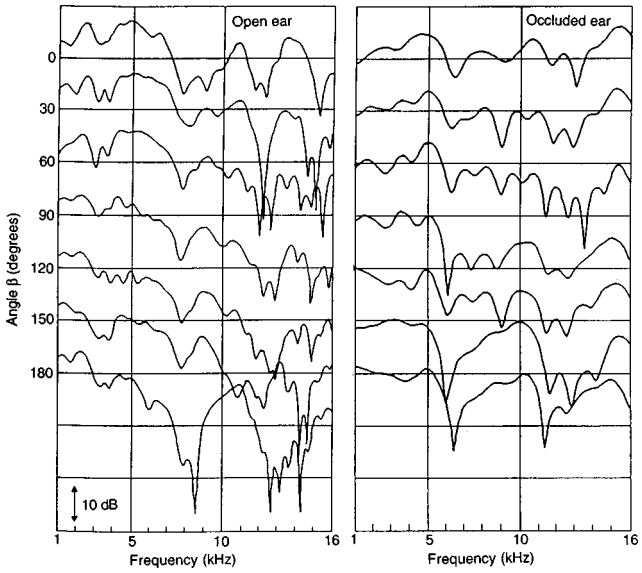


FIG. 4. Measured amplitudes of head-related transfer functions of an open ear (left) and an occluded ear (right). Left ear of one of three subjects. Source angle α=90°.

Another possible effect of the occlusion was to cause possible lateral displacement of a sound image, because occlusion can affect binaural disparity cues, i.e., interaural time (phase) differences and interaural level differences. However, Gardner (1973) stated that if a passageway was provided through the occluding plug, lateral displacements were avoided and the apparent location of the signal remained within, or very close to, the median plane. Accordingly, the occlusion of pinna cavities could be regarded as an adequate manipulation to achieve the purpose of the present paper.

D. Procedure

Each subject was tested individually while seated, with his head fixed in a stationary position, in a partially darkened anechoic chamber. Seventy-seven recording sheets, each with a circle and an arrow on it, were supplied to the subject. An arrow on the top of the circle indicated the 0° of angle β. The subject's task was to mark down the perceived angle β for each one-second stimulus presentation. The 4 s interstimulus interval allowed the subject to pick up the next record sheet. The only light in the chamber was placed such that it provided just enough illumination for the subject to see and utilize the record sheets.

Localization tests for each subject were performed in two days. On the first day, tests for both ears open condition were carried out to confirm the localization ability of the subjects. On the second day, tests for the other pinna conditions were carried out, separated by breaks of about 30 min each in the order right ear open, both ears occluded, and left ear open. For each pinna condition, a subject was tested with each angle α in random order, separated by a break of at least 5 min while he was seated. For each angle α, 11 sets of stimuli were presented successively to the subject. In a set of stimuli, the stimulus was presented once from a loudspeaker at each of seven angles β in random order. The presentation order of the seven angles β was different between sets. Thus each subject made a total of 1540 judgments for the entire task (4 pinna conditions, 5 angle α settings, 7 angle β settings, and 11 presentations for each angle β).

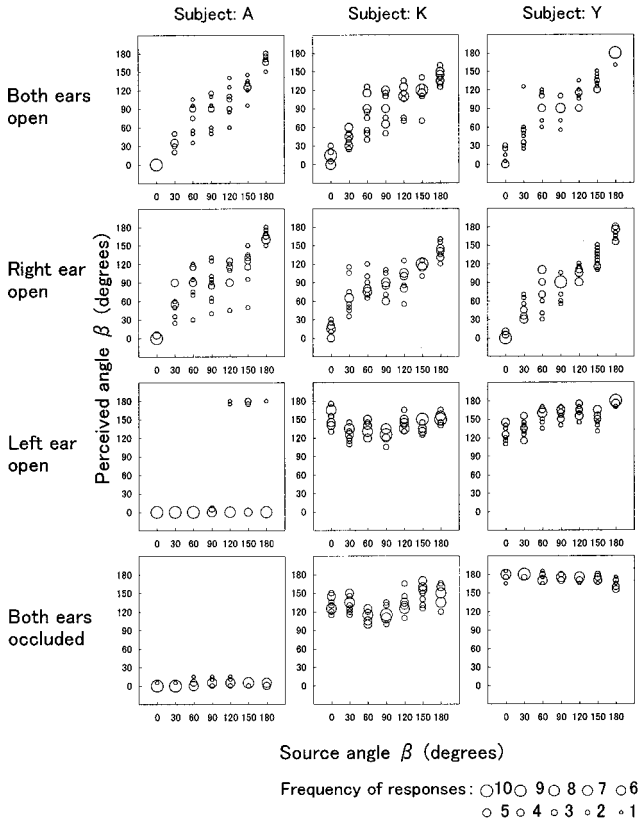


FIG. 5. Subjects localization responses β as a function of source angle β. Each column displays the results for a different subject. Each row displays the results for a different pinna condition (labeled to the left). The diameter of the data points indicates the number of responses, as shown by the key at the bottom. Source angle α=30°.

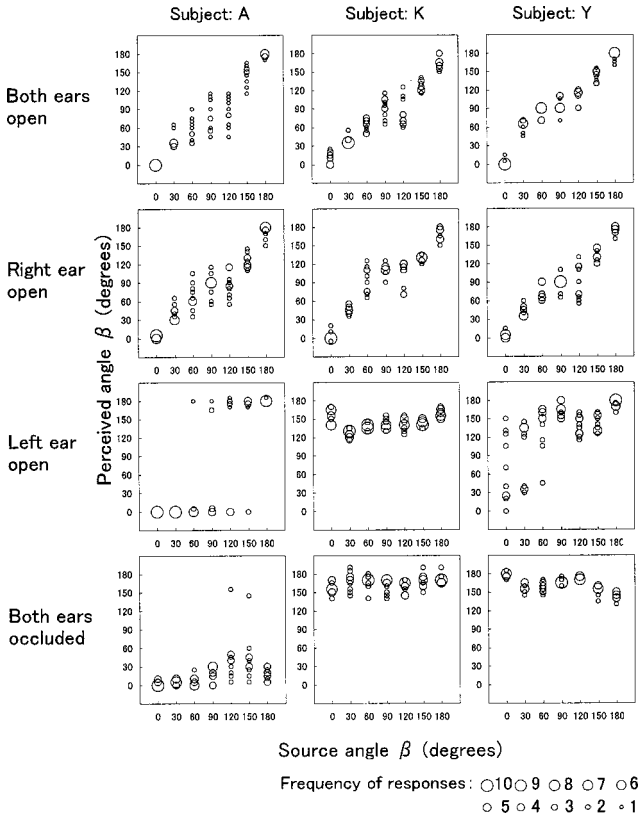


FIG. 6. As Fig. 5 for the angle α=60°.

III. RESULTS

Sometimes subjects reported that sound images for sound sources in the median plane (angle α=90°) shifted a little to the open ear side. It appears that the pinna occlusion slightly altered binaural disparity cues [i.e., interaural time (phase) differences and/or interaural level differences] as described in Sec. II C. However, Morimoto and Aokata (1984) and Middlebrooks (1992) made clear that angle α and angle β are independently determined. Moreover, subjects who perceived a lateral displacement of a sound image stated that they could judge angle β without hindrance because of the slight displacement. Therefore, the lateral displacement was not regarded as an obstacle to achieving the purpose of the present paper.

Responses to the first test sets for each angle α under each pinna condition were regarded as practice and were excluded from the results. The angle marked by the subject was read with a protractor to an accuracy of one degree.

The effects of pinna occlusion on the perception of angle β in sagittal planes were investigated with distribution charts of subjects' responses. Results are displayed in Figs. 5–9 for angles α=30°, 60°, 90°, 120°, and 150°, respectively. Within each figure the three columns show the individual data for the three subjects, and the four rows display the four pinna conditions. The diameter of each circle plotted is proportional to the number of responses within 5°. The ordinate of

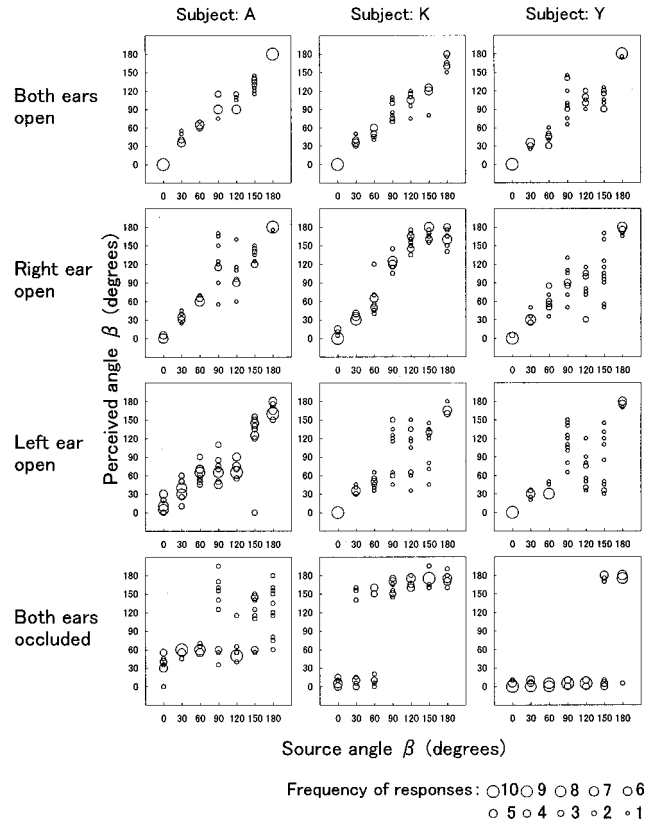


FIG. 7. As Fig. 5 for the angle α=90°.

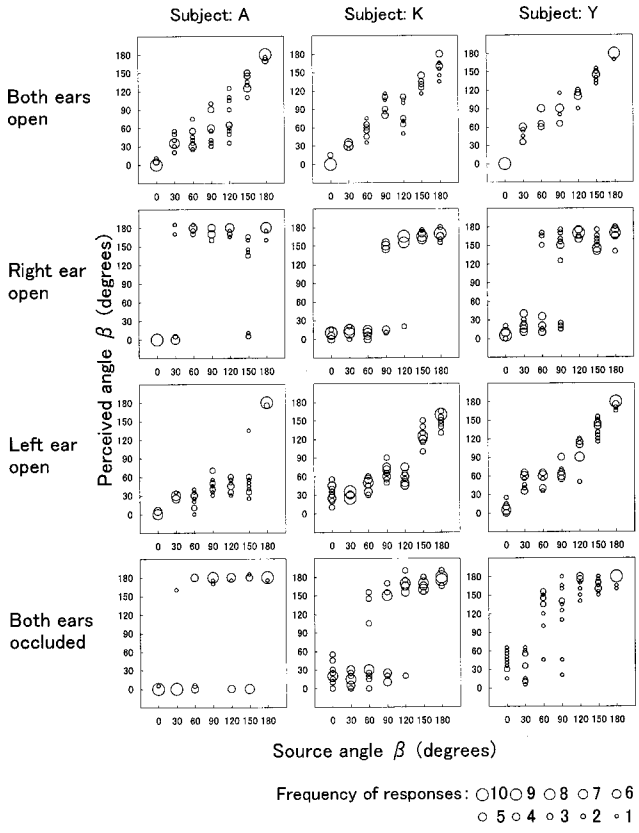


FIG. 8. As Fig. 5 for the angle α=120°.

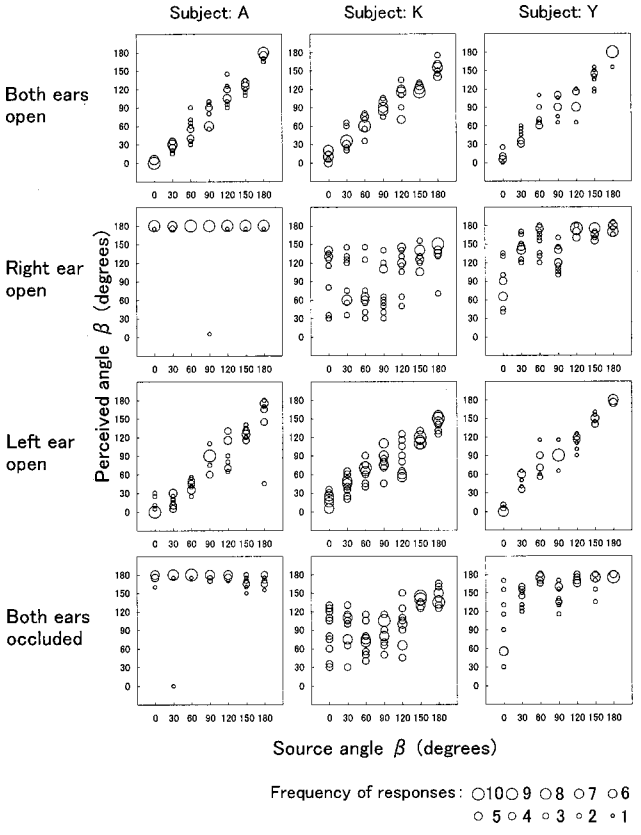


FIG. 9. As Fig. 5 for the angle α=150°.

each panel is the perceived angle β, and the abscissa is the source angle β.

Distinct individual differences in localization performance appeared in only two cases: the responses of subject A for angle α=150° under the right-ear-open condition (Fig. 9, second row) and those of subject K for angle α=150° under the both-ears-occluded condition (Fig. 9, bottom row). Although the responses of subject A for angle α=30° under the left-ear-open condition (Fig. 5, third row) and for angle α=30° and 60° under the both-ears-occluded condition (Figs. 5 and 6, bottom row) seem to be different from those of subjects K and Y, a tendency of responses to concentrate in one direction was common to all three subjects. Furthermore, although the responses of the three subjects for angle α=60° under the left-ear-open condition (Fig. 6, third row) seem to differ from each other, a tendency for few responses to appear around angle α=90° and most to be near 0° (front) and 180° (rear) was common to all three subjects. In summary, some individual differences were observed, yet one could see certain general trends.

Figure 7 shows the perceived angles β for sound sources at angle α=90°, i.e., in the median plane. When both ears were open (top row), the perceived angles β practically coincided with the source angles β. The variances of the responses for the source angles β around 90° were larger than those for angle β near 0° (front) or 180° (rear). This tendency coincides with localization blur in the median plane (Damaske and Wagener, 1969; Kurosawa *et al.*, 1982). This means that the response method with recording sheets used in the present tests was adequate. When both ears were occluded (bottom row), the perceived angles β were practically independent of the source angles β: most responses shifted to 0° (front) or 180° (rear). Some front–back confusion occurred for every position of source angle β. As a result, the pinna occlusion effectively removed localization cues for the perception of angle β. When either of the ears was open (second and third rows), regardless of the side of the open ear, the source angles β were correctly perceived on the whole, although the variances of the responses for the source angles β from 90° to 150° were larger than those when both ears were open, and some front–back confusion occurred at β=120° and 150°.

Figures 5 and 6 show the perceived angles β for sound sources at angles α=30° and 60° on the right of the median plane. When both ears were open (top rows), the perceived angles β practically coincided with the source angles β and front–back confusions hardly occurred, although the variances of the responses were somewhat larger than those in the case of angle α=90° (Fig. 7). When the right ear on the source side was open (second rows), the angles β could be perceived with much the same accuracy as when both ears were open. When the left ear was open, that is, the right ear on the source side was occluded (third rows), few responses appeared around 90° (above) and most of them shifted to 0° (front) or 180° (rear). A considerable number of front–back confusions occurred for every position of source angle β. When both ears were occluded (bottom rows), the angles β were perceived as inaccurately as when the right ear on the source side was occluded.

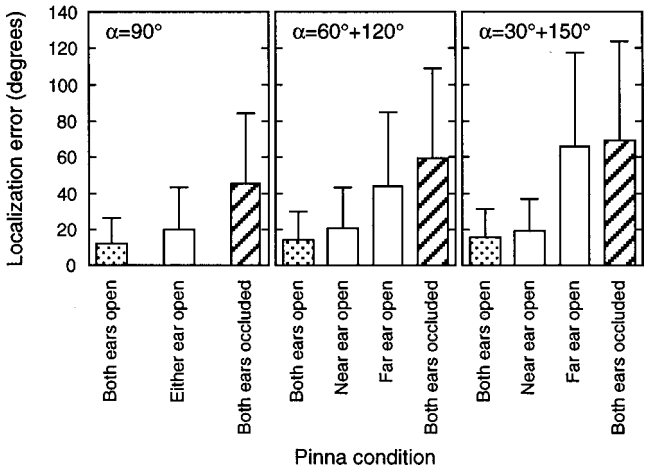


FIG. 10. Effects of pinna condition on localization error. Bars indicate standard deviations.

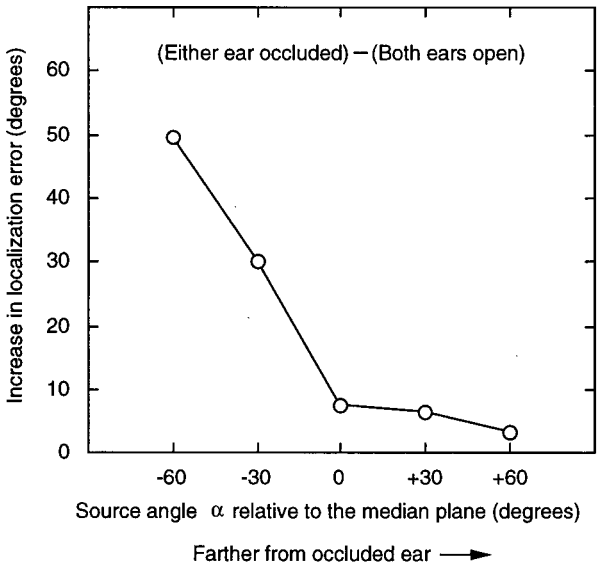


FIG. 11. Effects of either ear occlusion on localization error. Increase in error with either ear occluded relative to the both-ears-open condition.

Figures 8 and 9 show perceived angles β for sound sources at the angles $\alpha=120°$ and 150° on the left side of the median plane. The results in these cases show a similar tendency to those in the cases of $\alpha=30°$ and 60°. Here, of course, the left ear open conditions (third rows) are the cases in which the open ear is one the same side as the source, and are comparable to the right-ear-open conditions considered in the previous paragraph (second rows in Figs. 5 and 6). Similarly, the right ear open conditions (second rows) are comparable to the left-ear-open conditions (third rows in Figs. 5 and 6) for angle $\alpha=30°$ and 60°.

The effects of pinna occlusion on the perception of angle β in sagittal planes may be summarized as follows. When only the ear on the source side was occluded or both ears were occluded, few sound images appeared around $\beta=90°$ (above), and front–back confusions occurred. When only the ear on the side opposite the sound source was occluded, the angle β could be correctly perceived on the whole, while the variances of responses were larger than when both ears were open. These tendencies are very similar to those of sound localization for low-pass noises of less than 4.8 kHz in the median and the traverse planes described in previous papers (Morimoto, 1981; Morimoto and Aokata, 1984).

Figure 10 shows the effects of pinna condition on the localization error in each sagittal plane. The error is defined as the mean absolute deviation of perceived angle from source angle (Morimoto and Ando, 1980). Here, the error was obtained from the responses of the three subjects together, because individual differences among them were not remarkable, as shown in Figs. 5–9. In this figure, the results for $\alpha=60°$ and 120° are combined, and the results for $\alpha=30°$ and 150° are combined, because the results for those two sets of angles can be considered to be symmetrical about the median plane, according to the results shown in Figs. 5 and 9 and Figs. 6 and 8. The near ear and the far ear refer to the ear on the source side and the opposite ear to the source side, respectively.

In the median plane (angle $\alpha=90°$), the error increased in the order both ears open, either ear open, and both ears occluded. For $\alpha=60°$ and 120°, the error increased in the order both ears open, near ear open, far ear open, and both ears occluded. For the more lateral sagittal planes defined by angles $\alpha=30°$ and 150°, the error increased in the order both ears open, near ear open, far ear open, and both ears occluded.

Statistical tests were performed to determine whether or not a difference between mean localization errors for two pinna conditions in each sagittal plane was statistically significant. The statistical method used for the test was the z-test.[1] All three pairwise tests were done for $\alpha=90°$ and all six pairwise tests were done for $\alpha=60°$ and 120° and for $\alpha=30°$ and 150°. The sample size of each group was 420, but it was 210 for only both ears open and both ears occluded in the median plane (angle $\alpha=90°$).

For $\alpha=90°$ (median plane) and for $\alpha=60°$ and 120°, p-values obtained less than 1.0×10^{-6} for all pairs; that is, there were statistically significant differences between all of the conditions. This means that both ears contributed to the perception of angle β, even in the median plane, and suggests that the near ear contributed to the perception of angle β more than the far ear did. For $\alpha=30°$ and 150°, p-values were less than 1.0×10^{-6} for all pairs except two: for the test of far ear open versus both ears occluded the p-value was 0.2717, and for the test of near ear open versus both ears open the p-value was 0.0054. These values suggest that the far ear did not contribute measurably to the perception of angle β in these sagittal planes, although the latter value is similar than a value which is conventionally considered to be a significant level.

Figure 11 shows the effect of occluding either ear on localization error. The ordinate indicates the increase in the error when either of two open ears is occluded relative to the case with both ears open. The abscissa indicates the source angle α as a relative angle to the median plane. Zero means that a sound source is in the median plane. Plus or minus means that a sound source is on the open ear side or the occluded ear side relative to the median plane, respectively.

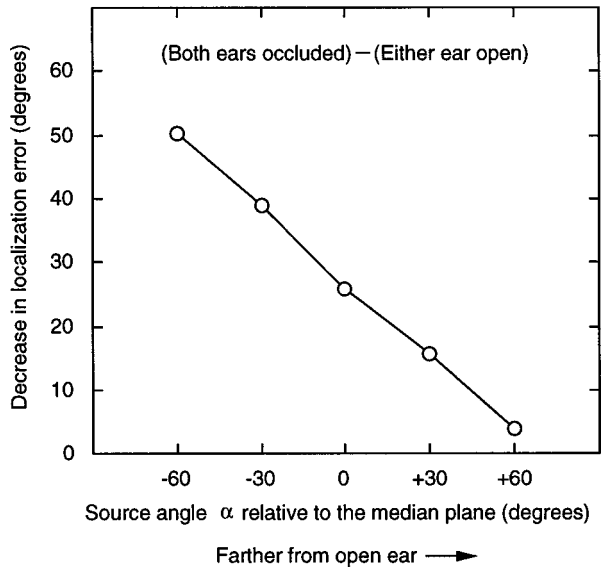


FIG. 12. Effects of either ear open on localization error. Decrease in error with either ear open relative to the both-ears-occluded condition.

In other words, the axis means how far from the occluded ear a sound source is. As a sound source shifts from the occluded ear, the increase in the error becomes small. At +60°, the increase is very small, as mentioned previously.

Figure 12 shows the effect of having either ear open on localization error. Here the ordinate indicates the decrease in the error when either of two occluded ears is open, relative to when both ears are occluded. In this case, plus or minus means that a sound source is on the occluded ear side or the open ear side, respectively. In other words, the abscissa indicates how far from the open ear a sound source is. As a sound source shifts from the open ear, the decrease in the error becomes small. At +60°, the decrease is unmeasurable, as mentioned above.

From these results, it can be inferred that the far ear no longer contributes measurably to the perception of angle β, when the shift of a sound source from the median plane exceeds 60°.

IV. DISCUSSION

Gardner (1973) investigated the effects of pinna cavities on localization ability using signals that originated in or near the horizontal plane in the anterior sector of the median plane. The result demonstrated that when the cavities of only one pinna were occluded, a considerable amount of localization ability still remained, although the localization ability was reduced compared with that achieved with no occlusion in either ear. The implication of his result was that the influence of the cavities was monaural in nature, at least insofar as localization in this sector of the median plane is concerned.

Meanwhile, the results of the present study demonstrated a significant difference in localization ability in the median plane between either ear occluded and both ears open. Moreover, the occlusion of either ear caused a significant difference in localization ability of the angle β in the saggital planes of angle α=60° and 120° which were outside of the median plane. Actually, the spectral shapes of input signals to the left and right ears from a sound source in the median plane are not identical which is also true in saggital planes outside of the median plane. Therefore, it is not reasonable to regard the perception of the angle β in only the median plane as a monaural phenomenon exclusively. In other words, both ears contribute to the perception of the vertical angle even n the median plane.

Musicant and Butler (1984) and Humanski and Butler (1988) investigated the contribution of the near and far ears to the localization of sound. The experimental results of both studies demonstrated unequivocally that the near ear contributes powerfully to localization accuracy. The results of the present paper support their findings. Musicant and Butler inferred that the far ear contributed, to some extent, to front–rear discrimination. This corresponds to the perception of the angle β in the present study, from the results in which there were fewer front–rear reversals when the far ear was open than when both ears were occluded. Such an inference was obtained by pooling across horizontal locations, which corresponded to the angle α in the present study. Careful observation of their results [see Table I in Musicant and Butler (1984)] indicates that, for locations close to the median plane, front–rear reversals when the far ear was open were fewer than when both ears were occluded, but that for locations far from the median plane, front–rear reversals when the far ear was open and when both ears were occluded were almost equal. Such a tendency coincides with the results of the present localization tests.

V. CONCLUSIONS

Localization tests of sound sources arranged throughout the upper hemisphere were performed in order to clarify the degree of contribution of each of two ears to the perception of angle β, that is, the vertical angle in sagittal planes.

The results confirm that: (1) Both ears contribute to the perception of the vertical angle even in the median plane, and (2) The near ear prominently contributes to the perception of angle β, as indicated by Humanski and Butler (1988). In addition, the results clarify that: (3) Not only the near ear but also the far ear contributes to the perception of angle β, and (4) As a sound source shifts laterally from the median plane, the contribution of the near ear increases, and, conversely, that of the far ear decreases. Furthermore, it can be inferred that: (5) When the deviation of the angle α of a sound source from the median plane is larger than a certain amount, the far ear no longer contributes measurably to the perception of angle β. This amount was 60° under the test conditions in the present study.

ACKNOWLEDGMENTS

The author thanks H. Aokata for his cooperation in localization tests and Y. Kuroki for her analysis for the experimental results.

APPENDIX

In order to obtain the amplitude and the phase parts of the head-related transfer function (HRTF), the pulse tech-

nique (Mellert *et al.*, 1974) was used adding 100 repetitions. This yielded an increase in the signal-to-noise ratio of 20 dB. The subject was seated in an anechoic chamber, and his head was stabilized. The tip of a probe-microphone (Brüel & Kjaer, Type 4170) was carefully set at the entrance to the auditory canal of the open ear or at the entrance of the passageway of the occluded ear. A pulse of 20 μs duration was fed through an amplifier into a loudspeaker. The probe-microphone output was led to a digital computer through an A-D converter (12 bit) with a sampling frequency of 50 kHz.

The HRTF, $H(\omega)$, is obtained by

$$H(\omega)=G(\omega)/F(\omega), \tag{A1}$$

where $F(\omega)$ is the Fourier transform of the waveform, $f(t)$, measured at the point corresponding to the center of the subject's head in a free field without a subject, and $G(\omega)$ is that measured at the entrance of the auditory canal of the open ear or at the entrance of the passageway of the occluded ear.

[1]The *z*-test is a test for assessing hypotheses about population means when their variances are not known and samples sizes are large ($n>100$). The test statistic z is

$$z=(\overline{x_1}-\overline{x_2})/\sqrt{s_1^2/n_1+s_2^2/n_2},$$

where $\overline{x_1}$ and $\overline{x_2}$ are the means of sample size n_1 and n_2 and s_1 and s_2 are the sample variance. z has a normal distribution (Bethea *et al.*, 1985).

Bethea, R. M., Duran, B. S., and Boullion, T. L. (**1985**). *Statistical Methods for Engineers and Scientists*, 2nd ed. (Marcel Dekker, New York), p. 193.

Blauert, J. (**1969/1970**). ''Sound localization in the median plane,'' Acustica **22**, 205–213.

Butler, R. A. (**1969**). ''Monaural and binaural localization of noise bursts vertically in the median sagittal plane,'' J. Aud. Res. **3**, 230–235.

Butler, R. A., and Belendiuk, K. (**1977**). ''Spectral cues utilized in the localization of sound in the median sagittal plane,'' J. Acoust. Soc. Am. **61**, 1264–1269.

Damaske, P., and Wagener, B. (**1969**). ''Richtungshörversuche der einen nachgebideten Kopf,'' Acustica **21**, 30–35 (in German).

Duda, R. O. (**1997**). ''Elevation dependence of the interaural transfer function,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, New York), Chap. 3, pp. 49–75.

Gardner, M. B., and Gardner, R. S. (**1973**). ''Problem of localization in the median plane: Effects of pinna cavity occlusion,'' J. Acoust. Soc. Am. **53**, 400–408.

Gardner, M. B. (**1973**). ''Some monaural and binaural facets of median plane localization,'' J. Acoust. Soc. Am. **54**, 1489–1495.

Hebrank, J., and Wright, D. (**1974a**). ''Are two ears necessary for localization of sound sources on the median plane?,'' J. Acoust. Soc. Am. **56**, 935–938.

Hebrank, J., and Wright, D. (**1974b**). ''Spectral cues used in the localization of sound sources on the median plane,'' J. Acoust. Soc. Am. **56**, 1829–1834.

Hebrank, J. (**1976**). ''Pinna disparity processing: A case of mistaken identity?,'' J. Acoust. Soc. Am. **59**, 220–221.

Hofman, P. M., van Riswick, J. G. A., and van Opstal, A. J. (**1998**). ''Relearning sound localization with new ears,'' Nat. Neurosci. **1**, 417–421.

Humanski, R. A., and Butler, R. A. (**1988**). ''The contribution of the near and far ear towards localization of sound in the sagittal plane,'' J. Acoust. Soc. Am. **83**, 2300–2310.

Kurosawa, A., Takagi, T., and Yamaguchi, Z. (**1982**). ''On transfer function of human ear and auditory localization,'' J. Acoust. Soc. Jpn. **38**, 145–151 (in Japanese).

Mellert, V., Siebrasse, K. F., and Mehrgardt, S. (**1974**). ''Determination of the transfer function of the external ear by an impulse response measurement,'' J. Acoust. Soc. Am. **56**, 1913–1915.

Middlebrooks, J. C. (**1992**). ''Narrow-band sound localization related to external ear acoustics,'' J. Acoust. Soc. Am. **92**, 2607–2624.

Morimoto, M., and Ando, Y. (**1980**). ''On the simulation of sound localization,'' J. Acoust. Soc. Jpn. (E) **1**, 167–174.

Morimoto, M. (**1982**). ''A study on sound localization,'' Doctoral Dissertation, University of Tokyo, Japan (in Japanese).

Morimoto, M., and Nomachi, K. (**1982**). ''Binaural disparity cues in median plane localization,'' J. Acoust. Soc. Jpn. (E) **3**, 99–103.

Morimoto, M., and Aokata, H. (**1984**). ''Localization cues in the upper hemisphere,'' J. Acoust. Soc. Jpn. (E) **5**, 165–173.

Musicant, A. D., and Butler, R. A. (**1984**). ''The influence of pinnae-based spectral cues on sound localization,'' J. Acoust. Soc. Am. **75**, 1195–1200.

Roffler, S. K., and Butler, R. A. (**1968**) ''Factors that influence the localization of sound in the vertical plane,'' J. Acoust. Soc. Am. **43**, 1255–1259.

Searle, C. L., Braida, L. D., Cuddy, D. R., and Davis, M. F. (**1975**). ''Binaural pinna disparity: another auditory localization cue,'' J. Acoust. Soc. Am. **57**, 448–455.

Watkins, A. J. (**1978**). ''Psychoacoustic aspects of synthesized vertical locale cues,'' J. Acoust. Soc. Am. **63**, 1152–1165.

Sensitivity to brief changes of interaural time and interaural intensity

Leslie R. Bernstein, Constantine Trahiotis, Michael A. Akeroyd, and Klaus Hartung
Department of Neuroscience and Department of Surgery (Otolaryngology), University of Connecticut Health Center, Farmington, Connecticut 06030

(Received 6 July 2000; accepted for publication 12 January 2001)

The purpose of this study was to measure listeners' abilities to detect brief changes in interaural temporal disparities (ITDs) or interaural intensitive disparities (IIDs) conveyed by bursts of noise (probes) temporally and symmetrically flanked by segments of diotic or uncorrelated noise. Thresholds were measured using a four-interval, two-alternative, forced-choice adaptive task and the total duration of the bursts of noise was either 20, 40, or 100 ms. Probes were temporally centered within each burst and the durations of the probes ranged from 2 to 100 ms, depending upon the duration of the (longer) total burst of noise within which they were embedded. The results indicate that, for a given total duration of noise, there is a rapid decrease in threshold ITD or threshold IID as the duration of the probe is increased so that it occupies a larger portion of the total burst of noise. Mathematical analyses revealed that both threshold ITDs and threshold IIDs could be well accounted for by assuming that the listener processes both types of binaural cues via a single, symmetric, double-exponential temporal window. Interestingly, the shapes of the temporal windows that fit the data obtained from the human listeners resemble the shapes of the temporal windows derived by Wagner [H. Wagner, J. Comp. Physiol. A **169**, 281–289 (1991)], who studied the barn owl. The time constants and relative weightings yielded temporal window functions that heavily emphasize information occurring within the very temporal center of the window. This temporal window function was found to be generalizable in the sense that it also accounts for classic data reported by Grantham and Wightman [D. W. Grantham and F. L. Wightman, J. Acoust. Soc. Am. **63**, 511–523 (1978)] concerning sensitivity to dynamically changing interaural disparities. [DOI: 10.1121/1.1354203]

PACS numbers: 43.66.Pn, 43.66.Mk, 43.66.Ba [SPB]

I. INTRODUCTION

In 1991, Wagner published behavioral data concerning the ability of barn owls to process brief changes in interaural time delay (ITD). Wagner's stimuli consisted of broadband noises that were interaurally uncorrelated save for a short portion termed the ''probe.'' The probe contained a fixed value of ITD that favored either the left or the right ear and the barn owl was trained to make a ''head turn'' to indicate the leading side. Wagner measured discrimination of ITD as a function of the duration of the probe and the duration of the temporally flanking interaurally uncorrelated noise. He reported that the barn owl was able to lateralize a very brief change in ITD (e.g., an ITD occurring for 10 ms or so that was temporally centered within two bursts of interaurally uncorrelated noise that were each 45 ms long). A quantitative analysis revealed that the ''temporal window'' that permitted a fit to the data of the barn owl was composed of *two* time constants: one on the order of 3–5 ms that described the central portion of the window and one of about 30–50 ms that determined the ''skirts'' of the window.

The owl's ability to discriminate such brief changes in ITD appears to be inconsistent with several studies that characterized the human's binaural system as being ''sluggish'' (e.g., Grantham and Wightman, 1978, 1979; Grantham, 1982, 1984). Such studies have routinely reported ''time constants'' of binaural processing ranging from 50 to 250 ms or so, depending on the experiment and on the individual listener. It seemed useful and important to investigate whether human listeners would also be able to detect very brief changes in ITD when those ITDs were presented in stimulus conditions similar to those utilized by Wagner (1991). To our knowledge, experiments similar to the one reported by Wagner have never been conducted with humans. Consequently, we measured human listeners' sensitivities to brief interaural disparities temporally embedded within longer bursts of noise.

It will be shown that the data from human listeners, like those obtained from the barn owl, indicate rapid processing of dynamic changes of ITDs. Furthermore, not unlike the findings for the barn owl, the new data are well described by a temporal window composed of two time constants. As was found for the barn owl, a short time constant describes the central portion of the window and a longer time constant describes the skirts of the window. In addition, we found that only one temporal window function was required to account both for sensitivity to brief changes of ITD and brief changes of interaural intensitive disparity (IID). Finally, it will be shown that the temporal window obtained using the paradigm employed here also accounts for Grantham and Wightman's (1978) classic data concerning listeners' abilities to ''track'' dynamically changing ITDs.

II. EXPERIMENT

A. Procedure

In separate experiments, the listener's task was to detect a change in either ITD or IID imposed during a segment of a

burst of noise. The segment containing the ITD or IID to be detected (i.e., the probe) was temporally centered within a longer burst of noise that was either diotic or interaurally uncorrelated. The durations of the probes and of the surrounding portions of noise were similar to those employed by Wagner (1991). The total duration of the stimuli (probe plus surrounding noise) was either 20, 40, or 100 ms. The duration of the probes was either 2, 4, 8, 16, 32, or 64 ms, depending on the total duration. Data were also collected for conditions in which the duration of the probe was equal to the total duration of the stimulus (i.e., durations of 20, 40, or 100 ms).

All stimuli were generated digitally with a sampling rate of 20 kHz via a TDT array processor and PowerDac. The noises were bandlimited between 100 and 3000 Hz and independent tokens of noise were used, respectively, for the leading, probe, and trailing segments of each total burst of noise. Each segment was constructed using inverse FFTs that yielded 8192 points in the time domain. The segments were truncated to the durations appropriate for the condition to be tested. The beginning and ending of each total burst of noise was shaped with a 1-ms $\cos^2$ ramp and the transition between surrounding noises and probes was essentially instantaneous. Stimuli were low-pass filtered at 8.5 kHz (TDT FLT2) and presented via TDH-39 earphones (mounted in MX/41-AR cushions) to listeners seated in single-walled, IAC booths. The spectrum level of each noise was equivalent to 50 dB SPL.

The stimuli were presented in a four-interval, two-cue, two-alternative temporal forced-choice adaptive task. Each trial consisted of a warning interval (500 ms) and four observation intervals separated by 450 ms. The observation intervals were marked by a visual display on a computer monitor. In the first and fourth ''cuing'' intervals, the portion of the total burst of noise that defined the probe was presented diotically. The interaural disparity to be detected (either ITD or IID) was imposed on the probe in either the second or the third interval with equal *a priori* probability. The interaural conditions in the remaining interval were like those in the cuing intervals. Feedback was provided for approximately 400 ms after the listener responded.

In order to ensure that listeners could not identify the interval containing the interaural disparity on the basis of any stimulus artifacts, independent noises were concatenated in all four intervals of the psychophysical task. Further, when the listeners' task was to detect IIDs, each nonsignal interval contained either increments at both ears or decrements at both ears (with equal *a priori* probability) that were equal in magnitude to the symmetric increments/decrements composing the IID in the signal interval. These increments and decrements were made at identical times and for identical durations as the symmetric increments/decrements produced in the signal interval. This was done in order to prevent listeners from solving the task on the basis of energy-related cues.

The ITD or IID was varied adaptively in order to estimate 70.7% correct (Levitt, 1971). When threshold ITDs were measured, the initial step size for the adaptive track corresponded to a factor of 1.584 (2 dB) and was reduced to a factor of 1.122 (0.5 dB) after two reversals. When threshold IIDs were measured, the initial and final step sizes depended upon the magnitude of the IID. For IIDs greater than 3.0 dB, the initial and final step sizes were 1.0 and 0.5 dB, respectively. For IIDs less than or equal to 3.0 dB, the initial and final step sizes were 0.5 and 0.2 dB, respectively. For both the measurement of threshold ITDs and threshold IIDs, a run was terminated after 12 reversals and threshold was defined as the average level of the signal across the last ten reversals.

Three young adults, two females and one male who had served in previous experiments and who had no evidence or history of hearing loss, served as listeners. Because they were experienced, only the measurement of a few threshold ITDs when the surrounding noise was diotic was sufficient for them to become familiar with the task. Formal testing began with the measurement of threshold ITDs when the surrounding noise was diotic. Three estimates of threshold were obtained for each combination of burst duration and probe duration. The values of burst duration and probe duration were chosen randomly from the set of values described above. Once the 21 combinations of probe duration and total duration were tested, the process was repeated with a new random order of combinations of burst and probe durations with the surrounding noise now being interaurally uncorrelated. Note that this procedure resulted in thresholds for the conditions in which the duration of the probe was equal to the total duration of the stimulus (i.e., durations of 20, 40, or 100 ms) being measured twice, once in the context of the conditions in which the surrounding noise was diotic and once in the context of conditions in which the surrounding noise was interaurally uncorrelated. The listeners were given substantial practice for conditions in which the surrounding noise was interaurally uncorrelated before formal collection of data began. This was done because the perceptual nature of the task was very different as compared to the conditions in which the surrounding noise was diotic. Then, three more estimates of threshold were obtained as described above in ''counterbalanced'' fashion beginning with conditions in which the surrounding noise was interaurally uncorrelated and ending with the conditions in which the surrounding noise was diotic. When the surrounding noise was interaurally uncorrelated and the duration of the probe was very short, sensitivity to ITD was very poor. In those instances, stimulus conditions were retested in order to decide whether a valid estimate of threshold ITD could be obtained. When estimates of threshold ITD could be obtained, the final estimate of threshold was based on the average of the last six measures that could be made. The stimulus conditions for which threshold ITDs could not be obtained are indicated in the figures containing the data by vertical arrows atop points plotted arbitrarily at a value of ITD of 999 μs.

The measurements of threshold IID were conducted following the measurements of threshold ITD in exactly the same manner described above. Stimulus conditions which did not produce a valid measure of threshold IID are indicated in the figures containing the data by vertical arrows atop points plotted arbitrarily at a value of IID of 20 dB.

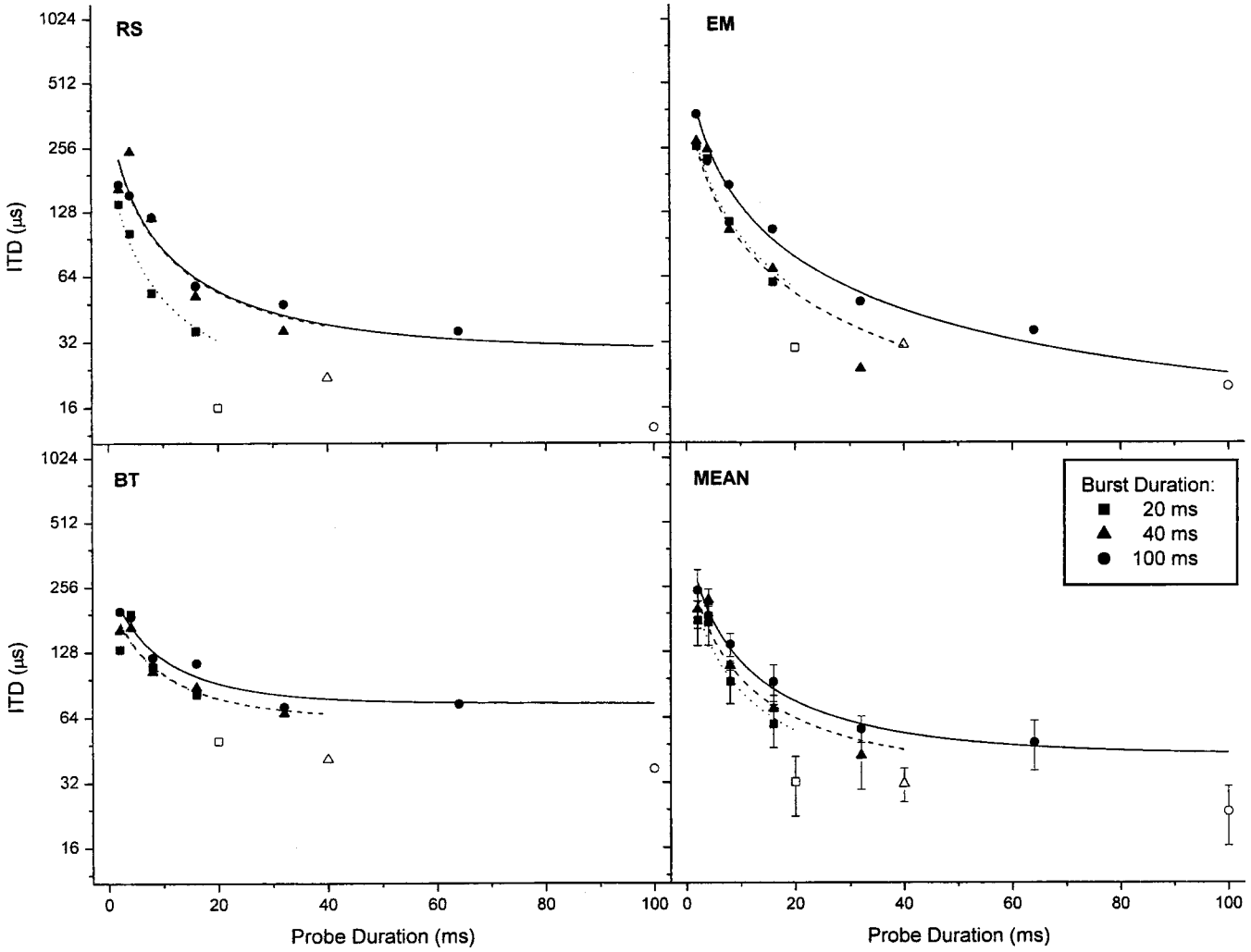


FIG. 1. Data obtained with probes surrounded by diotic noise. Threshold ITD (μs) is plotted as a function of the duration of the probe containing the ITD. The four panels display the data for the three listeners and the mean, respectively. The parameter within each plot is the total duration of the noise within which the probe was embedded. The lines passing through the data points represent predictions based on a symmetric, double-exponential window (see the text). Open symbols represent conditions in which the probe occupied the entire duration of the burst.

B. Results and discussion

1. Diotic noise surrounding the probe

Figure 1 contains the data obtained for conditions in which threshold ITDs were measured for probes surrounded temporally by diotic noise (closed symbols) and the data obtained when probes were presented in isolation (open symbols). Three of the panels display the data obtained from the individual listeners RS, EM, and BT, respectively. The lower right panel contains the means of the threshold ITDs obtained across the three listeners. The error bars represent ± 1 standard error of the mean. In each panel, threshold ITD is plotted as a function of the duration of the probe. The ordinate is scaled logarithmically because the psychometric function is approximately linear on log-ITD, as demonstrated by Yost *et al.* (1971) and as recently discussed by Saberi (1995) in the context of adaptive psychophysical procedures. The parameter in each plot is the total duration of the burst of noise in each interval. The lines passing through the closed symbols represent theoretical fits to the data and will be discussed below. Because the pattern of the data across listeners is similar, the discussion will focus on the means.

Beginning with the threshold ITDs obtained when total duration was 100 ms (circles), note that thresholds declined as the duration of the probe was increased from 2 to 100 ms. That is, threshold ITDs decreased as the portion of the total burst containing the ITD increased. The same trend occurred when the total duration was either 40 ms (triangles) or 20 ms (squares). The vertical displacements of the data indicate that, in general, threshold ITDs decreased for a constant duration of the probe as the total duration of the burst was decreased.

The thresholds obtained when the probe was presented without surrounding noise (open symbols) appear to represent ''discontinuities'' in otherwise smooth declines of threshold with increases in the duration of the probe. That is, beginning with a probe in isolation, replacing even a small portion of it with diotic surrounding noise results in a relatively large elevation in threshold. Thereafter, as the duration of the probe is decreased and more of the total duration of the burst is occupied by diotic noise, thresholds increase smoothly in a manner predicted using the temporal window. Perhaps this discontinuity occurs because the inclusion of even short segments of surrounding noise produces conflict-

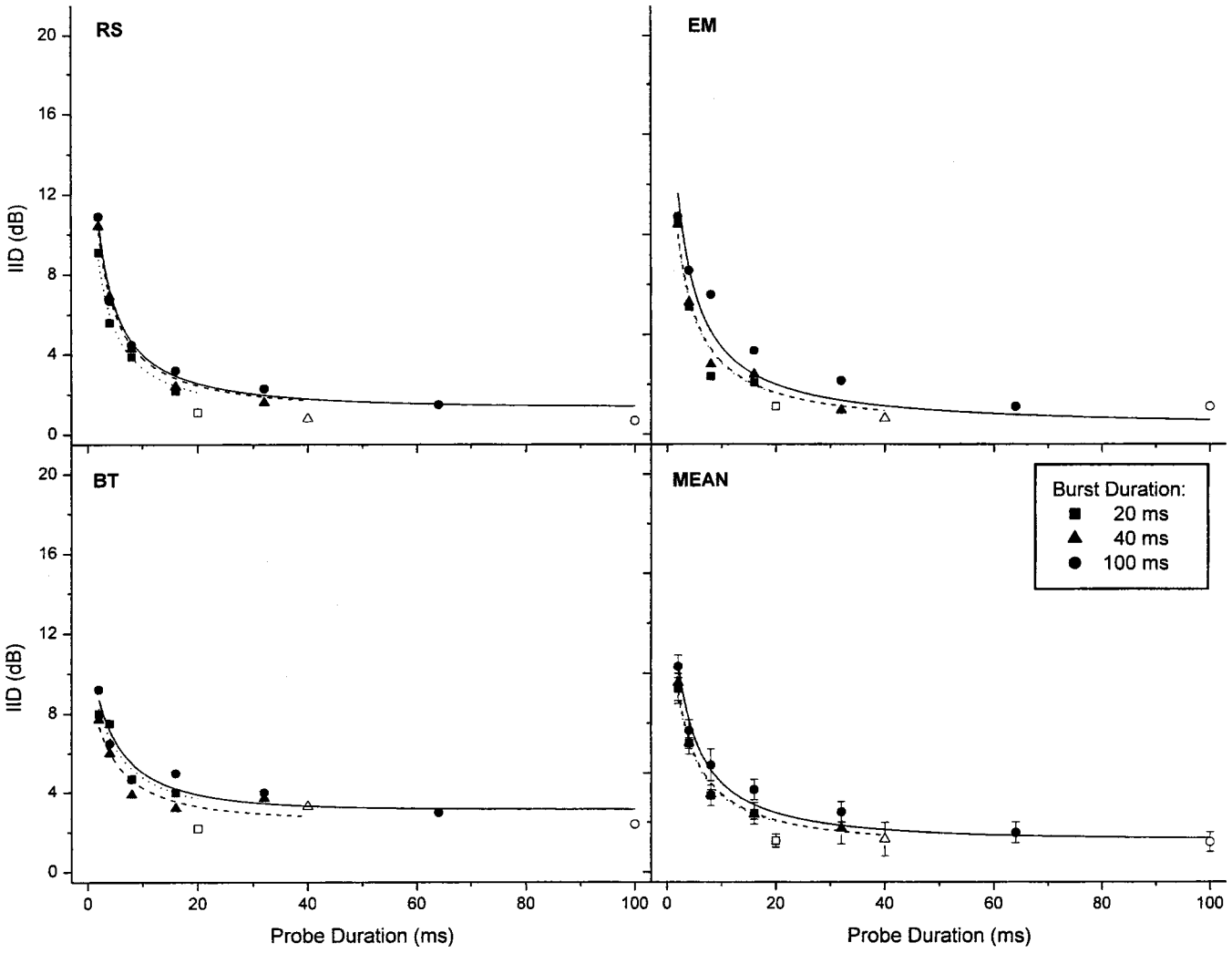


FIG. 2. Same as Fig. 1, but for threshold IIDs.

ing onset and/or offset information that degrades sensitivity to interaural cues over and above that produced by averaging binaural information over time.

The decreases in threshold ITDs with increases in probe duration do not appear to result from increases in the duration of the probe, *per se*. Note that threshold ITDs varied little with the duration of the probe when no surrounding noise was present (open symbols). Similar findings were reported by Houtgast and Plomp (1968), who observed only small changes in threshold ITDs for durations between 10 and 1000 ms when their noise ''probes'' were presented in isolation, as compared to when the probes were added to an independent noise that was lower in level. In order to explore this issue further, we conducted a supplementary experiment and found that, when the probe was presented in isolation, thresholds for all three listeners varied little for durations between 4 and 100 ms. Taken together, our data and Houtgast and Plomp's support the conclusion that threshold ITDs do not vary appreciably over a large range of duration when the information conveying the ITD is presented in ''isolation.''

The decreases in threshold ITDs observed in Fig. 1 are also not simply associated with the proportion of the duration of the total burst occupied by the probe. Replotting the data as a function of the ratio between the duration of the probe and the duration of the total burst did not result in a superimposition of the threshold ITDs. Instead, for any given constant proportion of probe and total duration, the threshold ITDs were lowest for the burst duration of 100 ms, and highest for the burst duration of 20 ms.

Figure 2 contains the data obtained when threshold IIDs were measured for the same stimulus conditions. Overall, the threshold IIDs display the same trends described earlier for the threshold ITDs. In fact, it will be seen that both sets of data are well accounted for by assuming that the stimuli are processed through a temporal window that weighs most heavily binaural information occurring at the temporal center of the stimulus while still reflecting the contribution of information stemming from the surrounding diotic noise.

2. Interaurally uncorrelated noise surrounding the probe

Figure 3 contains threshold ITDs obtained when the probe was surrounded temporally by interaurally uncorrelated noise. The reader is reminded that the portion of the total burst occupied by the probe was diotic during the three ''nonsignal'' intervals and contained an ITD during the

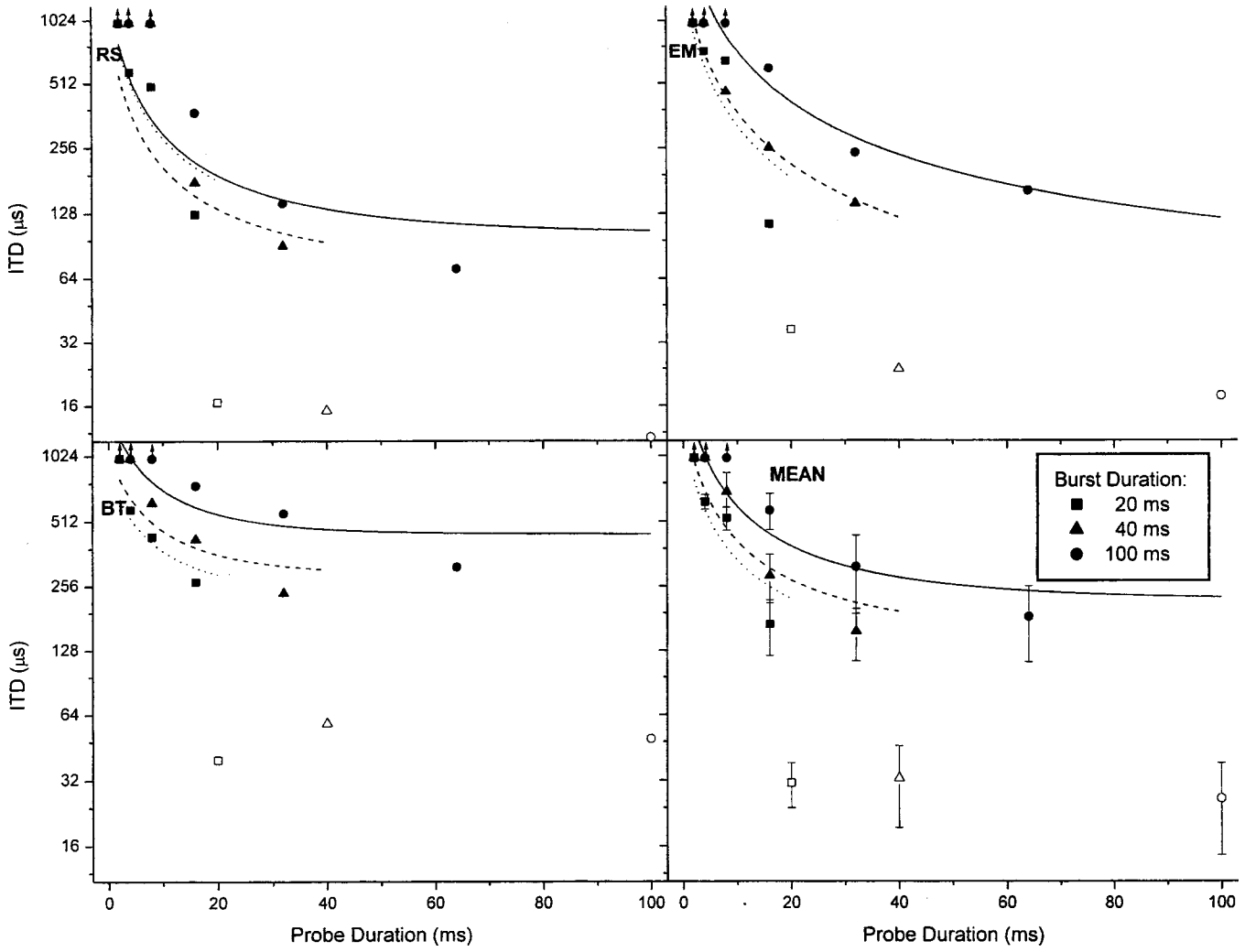


FIG. 3. Same as Fig. 1, but for conditions in which the noise that temporally surrounded the probe was interaurally uncorrelated. Arrows atop data points arbitrarily plotted at a value of 999 indicate that measures of threshold could not be obtained.

''signal'' interval. As a result, unlike the stimulus conditions in which the noises surrounding the probe were diotic, each of the four intervals composing a trial contained brief changes in the interaural disparities. Those changes occurred at the transitions between the interaurally uncorrelated surrounding noise and the probe. Consequently, the listener could not solve the task simply on the basis of whether a change in the interaural parameters occurred, as was the case when the temporally surrounding noise was diotic. Rather, the listener had to respond according to *which of the differences* was the result of the ITD. Note that, because these stimulus conditions are directly comparable to those employed by Wagner (1991), his barn owls also had to make discriminations based on changes in ITD under similar circumstances.

Keeping these issues in mind, note that the pattern of the data in Figure 3 departs in important ways from the pattern of the data observed in Figs. 1 and 2. First, threshold ITDs often could not be measured for probes of 2–8 ms. This outcome is indicated by the vertical arrows atop the symbols plotted arbitrarily at a value of 999 μs. Second, although threshold ITDs decrease with increases in the duration of the probe, the thresholds are about two to four times larger than their counterparts with diotic surrounding noise. For example, the mean data indicate that for probes having durations of 16, 32, and 64 ms embedded, respectively, in total burst durations of 20, 40, and 100 ms, threshold ITDs were, respectively, 170, 159, and 184 μs. These values are approximately three times larger than the values of threshold ITD obtained with stimuli having the same temporal properties, but with diotic noise surrounding the probe. Accordingly, the differences between thresholds obtained with probes surrounded by noise (closed symbols) and probes presented in isolation (open symbols) are much larger than those observed when the surrounding noise was diotic. These large disparities can be understood by considering that, as discussed above, adding interaurally uncorrelated noise required the listener to respond according to *which of the differences* that occurred within in each of the four intervals was the result of the ITD. In contrast, when the probe was presented in isolation, thresholds would be expected to be lower because the *only* difference that occurred was across intervals and was produced by the imposition of the ITD on the probe. In summary, there are two different aspects of the data presented in Fig. 3 that, taken independently and together, appear to indicate that sensitivity to ITDs is degraded to a

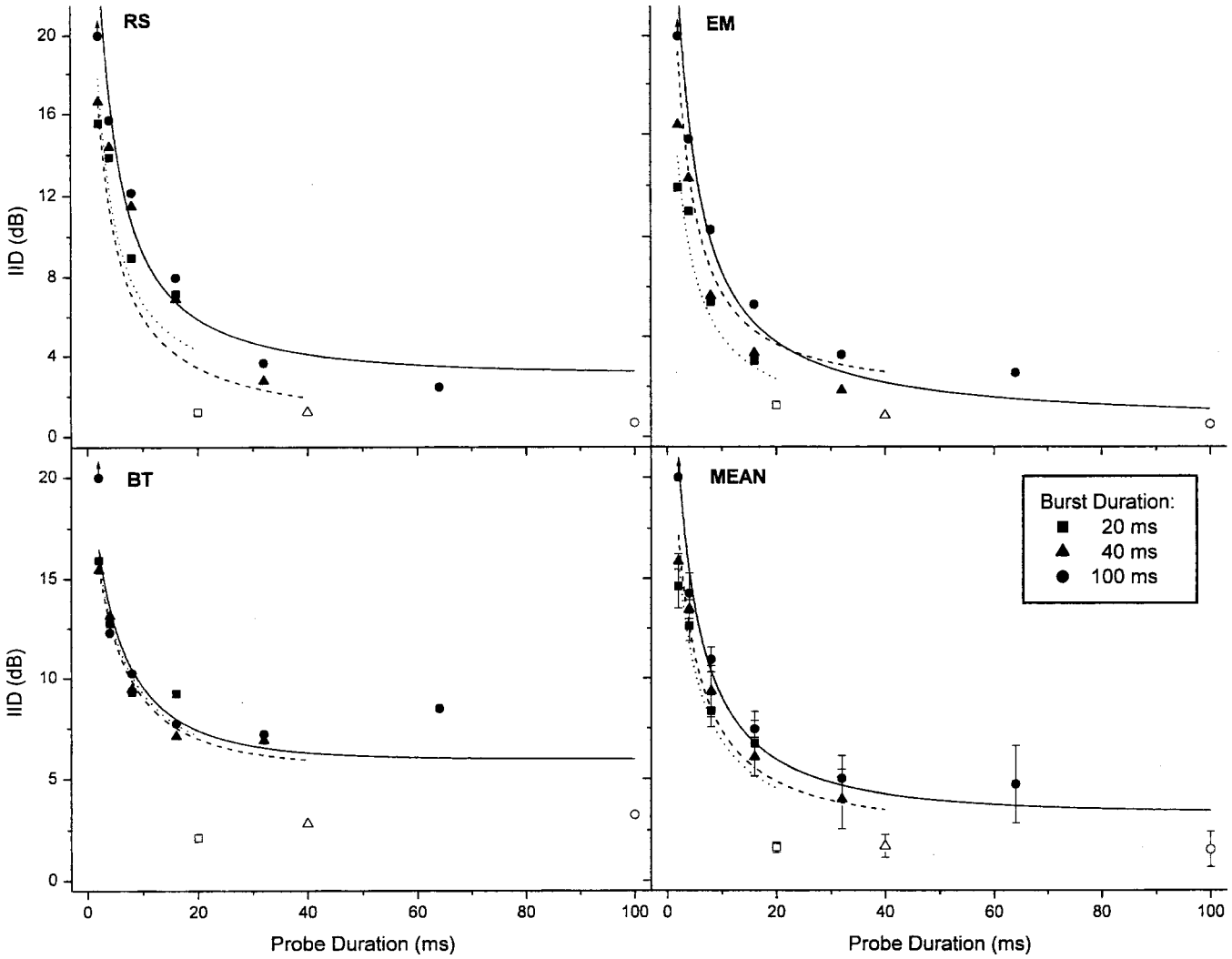


FIG. 4. Same as Fig. 2, but for conditions in which the noise that temporally surrounded the probe was interaurally uncorrelated. Arrows atop data points arbitrarily plotted at a value of 20 indicate that measures of threshold could not be obtained.

much greater degree when the noise that surrounds the probe is interaurally uncorrelated as compared to when it is diotic.

Figure 4 contains threshold IIDs measured with probes surrounded temporally by interaurally uncorrelated noise. Note that, for all three listeners, threshold IIDs could not be measured for probes of 2 ms when the total burst duration was 100 ms. This outcome is indicated by the vertical arrows atop the symbols plotted arbitrarily at a value of 20 dB. Once again, as compared to when the surrounding noise was diotic, employing surrounding noise that was interaurally uncorrelated produced an overall elevation in thresholds resulting in a much greater disparity between thresholds obtained when the probes were presented in isolation and thresholds obtained when the probes were surrounded by the noise. The greater disparity may result from the fact that, in these conditions, the listener had to respond according to *which of the differences* that occurred within each interval was the result of the IID.

The data obtained when the surrounding noise was uncorrelated, like the data obtained when the surrounding noise was diotic, are not simply associated with the proportion of the duration of the total burst occupied by the probe. When the ITD or IID thresholds in Figs. 3 and 4 were replotted as a function of the ratio between the duration of the probe and the duration of the total burst, it was again found that thresholds were lowest for the data obtained with the burst duration of 100 ms and highest for the data obtained with the burst duration of 20 ms. Nor are the data in Figs. 3 and 4 consistent with the idea that thresholds are determined by the duration of the probe, *per se*. Thresholds obtained with surrounding noise that was interaurally uncorrelated, like their counterparts with surrounding noise that was diotic, generally increased for a constant duration of the probe as the total duration of the burst was increased. As will be shown, the thresholds in Figs. 3 and 4 can be accounted for by the same temporal window functions that fit the thresholds obtained when the surrounding noise was diotic.

C. "Temporal window" analyses of the data

We begin by considering the data obtained when the surrounding noise was diotic. Our analysis is consistent with those performed by other investigators (e.g., Kollmeier and Gilkey, 1990; Akeroyd and Summerfield, 1998, 1999) in that the working assumption is that the listener's decisions are based on interaural information that has been passed through

a ''temporal window.'' The functional import of the temporal window is that it integrates binaural information across time, thereby ''smearing over'' dynamic changes of the interaural information.

The general approach was to apply a temporal window *to the values of the interaural disparities* carried by the probe and the surrounding noise, respectively. In the case of ITDs, the temporal window integrates values of ITD of zero conveyed by the diotic surrounding noise with nonzero values of ITD conveyed by the probe. The integration results in an internal, or effective, ITD that is smaller than the external ITD carried by the probe. The consequence is that the value of external ITD must be increased so that the *internal, effective ITD* is of sufficient magnitude to support detection. Following Akeroyd and Summerfield (1999), the simplifying assumption was made that the temporal window is symmetric and is placed such that its peak coincides with the temporal midpoint of the probe. This assumption is equivalent to assuming that the same symmetric temporal window is convolved in time with the stimulus and that the listener's decisions are determined by the maximum of the processed output.

The steps carried out to arrive at predicted thresholds were as follows. First, for a given total burst duration, the integral of the particular window function to be utilized was normalized to be 1.0 (see Kollmeier and Gilkey, 1990). Next, the integral was computed for the portion of the temporal window function occupied by the probe. That value represents the relative weighting applied to the information in the probe via the window function. If one begins with an effective, internal ITD (or IID) required for detection, then, by dividing that value by the weighting given the probe, one will obtain an estimate of the measured threshold ITD (or IID) for that condition. That is, the outcome of this process is the value of external ITD (or IID) in the probe required to overcome the ''dilution'' produced within the temporal window by the surrounding diotic noise.

We sought to determine the best-fitting parameters of a particular window function for the data obtained from each listener and for the mean data in Figs. 1 and 2. The thresholds obtained when the probe was presented without surrounding noise (open symbols) were excluded from the fitting procedure. As discussed above, those thresholds appear to represent ''discontinuities'' in otherwise smooth declines of threshold with increases in the duration of the probe. The fitting procedure was performed iteratively so as to minimize, in a least-squares sense, the errors of prediction. It is important to note that the data obtained with all three total durations were fit simultaneously. The parameters that were free to vary were those that defined the particular type of window function being applied and the ''criterion'' amount of ITD (or IID) that was necessary to reach behavioral threshold for each burst duration, respectively. These criterion values of ITD or IID represent overall sensitivity for each binaural cue within each burst duration condition. The fits were accomplished using the MATLAB Optimization Toolbox function, ''lsqnonlin,'' a routine specifically designed to perform nonlinear, least-squares minimization. When threshold ITDs were fit, the least-square errors were computed using the logs of the values of ITD as the dependent variable. When threshold IIDs were fit, the least-square errors were computed using the values of IID in dB as the dependent variable.

Using these procedures, we attempted to fit the data using a variety of temporal window functions including symmetric Gaussian, symmetric single- and double-exponential, asymmetric single exponential, and symmetric single- and double-rounded exponentials (see Moore *et al.*, 1988 and Culling and Summerfield, 1998 for the relevant equations). We chose to investigate the applicability of this set of window functions because it has been used by other investigators to fit data obtained in experiments in which dynamically changing interaural cues were utilized (Kollmeier and Gilkey, 1990; Culling and Summerfield, 1998). We found that the best fits were obtained using either the symmetric double-exponential function or the double rounded-exponential function. For those two functions the fitting procedure yielded temporal windows that were not only similar in shape, but which produced predicted thresholds that were, for practical purposes, indistinguishable. Using the mean data as the best estimate of binaural capability, those two functions were the ''best ones'' because they accounted for between 96% and 98% of the variability in both ITD and IID thresholds. The other functions fit the IID data about as well, but only accounted for between 83% and 91% of the variability in the mean ITD data. Furthermore, the parameters that best fit the data using the symmetric double-exponential window and the double-rounded exponential resulted in temporal windows that were remarkably similar in form to the temporal windows reported by Wagner (1991) for the barn owl.

Because the symmetric double-exponential window is both mathematically more fundamental and more amenable to analytic analysis as compared to the double-rounded exponential, we restrict description of the details of the fitting procedure to it.

Specifically, we used a symmetric double-exponential window of the form

$$W(t) = we^{(-t/\tau_1)} + (1-w)e^{(-t/\tau_2)}, \tag{1}$$

where τ_1 and τ_2 represent the two time constants associated with the temporal window function and w and $(1-w)$ represent the weights applied to the two time constants, respectively. Using this form of function, we fit, simultaneously, the threshold ITDs and IIDs obtained when the surrounding noise was diotic. In order to allow threshold ITDs and threshold IIDs to weigh equally in the determination of the parameters of the best-fitting window, IIDs were scaled so that the range of threshold IIDs (in dB) was approximately equal to the range of values of log-threshold ITD. Once the best-fitting window was determined, predictions of threshold were made in the original, ''unscaled'' space.

The lines passing through the data in Figs. 1 and 2 are the predicted ITD and IID thresholds, respectively. The predictions were made using one temporal window function per listener (and a separate temporal window function for the average of the data) to fit both threshold ITDs and IIDs. As discussed above, the symmetric double-exponential temporal

TABLE I. Time constants, relative weighting, and criterion values for the symmetric, double-exponential windows used to fit the data obtained from each listener and the mean. Criterion values of ITD (μs) and IID (dB) are listed separately for the 20-, 40-, and 100-ms bursts of noise for the conditions in which the noise surrounding the probe was either diotic or interaurally uncorrelated. The percentage of variance accounted for by the fits for each condition is shown in the right-most column of each section.

Window Parameters

Listener	$\tau_1(ms)$	$\tau_2(ms)$	w
RS	0.02	14.17	0.98
EM	0.03	64.21	0.97
BT	0.12	7.48	0.96
Mean	0.09	13.81	0.94

DIOTIC

	ITD				IID			
	Criteria (ms)			Variance accounted	Criteria (dB)			Variance accounted
Listener	20 ms	40 ms	100 ms	for (%)	20 ms	40 ms	100 ms	for (%)
RS	33	38	31	91	2.1	1.7	1.4	99
EM	57	30	23	96	2.0	1.2	0.7	93
BT	78	67	74	87	3.7	2.8	3.2	91
Mean	55	45	43	96	2.5	1.8	1.7	98

UNCORRELATED

	ITD				IID			
	Criteria (ms)			Variance accounted	Criteria (dB)			Variance accounted
Listener	20 ms	40 ms	100 ms	for (%)	20 ms	40 ms	100 ms	for (%)
RS	183	94	107	59	4.3	3.2	3.3	85
EM	187	123	122	83	2.9	1.9	1.4	94
BT	291	307	449	77	7.1	5.9	6.0	88
Mean	223	194	226	85	4.5	3.5	3.4	96

windows appear to yield excellent fits to both threshold ITDs and threshold IIDs. Table I summarizes the results of applying such windows to the thresholds obtained from each listener and to the means. The top portion of the table displays the time constants and weighting parameters that define the best-fit windows and the amounts of variance accounted for by the fits. The middle portion of the table contains the entries for the conditions in which the surrounding noise was diotic and will be discussed first. The bottom portion of the table contains the entries for the conditions in which the surrounding noise was interaurally uncorrelated and will be discussed later.

As indicated in the table, these windows are each described by one time constant (τ_1) that is only a fraction of a millisecond long and another (τ_2) that is much longer, ranging from 7 to 64 ms, depending on the individual listener. The values of time constants for the mean are 0.09 ms (τ_1) and 13.8 ms (τ_2). Note that τ_1 carries, by far, the greater relative weight (w), being between 0.94 and 0.98 across individual listeners and the mean. The amounts of variance accounted for are remarkably large being typically above 90%.[1] Specifically, the amount of variance accounted for in the mean data is 96% for threshold ITDs and 98% for threshold IIDs. Clearly, based on these findings, there is little reason to postulate that sensitivity to dynamic changes in ITDs and IIDs is processed differentially.

Figure 5 displays the window functions calculated on the basis of the time constants and their respective weights. Note that the temporal windows place relatively great emphasis on information within their temporal centers and far less, but, as demonstrated by the fits to the data, important emphasis on the temporally surrounding binaural information. The shapes of these temporal windows, which were derived from data obtained with diotic surrounding noise, are similar to the shapes of the temporal windows derived by Wagner (1991) and shown in his Fig. 7. The reader is reminded that Wagner used interaurally uncorrelated surrounding noise.

We wished to determine how strongly our data favor a temporal window having one very short, highly weighted time constant acting in combination with a longer time constant which carries a lesser weight. In order to do so, we recomputed the fits to our averaged data while varying, in stepwise fashion, the value of one of the time constants and allowing the fitting procedure to determine the other time constant and the relative weighting between the two time constants. It was consistently found that the fitting procedure returned both a shorter and a longer time constant. Importantly, for the generality of the nature of our temporal window, the two time constants consistently differed by about two orders of magnitude. We interpret this outcome as verification of the validity of the form of the temporal window depicted in Fig. 5 which is similar to the ones derived for the barn owl by Wagner (1991). Later, it will be seen that this

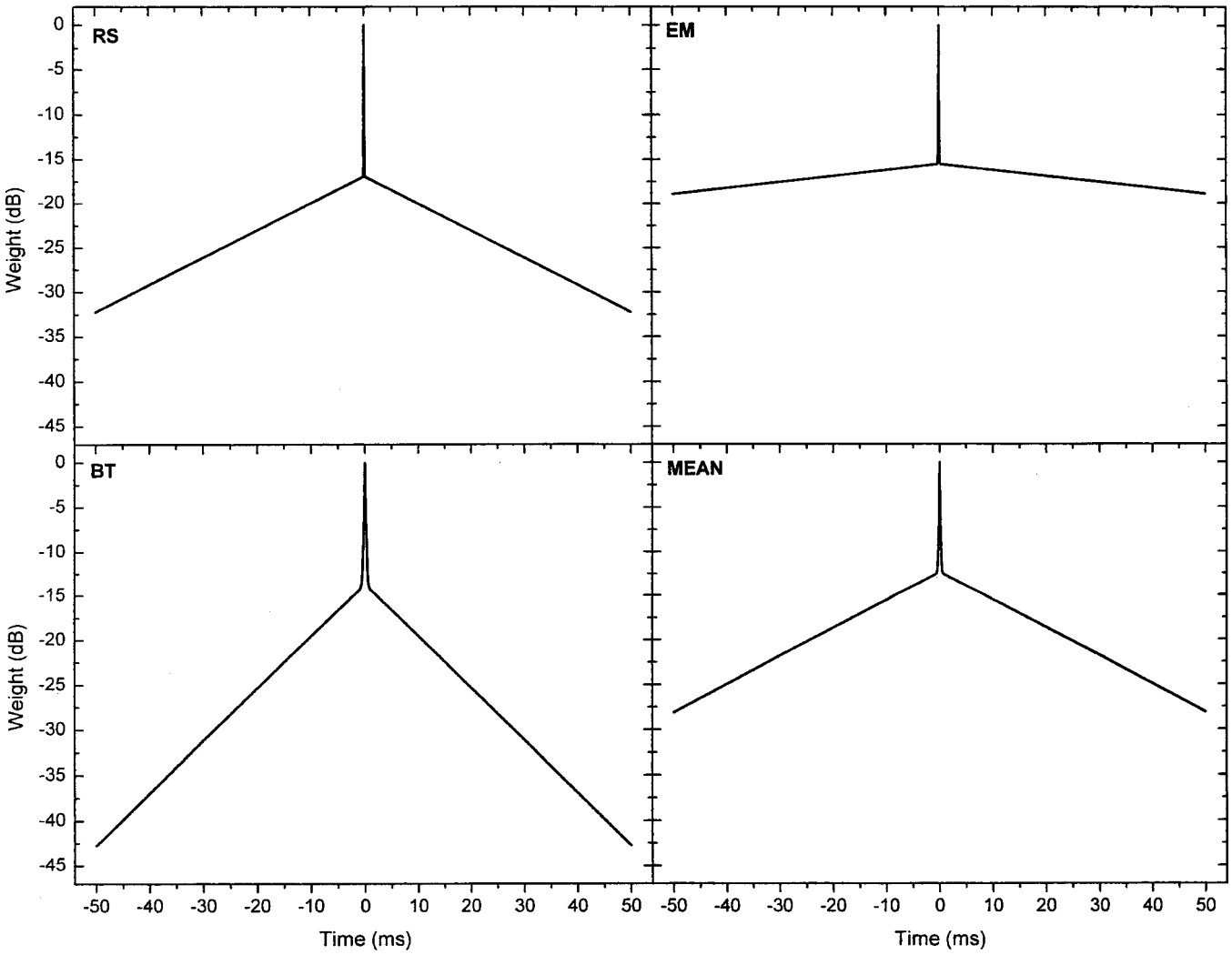


FIG. 5. Plots of the symmetric, double-exponential windows that best fit the data for the three listeners and the mean, respectively. The functions were calculated on the basis of the parameters listed in the top portion of Table I.

temporal window also accounts for the classic ''sluggishness'' data of Grantham and Wightman (1978).

We will now show that the temporal windows we derived using diotic surrounding noise also account for the thresholds we obtained when the surrounding noise was interaurally uncorrelated. In obtaining the predictions, we used the expected value of zero ITD (or IID) to represent the binaural information in the interaurally uncorrelated noise. In order to fit the data, the thresholds obtained with the interaurally uncorrelated surrounding noise were accounted for by allowing the iterative procedure to determine new criterion values of ITD (or IID) while fitting the data with the same temporal window function that best fit the data obtained when the surrounding noise was diotic. New criterion values were necessary because the thresholds obtained with interaurally uncorrelated noise were larger than their diotic counterparts. We believe the thresholds are larger for interaurally uncorrelated noise because the variability of short-term estimates of ITD (or IID) about their mean of zero is inherently large. Therefore, interaurally uncorrelated surrounding noise would be expected to result in larger *mean* differences in ITD (or IID) being required in order to produce a mean-to-sigma ratio large enough to support detection. As was the case when the surrounding noise was diotic, the thresholds obtained when the probe was presented without surrounding noise were excluded from the analysis. Also excluded were the few stimulus conditions for which thresholds could not be measured, which are represented by arrows in Figs. 3 and 4.

The lines passing through the data in Figs. 3 and 4 are the predicted thresholds resulting from using the best-fitting temporal window derived from the data obtained when the surrounding noise was diotic. The open symbols represent the thresholds obtained when the probe was presented without surrounding noise. The reader is reminded that these values were excluded from the fitting procedure, for reasons discussed above. The symmetric double-exponential temporal windows derived using diotic surrounding noise appear to account for a substantial portion of the variability in threshold ITDs and threshold IIDs obtained when the surrounding noise was interaurally uncorrelated. The bottom portion of Table I summarizes the results. The amounts of variance accounted for are somewhat smaller than with diotic noise, ranging between 59% and 94% for data obtained from individual listeners. The amount of variance accounted for in the

mean data is 85% for threshold ITDs and 96% for threshold IIDs.

These values appear to be reasonably high when it is considered that the values of the parameters defining the temporal windows were not free to vary and the mean data represent thresholds obtained from only three listeners. In our view, the fits attest to the robust nature of the analysis and provide better evidence for the validity of the temporal window functions than would be the case if we chose to fit the data obtained using interaurally uncorrelated noise with their own temporal windows. No doubt, the latter strategy would have accounted for more of the variance but at the expense of the generality of the analyses presented herein. Furthermore, the amount of variance accounted for in the data indicates the utility of characterizing the interaurally uncorrelated noise in terms of its expected values of zero ITD and zero IID.

III. GENERAL DISCUSSION

We have demonstrated that the sensitivity of the human binaural system to brief changes in ITD or IID can be accounted for by assuming a symmetric double-exponential temporal window. The shape of the temporal windows that fit our data best is similar to the shape of the temporal windows derived by Wagner (1991) for the barn owl. The time constants and relative weights associated with the windows are quite interesting in that they yield functions that heavily emphasize information occurring within the very temporal center of the window, relative to information occurring just before and after. An issue of great concern to us was whether our temporal windows which incorporate a heavily weighted, but very small time constant in combination with a longer time constant would generalize in the sense that they also could account for data obtained by Grantham and Wightman (1978). Their data are often cited as evidence that binaural processing is ''sluggish.''

In their classic study, Grantham and Wightman (1978) measured listeners' abilities to discriminate between 440-ms bursts of broadband noise that were either diotic or that contained an ITD that varied sinusoidally about a value of zero. As Grantham and Wightman noted, the sinusoidally varying ITDs produced an intracranial image that, for suitably low sinusoidal frequencies, seemed to ''move'' back and forth about the middle of the head. The main parameter of interest was the rate of sinusoidal variation, or ''modulation frequency,'' of the ITD. The dependent variable was the maximum or ''peak value'' of the ITD necessary to discriminate the ''moving'' stimulus from the one that was diotic.

The squares in Fig. 6 represent the average of the thresholds read from Grantham and Wightman's (1978) Fig. 1 for frequencies of modulation between 2.5 and 20 Hz. At these frequencies, movement of the acoustic image appears to be the cue for detection, as indicated both by the pattern of the data and by listeners' subjective reports. The line through the data points represents the predictions obtained using the temporal window function that simultaneously fit the means of our threshold ITDs and IIDs.

In order to make the predictions, our temporal window function was evaluated over a 2-s duration (from −1.0s to 1.0 s) at points determined by a sampling frequency of 20 kHz. Next, the discrete Fourier transform (MATLAB) of the temporal window was calculated. The resulting frequency response of the temporal window is shown in Fig. 7. The frequency response was used to obtain the expected ratios between the average threshold ITD Grantham and Wightman obtained at frequency of modulation of 2.5 Hz and the threshold ITDs they obtained at frequencies of modulation of 5, 10, and 20 Hz, respectively. Those ratios were calculated as $10^{d/20}$ where d represents the *difference* in dB between the amplitude of the response at 2.5 Hz and the amplitude of the response at the frequency being considered.

Finally, the predictions plotted in Fig. 6 were calculated by multiplying those ratios by the threshold ITD obtained by Grantham and Wightman at 2.5 Hz. In essence, we used the attenuation characteristic of the temporal window in the fre-

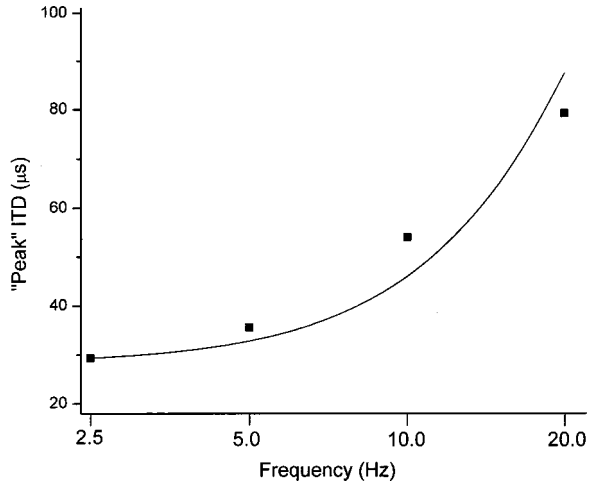


FIG. 6. The squares indicate the average ''peak'' ITDs (μs) required by Grantham and Wightman's (1978) listeners in order to discriminate a broadband diotic noise from one in which the ITD was varied sinusoidally. The thresholds are plotted as a function of the frequency at which the ITD was varied. The solid line represents predicted thresholds calculated via the magnitude spectrum of the symmetric double-exponential window that was derived from the mean data obtained in the present study (see Fig. 5 and Table I).

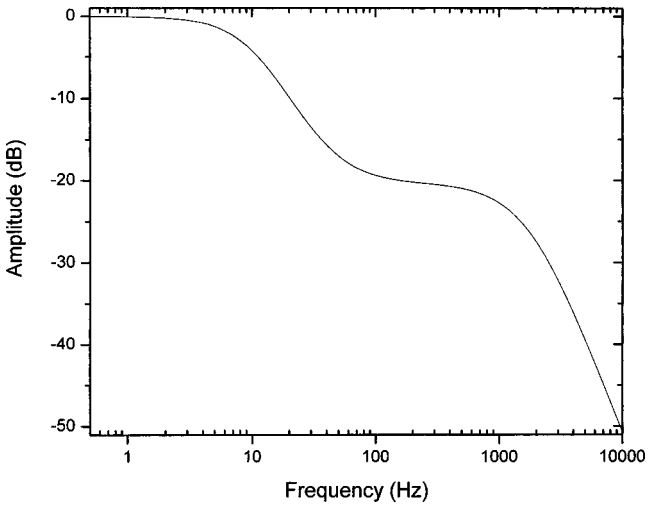


FIG. 7. The magnitude spectrum, or frequency response of the symmetric, double-exponential window that was derived from our mean data (see Fig. 5 and Table I).

quency domain to determine the increase in threshold required to maintain constant detectability as the rate of modulation of the ITD was increased. The predictions of Grantham and Wightman's data appear to be excellent and 91% of the variation in their data is accounted for by using the symmetric double-exponential window used to fit our listeners' mean data.

No doubt, the important factor contributing to the successful predictions of Grantham and Wightman's (1978) data is the initial low-pass segment of the frequency response of our temporal window (Fig. 7). Perhaps it is counterintuitive, but the initial low-pass segment is the segment that reflects the contribution of the *lower-weighted, longer* member of the pair of time constants that describe the temporal window. It is the longer time constant of 13.8 ms that accounts for, by far, the greater relative power in the frequency response. The more heavily weighted, shorter time constant of 0.09 ms is associated with a second low-pass segment that begins to decline in the region of 1 kHz. The relative power of the portion of the frequency response reflecting the shorter time constant is quite small, being at least 20 dB below the power of the initial low-pass segment.

Our longer time constant of 13.8 ms is quite close to the 15.9-ms time constant one obtains from Grantham and Wightman's (1978) average data by noting, as they did, that the ''3-dB-down'' point in their behavioral thresholds occurs at rate of modulation of 10 Hz. The reader is cautioned that Grantham and Wightman derived their time constant (100 ms) by simply taking the inverse of the 10-Hz cutoff frequency. When their time constant is recalculated by including the factor of 2π (i.e., $\tau = 1/2\pi f$) one then obtains a value of 15.9 ms, rather than the value of 100 ms which has been used to characterize the binaural system as being sluggish.

Our data and analyses suggest the intriguing possibilities that, with regard to the processing of dynamically changing ITDs and IIDs (1) one underlying temporal window function may underlie performance and (2) the binaural system may not be necessarily or intrinsically sluggish. Nevertheless, it is certainly the case that binaural time constants much larger than those measured here have been obtained in experiments concerning the detection of signals in noise on the basis of binaural cues (e.g., Grantham and Wightman, 1979; Kollmeier and Gilkey, 1990; Culling and Summerfield, 1998) and the discrimination of changes in interaural correlation (Akeroyd and Summerfield, 1999). Those larger time constants *do* appear to indicate relatively sluggish binaural performance. For example, Culling and Summerfield (1998) fitted double-exponential windows to their detection data. The average value of the shorter of the two time constants defining the window was 36 ms; the average value of the longer of the two time constants was 210 ms. Not surprisingly, these double-exponential windows do a poor job of accounting for our data. When applied to the averaged data in Figs. 1 and 2, the amounts of variance accounted for were only 50% and 67% for the threshold ITDs and threshold IIDs, respectively. It is interesting to note that the stimulus conditions in the several studies cited immediately above were such that the dynamically varying interaural disparities would have been expected to have been conveyed by a relatively narrow spectral region. In contrast, for the experiments reported here and for those reported by Grantham and Wightman (1978) and by Wagner (1991), dynamically varying interaural disparities were conveyed by a relatively broad spectral region.

We concur with Kollmeier and Gilkey (1990), who suggested that different tasks may result in different time constants because they tap different aspects of binaural temporal processing. From this point of view, there may be no single time constant that describes the binaural system's ability to process dynamically changing interaural disparities. Said differently, it appears that the binaural system may not be universally constrained by a single integrator with a fixed time constant. It may also be the case that different experiments have appeared to yield shorter or longer time constants because other factors, which are task dependent, either act in concert with the underlying temporal window and/or determine the strategies or manners by which the underlying temporal window is applied. For example, Akeroyd *et al.* (2000) recently discussed how data typically explained by invoking the ''precedence effect'' (e.g., Zurek, 1980), can be accounted for by invoking a symmetric, binaural temporal window that is used in conjunction with a function that emphasizes the binaural information in the very onset of the stimulus (Houtgast and Aoki, 1994). The important point is that both types of weighing functions were needed to account for the data. In addition, Pollack (1978) obtained data that indicate that listeners can process very rapid (up to 500 Hz or so) changes in interaural phase conveyed within trains of short pulses. As discussed by Grantham (1982), however, the magnitude of the cues present in Pollack's experiment vitiates a simple accounting for the data in terms of time constants.

This is not to say that all binaural experiments employing dynamically changing interaural cues are amenable to explanations based on temporal windows. Notably, $\mathrm{NoS}\pi$ thresholds *decrease* when ''forward fringes'' of diotic masking noise are added to coincident bursts of signal and noise, as compared to when the coincident bursts of signal and masker are presented in isolation (e.g., McFadden, 1966; Robinson and Trahiotis, 1972; Yost, 1985). It is difficult to reconcile such an improvement in performance with the notion that binaural processing is limited by the averaging that would occur within a temporal window. It would appear that such averaging could only degrade performance.

In summary, we have presented data concerning listeners' abilities to detect brief changes in ITD or IID conveyed by bursts of noise (probes) temporally embedded within longer segments of diotic or uncorrelated noise. Threshold ITDs or IIDs decreased as the probes occupied larger and larger portions of the total bursts of noise. We found that both threshold ITDs and threshold IIDs obtained for each listener and the mean data, respectively, could be predicted using one symmetric, double-exponential temporal window. In each case, the same temporal window fit the data obtained both when the temporally surrounding noise was diotic and when it was interaurally uncorrelated. The shapes of the temporal windows that fit our data resembled those derived by Wagner (1991), who employed a similar psychophysical task

while studying binaural processing in the barn owl. Finally, we found that the temporal window function that accounted for our mean data also accounted for the mean of a subset of the data of Grantham and Wightman (1978).

ACKNOWLEDGMENTS

Supported by research Grant Numbers NIH DC-04073 and NIH DC-00234 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health. The authors wish to thank Dr. Wes Grantham, one anonymous reviewer, and Dr. Sid Bacon, the Associate Editor, who provided comments that helped strengthen the manuscript.

[1]The formula used to compute the percentage of the variance for which our predicted values of threshold accounted was $100\times(1-[\Sigma(O_i-P_i)^2]/[\Sigma(O_i-O)^2])$, where O_i and P_i represent individual observed and predicted values of threshold, respectively, and O represents the mean of the observed values of threshold.

Akeroyd, M. A., Bernstein, L. R., and Trahiotis, C. (**2000**). ''Discrimination of brief interaural temporal disparities embedded within diotic bursts of broadband noise,'' J. Acoust. Soc. Am. **107**, 2822.

Akeroyd, M. A., and Summerfield, A. Q. (**1999**). ''A binaural analog of gap detection,'' J. Acoust. Soc. Am. **105**, 2807–2820.

Akeroyd, M. A., and Summerfield, Q. (**1998**). ''Predictions of signal thresholds in frozen-noise masker using monaural and binaural windows,'' in *Psychological and Physiological Advances in Hearing*, edited by A. Palmer, A. Rees, Q. Summerfield, and R. Meddis (Whurr, London, 1998).

Culling, J. F., and Summerfield, Q. (**1998**). ''Measurements of the binaural temporal window using a detection task,'' J. Acoust. Soc. Am. **103**, 3540–3553.

Grantham, D. W. (**1982**). ''Detectability of time-varying interaural correlation in narrow-band noise stimuli,'' J. Acoust. Soc. Am. **72**, 1178–1184.

Grantham, D. W. (**1984**). ''Discrimination of dynamic interaural intensity differences,'' J. Acoust. Soc. Am. **76**, 71–76.

Grantham, D. W., and Wightman, F. L. (**1978**). ''Detectability of varying interaural temporal differences,'' J. Acoust. Soc. Am. **63**, 511–523.

Grantham, D. W., and Wightman, F. L. (**1979**). ''Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation,'' J. Acoust. Soc. Am. **65**, 1509–1517.

Houtgast, T., and Aoki, S. (**1994**). ''Stimulus-onset dominance in the perception of binaural information,'' Hear. Res. **72**, 29–36.

Houtgast, T., and Plomp, R. (**1968**). ''Lateralization threshold of a signal in noise. J. Acoust. Soc. Am. **44**, 807–812.

Kollmeier, B., and Gilkey, R. H. (**1990**). ''Binaural forward and backward masking: Evidence for sluggishness in binaural detection,'' J. Acoust. Soc. Am. **87**, 1709–1719.

Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

McFadden, D. M. (**1966**). ''Masking-level differences with continuous and with burst masking noise,'' J. Acoust. Soc. Am. **40**, 1414–1419.

Moore, B. C. J., Glasberg, B. R., Plack, C. J., and Biswas, A. K. (**1988**). ''The shape of the ear's temporal window,'' J. Acoust. Soc. Am. **83**, 1102–1116.

Pollack, I. (**1978**). ''Temporal switching between binaural information sources,'' J. Acoust. Soc. Am. **63**, 550–558.

Robinson, D. E., and Trahiotis, C. (**1972**). ''Effects of signal duration and masker duration on detectability under diotic and dichotic listening conditions,'' Percept. Psychophys. **12**, 333–334.

Saberi, K. (**1995**). ''Some considerations on the use of adaptive methods for estimating interaural delay thresholds,'' J. Acoust. Soc. Am. **98**, 1803–1806.

Wagner, Hermann (**1991**). ''A temporal window for lateralization of interaural time difference by barn owls,'' J. Comp. Physiol. A **169**, 281–289.

Yost, W. A. (**1985**). ''Prior stimulation and the masking-level difference,'' J. Acoust. Soc. Am. **78**, 901–907.

Yost, W. A., Wightman, F. L., and Green, D. M. (**1971**). ''Lateralization of filtered clicks,'' J. Acoust. Soc. Am. **50**, 1526–1530.

Zurek, P. M. (**1980**). ''The precedence effect and its possible role in the avoidance of interaural ambiguities.'' J. Acoust. Soc. Am. **67**, 952–964.

Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees

Ronald C. Scherer and Daoud Shinwari
Department of Communication Disorders, Bowling Green State University, 200 Health Center, Bowling Green, Ohio 43402

Kenneth J. De Witt
Department of Communication Disorders, Bowling Green State University, 200 Health Center, Bowling Green, Ohio 43402 and Chemical Engineering, University of Toledo, 3060 Nitschke Hall, Toledo, Ohio 43606-3390

Chao Zhang, Bogdan R. Kucinschi, and Abdollah A. Afjeh
Mechanical, Industrial, and Manufacturing Engineering, University of Toledo, 4006D Nitschke Hall, Toledo, Ohio 43606-3390

(Received 15 August 1999; accepted for publication 23 October 2000)

Human phonation does not always involve symmetric motions of the two vocal folds. Asymmetric motions can create slanted or oblique glottal angles. This study reports intraglottal pressure profiles for a Plexiglas model of the larynx with a glottis having a 10-degree divergence angle and either a symmetric orientation or an oblique angle of 15 degrees. For the oblique glottis, one side was divergent and the other convergent. The vocal fold surfaces had 14 pressure taps. The minimal glottal diameter was held constant at 0.04 cm. Results indicated that for either the symmetric or oblique case, the pressure profiles were different on the two sides of the glottis except for the symmetric geometry for a transglottal pressure of 3 cm H_2O. For the symmetric case, flow separation created lower pressures on the side where the flow stayed attached to the wall, and the largest pressure differences between the two sides of the channel were 5%–6% of the transglottal pressure. For the oblique case, pressures were lower on the divergent glottal side near the glottal entry and exit, and the cross-channel pressures at the glottis entrance differed by 27% of the transglottal pressure. The empirical pressure distributions were supported by computational results. The observed aerodynamic asymmetries could be a factor contributing to normal jitter values and differences in vocal fold phasing.
[DOI: 10.1121/1.1333420]

PACS numbers: 43.70.Aj, 43.70.Bk [AL]

I. INTRODUCTION

The vocal folds vibrate during phonation because of a number of simultaneous factors. Primary among these are the biomechanical tissue characteristics of the vocal folds (Ishizaka and Flanagan, 1972; Titze, 1974), glottal configuration (especially the degree of glottal adduction: Scherer *et al.*, 1995; Titze and Talkin, 1979), and aerodynamic air pressures surrounding the vocal folds (Ishizaka and Matsudaira, 1972). Typically, normal phonation is considered to involve symmetric activity of the two vocal folds whereby the motion of one vocal fold mirrors the motion of the other across the midline (Hirano, 1981; Svec and Schutte, 1996). This symmetric motion is reflected in schematics of the frontal (coronal) view of glottal shapes that are also symmetric across the midline (Hirano, 1981). On the other hand, some asymmetry can be seen in normal (Koike and Imaizumi, 1988; Svec *et al.*, 1999a) as well as abnormal or pathological conditions such as unilateral paralysis, polyps, arytenoid fixation, and webbing, corresponding to asymmetric conditions of tissue and configuration (von Leden *et al.*, 1960; Hirano and Bless, 1993; Svec and Schutte, 1996; Schutte *et al.*, 1998; Svec, Sram, and Schutte, 1999b). Asymmetry of motion of the two normal vocal folds may mean that they are out of phase with each other. In this situation, one vocal fold would be ahead of the other during the cycle, reaching the greatest width as well as the most medial position before the other one does.

The purpose of the present study was to examine the air pressures on the vocal fold surfaces for both a symmetric and an asymmetric glottal shape, where the glottis forms a divergent angle of 10 degrees. The 10-degree divergence angle should occur during most phonatory cycles, and is a potentially important angle due to the expected relatively large pressure rise within the duct (a duct with a 10-degree divergent angle creates an efficient diffuser to regain pressure, Kline, 1959). The obliquity of the two conditions is zero degrees for the symmetric condition and 15 degrees for the asymmetric condition, respectively.

A. Intraglottal air pressures

The focus of this report is on the air pressures within the glottis. The air pressures on the vocal fold surfaces act as the external driving forces on the vocal folds during phonation. It has long been known that the air pressures acting upon the vocal fold surfaces below and within the glottis have significant influence on the motion of the vocal folds (van den Berg *et al.*, 1957). Computer multimass models (Ishizaka and

Flanagan, 1972; Titze, 1974; Story and Titze, 1995; Liljencrants, 1996), in which the flow is treated analytically, have included aerodynamic pressure-flow-geometry equations stemming from basic fluid mechanics theory and experimentation using steady flow through rigid models. Pelorson *et al.* (1994) incorporated an important addition to the modeling in the form of an estimate of the dynamically changing location of the flow separation point in the glottis. The empirical studies with rigid models have assumed symmetric (across the midline) glottal shapes (Ishizaka and Matsudaira, 1972; Scherer *et al.*, 1983a; Scherer and Titze, 1983; Binh and Gauffin, 1983; Gauffin *et al.*, 1983; Scherer and Guo, 1990), the only apparent exception being a study of the effect of intraglottal obstructions (Scherer *et al.*, 1983b).

B. Asymmetric flow ducts

There appear to be no empirical data on the air pressures on the medial vocal fold surfaces in the nonsymmetric glottis. Asymmetric ducts studied in the fluid mechanics literature are known to have asymmetric pressures on their sidewalls and greater losses. For bent ducts, the pressures are higher on the outer wall and lower on the inner wall (Ward Smith, 1971; Parsons and Hill, 1973; So, 1976). The same may hold for the walls of the glottis, that is, that the distribution of air pressures on one side of the glottis would be different from the distribution of air pressures on the other side of the glottis for steady flow through rigid asymmetric glottal ducts. These asymmetric pressures would create different external driving forces on the two sides of the glottis. The study of the air pressures on the medial vocal folds for asymmetric glottal shapes would be important to better understand and more accurately model both normal and abnormal phonation.

C. Flow behavior within and downstream of a diverging duct

The flow through a divergent duct of 10 degrees (the divergence angle used in this study) may separate upstream in the diffuser section and, in addition, move closer to one of the two walls (Kline, 1959; Reneau *et al.*, 1967). When flow separation occurs in a symmetric diverging duct having straight walls, the pressures on the two walls may differ due to flow instability (Ashjaee and Johnston, 1980; Carlson *et al.*, 1967), resulting in a skewed flow. Also, flow from a centrally located slit into a larger rectangular duct is known to move toward one side of the downstream duct [Tsui and Wang, 1995; Cherdron *et al.*, 1978; this phenomena is also mentioned in Shadle *et al.* (1991) for an eccentric glottal orifice]. These findings suggest that the flow through the glottis of the model to be described here may be skewed to one side within the glottis, may have different pressure profiles on the two glottal sides, and may move toward one side of the downstream duct.

D. Features of the glottal geometry

The human glottis is a three-dimensional airway with two primary portions. The membranous (or anterior) glottis extends from the vocal processes of the arytenoid cartilages to the anterior commissure (Hirano *et al.*, 1988). The posterior (or cartilaginous) glottis extends from the vocal processes to the posterior mucosal wall between the two arytenoid cartilages (Hirano *et al.*, 1988). The study here will follow other empirical model studies by concentrating attention on the membranous glottis, where the normal coronal configuration during phonation alternates between converging and diverging shapes during the glottal opening phase and the closing phase, respectively (Hirano, 1981; Schönhärl, 1960). This alteration in shape change has been considered to significantly aid the intraglottal pressure variations between positive forces that help push the vocal folds out during the glottal opening phase, and negative pressures to help pull the vocal folds together during the glottal closing phase (Ishizaka and Matsudaira, 1972; Gauffin *et al.*, 1983; Titze, 1986; Scherer, 1991, 1995; Guo and Scherer, 1993). These force variations help to maintain oscillation of the vocal folds (Titze, 1988).

The membranous glottis has an anterior–posterior shape that usually is not uniform. When the folds are adducted just prior to phonation, the glottis often has a triangular shape depending upon the degree of adduction at the vocal processes (Zemlin, 1998). During phonation the glottis opens wider near the middle and tapers toward the anterior and posterior ends. Empirical rigid models used to obtain pressure profiles have had glottal shapes that do not taper anteriorly–posteriorly, most likely because of the difficulties in such modeling and the need for simplicity. The study reported here incorporates a glottis that also is uniform in the anterior–posterior dimension, and excludes the posterior glottis.

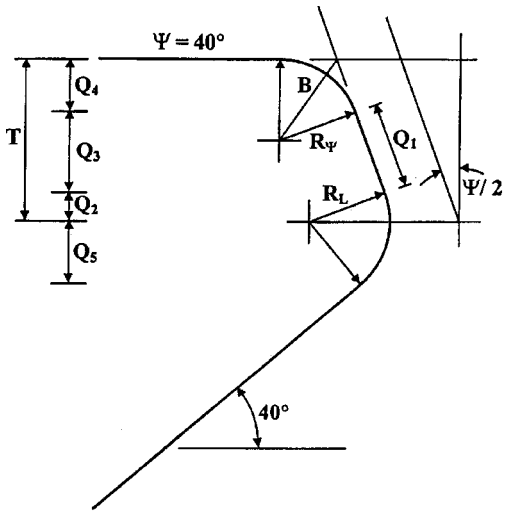
E. Glottal obliquity

The asymmetric modeling performed in this study examines a simple asymmetry that involves a change in the *direction* of the glottal duct. Two conditions are examined. The symmetric glottal duct has a divergent angle, and the direction of the center of the duct is ''on axis'' (parallel) with the tracheal or subglottal flow, that is, in an inferior to superior direction along the vertical plane following the midline of the trachea and glottis. The other condition involves an *oblique glottal duct* in which the direction along the center of the glottis is straight but slanted to the side, creating an oblique angle to the subglottal midline. This geometry could occur, for example, when the two vocal folds are vibrating out of phase with each other, as discussed above. That is, if one vocal fold is ahead of the other as it goes through its cycle, the direction of the center line of the glottal duct (its obliquity) will shift throughout the cycle. The ''instant'' of such vibration chosen for this study is when the vocal folds form a glottis that is both oblique and divergent.

II. METHODS

A. Model M5—Vocal folds and glottal angles

The model used in this research is called M5 and is a nonvibrating Plexiglas model of the simplified larynx with linear dimensions 7.5 times greater than a normal male larynx. The model uses replaceable vocal fold pieces situated in



General vocal fold surface design equations :

$R_o = 0.0987$ cm $T = 0.3$ cm $-40° \le \Psi \le 40°$

$R_\Psi = R_o / (1 - \sin (\Psi/2))$ $R_L = R_{40} = T/2$

$B = \sqrt{2}\, R_\Psi / \sqrt{1 + \sin (\Psi/2)}$

$= R_o \sec (\Psi/2) / \sqrt{(1 - \sin (\Psi/2)) / 2}$

$Q_1 = (T - R_\Psi) \sec (\Psi/2) + (R_\Psi - R_L) \tan (\Psi/2)$

$= (T - R_o - R_L \sin (\Psi/2)) \sec (\Psi/2)$

$Q_2 = R_L \sin (\Psi/2)$ $Q_3 = Q_1 \cos (\Psi/2)$

$Q_4 = R_o$ $Q_5 = R_L \sin 50°$

FIG. 1. Schematic of the design of the vocal fold pieces for model M5 with the design equations. The shape of the vocal fold shown is for a divergent-included glottal angle of 40°. Values for R_o and T are human values rather than model values.

a rectangular duct (''wind tunnel''). Figure 1 is a schematic of the general design of the vocal folds for M5. The shape of the vocal folds at the glottal entrance and exit is well defined by radii that depend upon the glottal angle. A straight section along the vocal fold wall connects the entrance and exit radii.[1] The vocal fold pieces of model M5 were milled by The University of Iowa Medical Instruments Shop. The glottal length was held constant at 1.2 cm human size (9.0 cm in M5). The length of the glottal duct (inferior to superior, or upstream to downstream) was 0.3 cm human size (2.25 cm in M5). Eighteen vocal fold pieces (nine pairs) were fabricated to provide nine symmetric glottal angles. The full set of included symmetric glottal angles are divergent 40°, 20°, 10°, and 5°; uniform 0°; and convergent 40°, 20°, 10°, and 5°. These are called ''included angles,'' indicating their measure between the two medial vocal fold walls.

B. Model M5—wind tunnel

Figures 2(a) and (b) show model M5 in detail. Figure 2(a) is a top view of the box showing the entrance to the wind tunnel, the upstream section, the two vocal folds with a diameter-determining shim, downstream section, and the exit tubing. Figure 2(b) is a three-dimensional view that also shows the large Plexiglas slab below the vocal folds. Each vocal fold is attached to the side of the wind tunnel by four screws. Figure 3 is a schematic that suggests the tubing connections for the 17 pressure taps, the pressure transducer (Validyne DP103-10) for the pressures taps, the pressure transducer (Validyne MP45-16) for the pneumotach (Hans Rudolph PNT 3719), and the vacuum source to pull the air through the model. Two voltmeters (Hewlett Packard 973A) were used to read the voltage outputs from the Validyne system during data collection.

It is noted that the human membranous glottis forms an eccentric constriction in the airway when the arytenoid cartilages are together. The air traveling up the posterior trachea will turn to then flow through the membranous glottis. This alteration of flow is not modeled in the present form of model M5. For this experiment, the duct upstream and downstream of the vocal folds had the same length as the vocal folds (1.2 cm human size). Effects of the presence of the arytenoid cartilages *per se* are absent (Alipour and Scherer, 1995).

C. Vocal fold pieces for this study

Three pairs of vocal fold pieces were used in this study. One pair corresponded to the symmetric glottis with the 10-degree divergence [Fig. 4(a)]. Each vocal fold of this pair had a divergence of 5 degrees, making the included angle 10 degrees. Because pressure taps were located only in one of the pair of vocal folds (see later in this work), there were two sets of vocal fold pairs for the study of the 15-degree oblique glottis with a divergence of 10 degrees [Fig. 4(b)]. For each pair, one side had a divergent angle of 20 degrees, and the other side had a convergent angle of 10 degrees. The glottal duct slanted toward one side for one pair, and to the other side for the other pair.

D. Pressure taps

One of each pair of the vocal folds had 14 pressure taps along its surface, as illustrated in Fig. 5. The pressure taps were made flush and perpendicular to the vocal fold surfaces. The inner diameter of each pressure tap was 0.033 cm (actual size in the model vocal folds). The minimal distance between all pairs of folds (that is, the minimal diameter of the glottis) was set at 0.04 cm human size (0.3 cm in M5).[2] The smallest ratio between the glottal duct diameter and the pressure tap diameter at the same cross section was 9.01 to 1. In addition to the pressure taps on the vocal folds, there was a pressure tap upstream in the side of the rectangular wind tunnel (the location of the subglottal reference pressure), and two downstream in the side of the rectangular wind tunnel (see Fig. 2).

The pressure taps on the surface of the vocal fold pieces were placed near the center in the inferior to superior direction (Fig. 5). They were staggered on either side of the centerline by 0.06 cm human size (0.45 cm in the model), which was 5% of the glottal length on either side of the centerline. The staggering was created so that any possible disturbance in the airflow caused by the presence of a pressure tap would not affect the next downstream pressure reading. Despite the staggering, the pressures were expected to be the same as they would be along the center of the vocal fold surface because of the uniform upstream and downstream ducts and the unchanging anterior–posterior glottal shaping. Figures

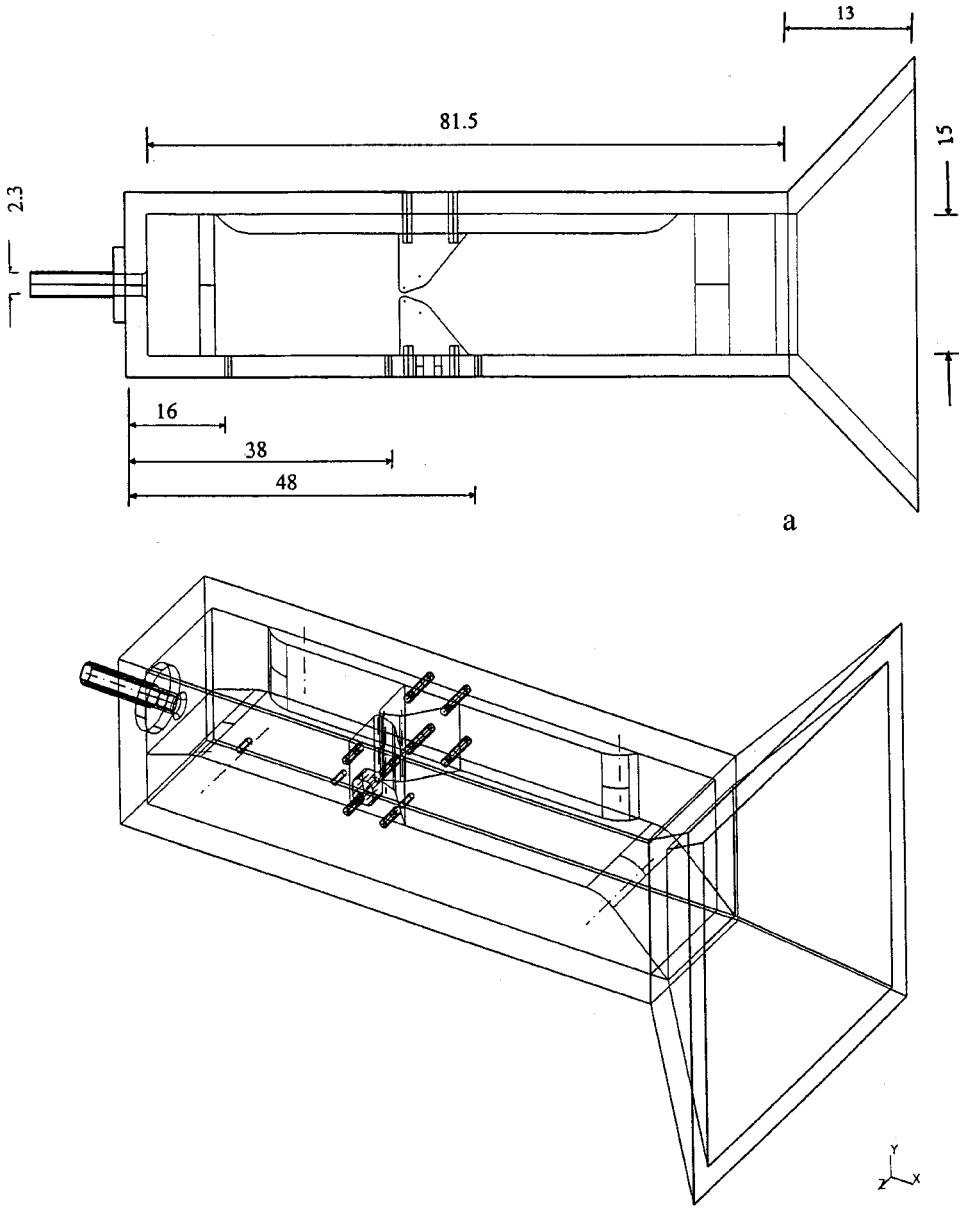


FIG. 2. (a) Schematic of the Plexiglas model M5 seen from above. The vocal folds are shown with a symmetric converging shape. Note the long shim (between one vocal fold piece and the lateral wall) that defines the closeness of the two vocal folds. All measures are actual model dimensions in cm. (b) Schematic in three dimensions. The opening for the multiple tubings for the pressure taps from the near-side vocal fold piece can be seen. This view also shows the large Plexiglas slab below the vocal folds.

4(a) and (b) show the relative axial locations of the pressure taps for the oblique and symmetric geometries, respectively.

E. Pressure and flow measures

The pressure taps led via silicon tubing to a pressure scanner (Scanivalve Corp. 24-port pressure scanner). The scanner was used to obtain pressure drops between the upstream pressure tap in the tunnel wall and the other pressure taps. The pneumotach was used to obtain the volume flow. The Validyne pressure transducers were calibrated using various methods, including a micromanometer (Dwyer Microtector Model 1430) and a wall *u*-tube manometer (Dwyer Flex-Tube). The pneumotach was calibrated with inline flowmeters (Gilmont Gf-1400, GF-1500, and GF-1660).

Four transglottal pressures, 3, 5, 10, and 15 cm H_2O (human size values), were used in this study. Because the Plexiglas model is larger than human size by a factor of 7.5, the corresponding pressures in the model were less than human size by a factor of $(7.5)^{-2}$, as determined by similitude analysis (Streeter and Wylie, 1975).[3] Also, the flows in the model were 7.5 times larger than human size values, again determined by similitude analysis.[3] Uncertainty analysis led to the pressure measurement accuracy estimate of ± 0.064 cm H_2O human size and a flow measurement accuracy of ± 4.24 cm^3/s human size.[4]

F. Glottal diameter

Each of the vocal fold pieces was held against the sidewalls in the model by four screws (see Fig. 2). A long Plexiglas shim was placed on the nonpressure taps side of the duct between the vocal fold piece and the sidewall (Fig. 2). The shim was tapered at its two ends near the entrance and exit of the wind tunnel, and was the same height in the tunnel as the length of the vocal folds. Feeler gages supplemented the shim to create the desired minimal glottal diameter for each condition studied.

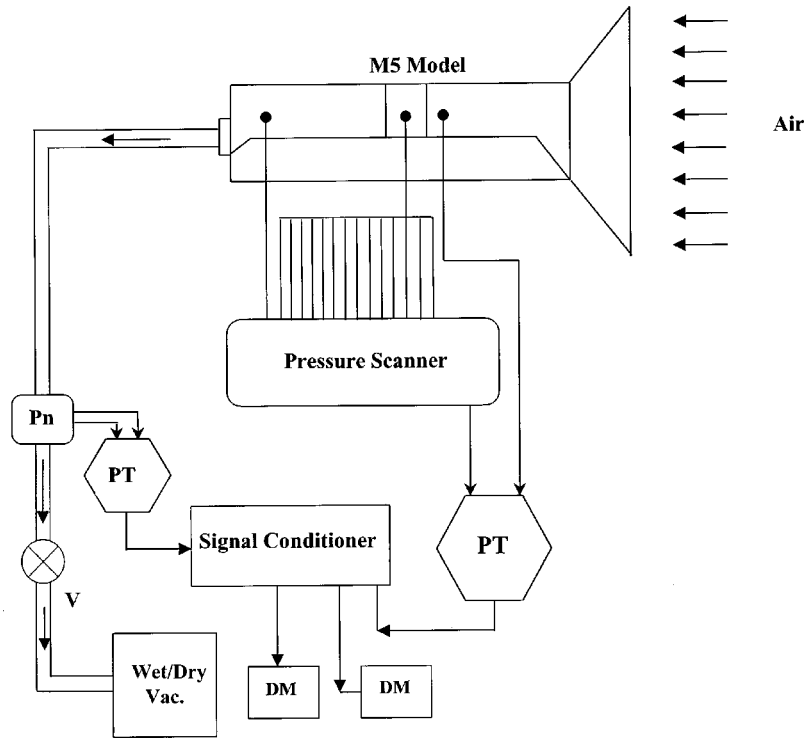


FIG. 3. Schematic of the experimental setup. Air is pulled through the system. Pressure transducers (PTs) are shown connected to the pressure scanner and the pneumotach (Pn). Two voltmeters (DMs) and a controlling valve (V) are also shown.

The upstream opening of model M5 allowed the manual measurement of the minimal glottal diameter. This was performed by using a stack of feeler gages placed snugly through the glottis. The 9 cm of glottal length in the model was not always completely uniform in diameter. Four locations were measured along this length, and feeler gage shims and screw adjustments were made until the nominal 0.04 cm human size (0.30 cm in the model) was obtained. The minimal glottal diameter for the symmetric case was 0.03996 cm human size. For the oblique condition with pressure taps on the divergent side, the diameter was 0.04013 cm human size, and 0.03996 cm human size for the oblique condition with the pressure taps on the convergent side. The accuracy of the measurement of the minimal diameter was ± 0.00034 cm human size (± 0.00254 cm in the model), that is, less than 1% of the 0.04 cm human size target diameter.

G. Procedure for obtaining pressure and flow readings

After the glottal configuration was set, the flow was increased until the desired transglottal pressure (between the tracheal pressure tap and the pressure tap furthest downstream) was obtained as seen on a voltmeter. The voltage from the pneumotach pressure transducer was then noted. Next, the pressure drop at each of the 16 pressure taps was obtained sequentially from the first vocal fold pressure tap to the sixteenth. For each of the pressure measures, the operator waited approximately 1 min before reading the voltage, and always after the voltage values stabilized around small fluctuations. Before taking pressure voltage readings by eye, the voltage for the flow was checked to determine if the flow had changed. For the occasional times when the flow voltage varied away from the prescribed value, the flow was adjusted by means of a gate valve downstream of the pneumotach. For each of the pressure tap locations, four or more readings were taken approximately 15 to 30 s apart and then averaged. The pressure profiles (the pressure drops from tap 1 to tap 16) for each of the prescribed transglottal pressures for all conditions were obtained in this manner.

The pressures on the two sides of the glottis for the symmetric condition were obtained in the following manner. Because the expectation was that the flow would be attached to one side of the glottis and separated from the other side, as well as tend to move toward one side of the downstream duct, a simple method was devised to determine which side the air flow favored. Human hairs were attached to the end of a narrow dowel rod and placed by hand in the upstream

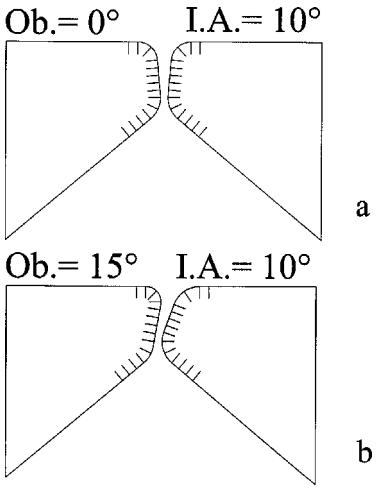


FIG. 4. The two glottal configurations of this study. (a) Outline of the symmetric glottis with a 10-degree divergent angle. (b) Outline of the oblique glottis with a 10-degree divergent angle and an obliquity of 15 degrees. The figures include the pressure tap locations. During the experimental runs, pressure taps existed only on one side (see text).

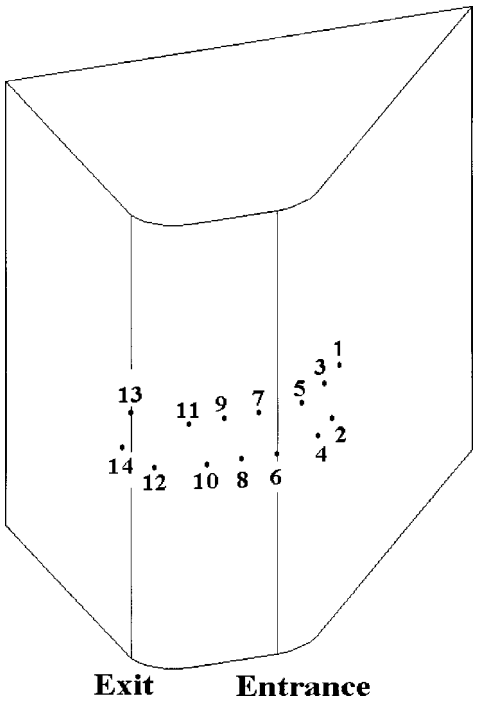


FIG. 5. Pressure taps on the vocal fold pieces. There were 14 taps, staggered near the center of the vocal fold in the axial direction. Tap 6 is at the entrance to the glottis, tap 12 is on the mid-region of the rounded surface of the glottal exit, and taps 13 and 14 are on the top of the vocal fold.

section of the wind tunnel (''trachea''). The hairs were placed through the glottis so that the hairs would follow the direction of the fast air flow. This method suggested the direction of the glottal flow (the presence of the rod and hairs did not appear to affect the direction of the flow). For the symmetric runs, then, the transglottal pressure was set to the desired level, and the rod-and-hair method was used to determine which direction the air was flowing downstream of the glottis. After the pressures were obtained, the flow direction was switched toward the other side by momentarily placing a paper guide in the glottis. The paper guide had an approximately 8.5-cm height and was placed slanted into the glottis to encourage air to flow toward the other side downstream of the glottis. The switch of air flow direction changed the pressures in the glottis except for the 3-cm H_2O condition. For the 3-cm H_2O condition, because of the lower inertia forces, any transverse pressure gradient resulting from the asymmetric flow downstream was not large enough to be detected using the pressure taps. This method allowed the measurement of the effective pressures on both sides of the symmetric glottis using the same flow. For all transglottal pressure conditions, the flow visualization method showed that the flow was not symmetric downstream of the glottis.

For the oblique glottis, there were two sets of vocal folds, as mentioned above. For each pair, there were pressure taps only on one side, requiring the run with one pair to obtain the pressures on the divergent side, followed by an additional run with the other pair to obtain the pressures on the convergent side, using the same glottal diameter, transglottal pressures, and flow rates for both pairs.

H. Computational methods

The commercial code FLUENT (produced by Fluent, Inc., Lebanon, NH) was used to numerically solve the Navier–Stokes equations for the air flow physics occurring inside the two-dimensional symmetric and asymmetric geometries. FLUENT is a computer program for modeling fluid flow in complex geometries and uses the control-volume technique. The grids used in the simulations contained both structured and unstructured meshes. Grids that contained between 15 000 and 126 000 nodes were used to obtain first- or second-order solutions for momentum and pressure which had residuals of less than 10^{-4}. In most cases, uniform inlet and outlet pressure boundary conditions were employed.

III. RESULTS

A. Symmetric glottis pressure profiles

1. Empirical results

The empirical data points for the symmetric glottis are shown in Fig. 6 for the 3-, 5-, 10-, and 15-cm H_2O inlet pressures. Also shown are computational results which will be discussed subsequently. Except for the 3-cm H_2O condition, the pressures on the two sides of the glottis were different downstream of the minimal diameter (downstream of pressure tap 6). The side with the lower pressures (i.e., a greater pressure drop from the trachea) is the side on which the flow was attached until it separated near the glottal exit (demonstrated in the computational results below). The other side (the ''nonflow wall'' or stalled side) had flow separation further upstream in the glottis. The pressures rose more quickly on the nonflow wall. On that side, the pressures were within 5% of the most downstream (supraglottal) pressure by tap 11 for 3 cm H_2O, tap 10 for 5 cm H_2O, tap 9 for 10 cm H_2O, and tap 8 for 15 cm H_2O (human values), suggesting a ''faster rise'' in pressure for larger flow rates (noting that the geometry was constant). The lower pressures on the flow wall, including the pressures at the rounded glottal exit (tap 12), as well as on the immediate superior surface of the vocal fold (taps 13 and 14) and in the side wall of the downstream duct immediately downstream of the glottal exit (tap 15), were below the pressure of the most downstream pressure tap. The largest differences between cross-channel pressures occurred at pressure taps 8 and 11 (or 12), with differences of approximately 5%–6% of the transglottal pressure. These pressures would create different forcing functions on the two vocal folds despite the symmetrical geometry, with lower pressures throughout the glottis (including the superior surface) for the flow wall. It is noted that the pressures on both surfaces upstream of the glottal entrance were sensibly the same regardless of the flow direction in the glottis, for this divergence angle.

Table I gives the transglottal pressure and flow values, dimensional and nondimensional, for the symmetric glottis results. The nondimensional pressure (the pressure coefficient) and the nondimensional flow (the Reynolds number) are included for comparison to other studies (a discussion not pursued in this report). The nondimensional pressure coefficient was defined as $P^* = P/(0.5\rho V^2)$, where the pressure drop P was taken between the upstream pressure tap and tap

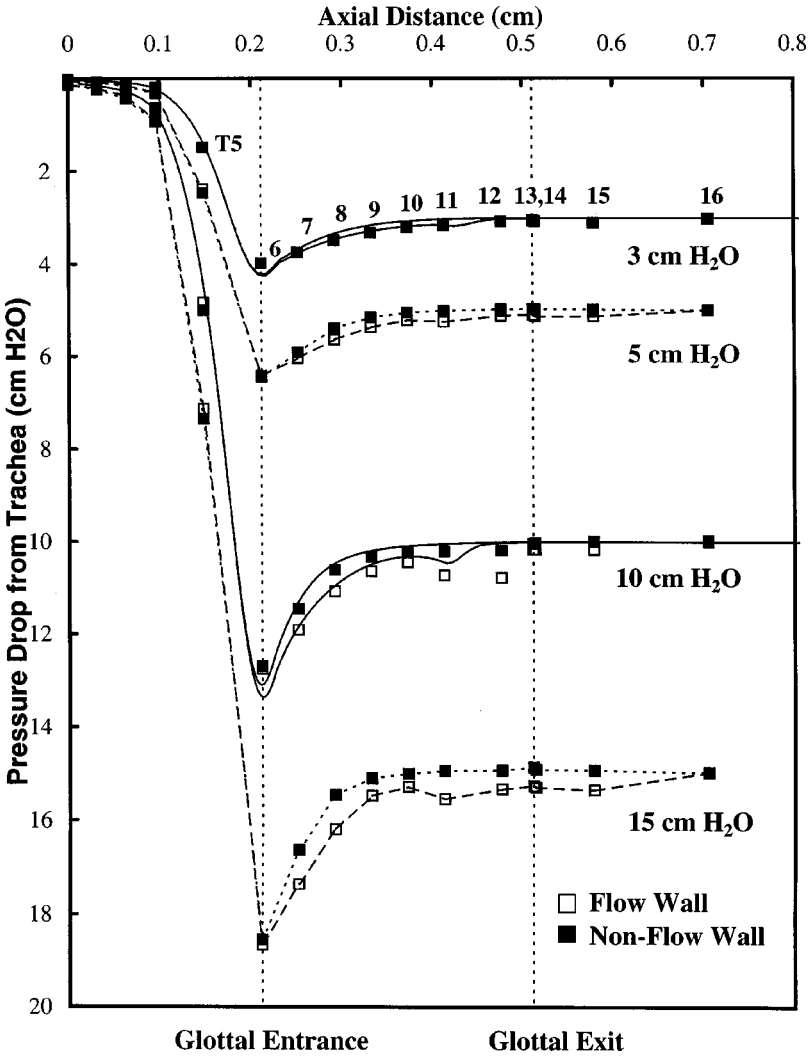


FIG. 6. Empirical pressure measurements for the 3-, 5-, 10-, and 15-cm H_2O (human values) transglottal pressure runs for the symmetric glottis with a 10-degree divergence. For the 3- and 10-cm H_2O transglottal pressure conditions, the predicted pressure profiles (using FLUENT) are shown by solid lines. For the 5- and 15-cm H_2O transglottal pressure conditions, no computed results were obtained and the empirical results are connected by straight dashed lines. The maximum pressure drop from the trachea is located at tap 6 for each transglottal pressure for both the empirical and predicted pressure profiles. The empirical pressures at taps 13 and 14 are staggered slightly at the glottal exit location in the figure, even though they are on the same plane. Tap 16 is downstream of tap 15 by 2.93 cm human size (22 cm in M5), a distance further than indicated in the figure. The zero of the horizontal axis refers to the axial location of the first pressure tap on the vocal fold surface (downstream of the tracheal pressure tap).

6 at the entrance to obtain the entrance pressure coefficient, or between the upstream tap and tap 16 to obtain the transglottal pressure coefficient; ρ is the air density (1.205×10^{-3} g/cm^3); and V is the average velocity obtained by dividing the flow rate by the minimal cross-sectional area of the glottis. The pressure coefficient values decreased in value with increase in volume flow. The Reynolds number was defined as $\mathrm{Re}=VD_h/\nu$, where V is the average velocity, D_h is the hydraulic diameter (four times the glottal area divided by the glottal perimeter), and ν is the kinematic viscosity (0.15 cm^2/s).

2. Computational results for the symmetric glottis

The FLUENT code predicts a symmetric flow through a symmetric glottis when the boundary conditions are symmetric and perpendicular (cross-stream) to the flow axis. Because the empirical experiment of this study did not measure downstream velocity or pressure values in the cross-stream direction, it was not possible to specify an empirical downstream boundary condition. To create a nonsymmetric or skewed flow within the symmetric glottis using FLUENT, it was necessary to impose a nonuniform pressure gradient as a boundary condition downstream of the glottis. This was an artificial way to induce the nonsymmetric flow, but was performed to illustrate how such flow skewness in the glottis produces pressure differences across the glottis that corresponded well to the empirical values, and to permit the qualitative examination of the pressure and flow fields. For the symmetric glottis, FLUENT was used to obtain nonsymmetric flows for the 3- and 10-cm H_2O conditions.

TABLE I. Empirical transglottal pressures and flows (dimensional human values and nondimensional values) for the symmetric condition, 10 degrees divergent glottis, glottal diameter of 0.039 96 cm (human value). The values are given for both the flow wall runs and the nonflow wall runs.

	Dimensional		Nondimensional		
	Transglottal pressure (cm H_2O)	Flow (cm^3/s)	Transglottal pressure coefficient	Entrance pressure coefficient	Reynolds no.
Flow wall	3.016	97.1	1.211	1.595	1045
	4.992	126.4	1.183	1.521	1360
	9.995	185.2	1.103	1.406	1994
	14.977	229.9	1.072	1.336	2475
Non flow wall	3.016	97.1	1.213	1.598	1044
	4.995	126.4	1.186	1.530	1359
	9.976	185.2	1.103	1.403	1992
	14.983	229.9	1.075	1.331	2474

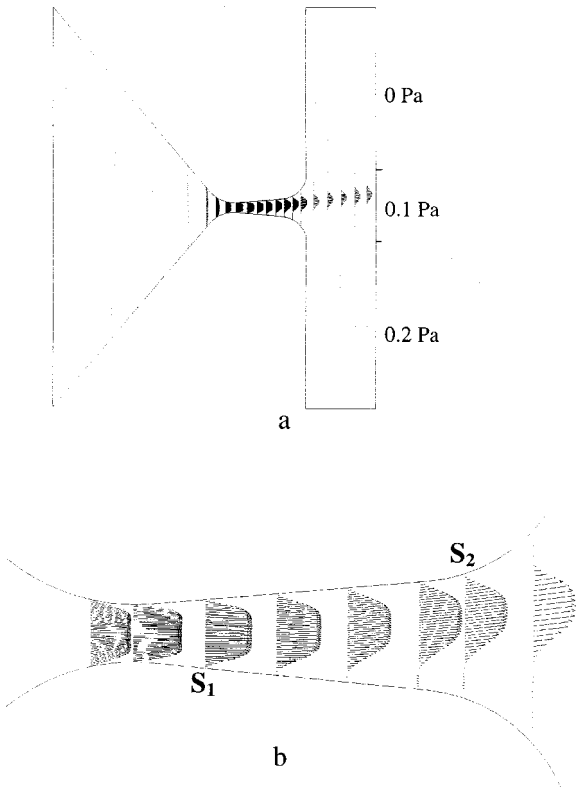


FIG. 7. (a) Predicted asymmetric velocity vectors for the 3-cm H_2O condition in the symmetric model with a nonuniform outlet pressure boundary condition (1 Pa$=1.02\times10^{-2}$ cm H_2O). (b) Detail of predicted asymmetric velocity vectors in the divergent section showing location of separation points on the nonflow wall (S_1) and on the flow wall (S_2).

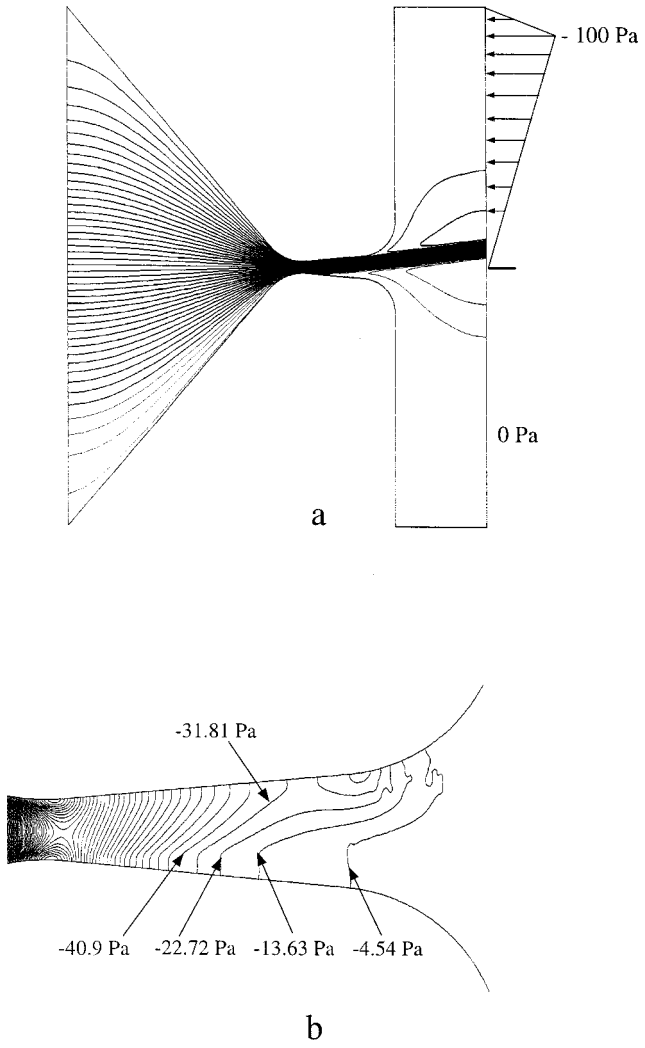


FIG. 8. (a) Predicted asymmetric streamlines for the 10-cm H_2O condition in the symmetric model with a nonuniform outlet pressure boundary condition (1 Pa$=1.02\times10^{-2}$ cm H_2O). (b) Predicted asymmetric pressure contours in the divergent section for the 10-cm H_2O condition in the symmetric model with a nonuniform outlet pressure boundary condition.

For the 3-cm H_2O condition for the symmetric glottis, the empirical pressures on both walls (Fig. 6) did not show any difference in value. This would suggest that the flow was symmetric throughout the glottis. However, good agreement between the empirical and calculated (FLUENT) pressure values could not be obtained using a uniform outlet pressure boundary condition. Since the flow visualization method showed that the fluid jet *was* skewed downstream of the glottis, a slight pressure gradient (0.0 to 0.2 Pa, where 1 Pa$=1.02\times10^{-2}$ cm H_2O) as the downstream boundary condition was sufficient to produce a skewed flow similar to that seen empirically for the glottis. Figures 7(a) and (b) show the resulting velocity vectors through the glottis, with the fastest flow closer to one wall and the exiting flow off center. The figures indicate that flow separated from the lower side (nonflow wall) at a location upstream of the flow separation from the upper side (flow wall). The computational pressure profiles for the 3-cm H_2O condition are shown in Fig. 6 as the solid lines, and suggest that the computational technique using this nonuniform outlet condition matched the pressures well on the flow wall. The computational technique predicted slight pressure differences between the sides whereas the empirical data in Fig. 6, as previously discussed, did not show pressure difference at this low flow rate since the supraglottal pressure variation needed to produce skewness was very slight.

For the 10-cm H_2O condition, a pressure gradient of 100 Pa (1.02 cm H_2O) over half of the downstream section was imposed, as shown in Fig. 8(a). The figure shows the flow streamlines through the glottis, again with the fastest flow closer to one wall and the exiting flow off center. Figure 8(b) indicates the corresponding pressure contours within the glottis, which show higher pressures near the nonflow wall, consistent with the pressure profiles for the 10-cm H_2O condition in Fig. 6. Compared with the empirical values, the computed pressures on the nonflow wall show a slightly lower pressure at the glottal entrance and a slightly higher pressure within the glottis. For the flow wall, the computed pressures again were lower at glottal entrance compared to the empirical values, and were reasonably matched within the glottis until taps 10 and 11, where the computational values did not follow the empirical dip. In general, the comparisons between the computational predictions and the empirical data were reasonable, despite the imprecise method of establishing the flow asymmetry. The computational results support the pattern and magnitude of the empirical pressure differences across the glottal duct. Similar computations were not performed for the 5- and 10-cm H_2O conditions.

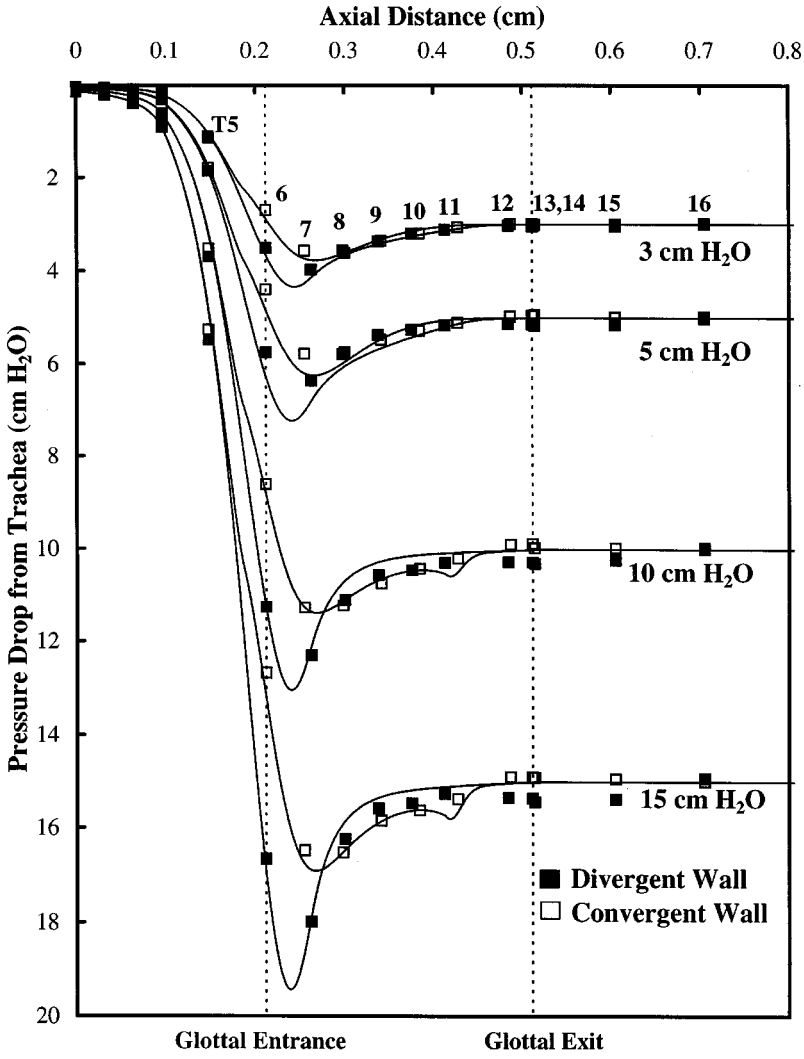


FIG. 9. Empirical pressure measurements and predicted (FLUENT) pressure profiles for the 3-, 5-, 10-, and 15-cm H_2O transglottal pressure runs for the glottis with an oblique angle of 15 degrees and included angle of 10 degrees. Pressures and distances are in human values rather than model values.

For these, the dashed straight lines in Fig. 6 were used to connect the empirical data points.

B. Oblique glottis pressure profiles

Figure 9 shows the pressure profiles for the two sides of the oblique glottis for the four transglottal pressures. The pressures at the glottal entrance (tap 6) were higher than at pressure tap 7 just downstream, for both sides of the glottis. This was because the minimal glottal diameter was downstream of pressure tap 6 by approximately 0.03 cm (human size), between taps 6 and 7 (closer to tap 7). The glottal diameter at pressure tap 6 was 7.3% larger than for the symmetric condition (due to the use of the existing vocal fold pieces created for model M5). The pressures upstream of tap 6 tended to converge to the same value at tap 5.

For all four transglottal pressure conditions, the cross-channel pressures at tap 6 were different by 27% of the transglottal pressure, with pressure being higher on the convergent wall. The pressure on the convergent wall was also higher at pressure tap 7 by 13.5%, 11.8%, 10.3%, and 10.1% for the 3-, 5-, 10-, and 15-cm H_2O conditions, respectively (with the understanding that tap 7 was 0.0076 cm human size more upstream for the convergent side). It is observed from Fig. 9 that the pressure on the convergent side near the entrance was higher than the supraglottal pressure, and the pressure on the diverging side was less than the supraglottal pressure. If the supraglottal pressure were taken to be zero, the convergent entry wall would be pushed in the direction of glottal opening, and the divergent entry wall would be pulled in the direction of closure. These opposite forces would promote motion of the two sides (at the glottal entrance) in the *same* direction, thus potentially enhancing glottal wall phasing differences.

The pressures on the two sides downstream of tap 7 were approximately the same (despite the divergence geometry) and rose within the diverging glottal duct over taps 8–11. The pressures diverged on the rounded portion of the exit radii (pressure tap 12), again with the pressure higher on the convergent side, ranging from 1.0% to 3.8% of the transglottal pressure (1.0%, 3.2%, 3.8%, and 2.9% for 3, 5, 10, and 15 cm H_2O, respectively). This difference, although minor, suggests slightly different forces at the glottal exit (as well as on the superior surface of the vocal folds). These pressure differences indicate that the oblique glottis with a 10-degree divergence and 15-degree obliquity produced larger entrance and exit pressures on the convergent side than on the divergent side, with the middle portion of approximately the same pressures for constant flow. The primary differences between the symmetric (Fig. 6) and oblique

TABLE II. Empirical transglottal pressures and flows (dimensional human values and nondimensional values) for the glottal condition of 15 degrees oblique and 10 degrees divergent. The glottal diameter was 0.039 96 cm (human value) for the experimental run with the pressure taps on the convergent side, and 0.040 13 cm (human value) for the run with the pressure taps on the divergent side.

	Dimensional		Nondimensional		
	Transglottal pressure (cm H_2O)	Flow (cm^3/s)	Transglottal pressure coefficient	Entrance pressure coefficient	Reynolds no.
Convergent wall	2.996	94.8	1.263	1.140	1020
	4.977	122.3	1.261	1.113	1315
	9.988	178.8	1.185	1.020	1922
	14.995	222.1	1.153	0.975	2388
Divergent wall	2.990	96.1	1.237	1.453	1033
	5.014	124.6	1.232	1.413	1340
	9.964	182.6	1.140	1.289	1964
	14.921	228.0	1.096	1.223	2451

(Fig. 9) pressures were (1) the pressures across the glottis were the same at glottal entry for the symmetric geometry but significantly different for the oblique. This was because the symmetric geometry caused the flow and pressure distributions to remain symmetric through the glottal entrance and just past the glottal minimum width. Pressure differences between the solid surfaces did not occur until the flow became skewed further downstream, and pressure disturbances due to this flow skewness did not affect the glottal entry region flow. For the oblique geometry, the flow and pressure distributions lost their symmetry near the minimal glottal width due to the change in the flow direction caused by the oblique geometry. The resulting asymmetric pressure disturbances propagated back and affected the flow behavior in the glottal entry region. Also, (2) the pressures were significantly different in the mid-glottis for the symmetric geometry but similar for the oblique. For the symmetric geometry, the unstable flow in the mid-glottal region resulted in flow skewness and widely different separation point locations on the solid surfaces. The difference in the size and position of these separation regions resulted in pressure differences in the mid-glottal region between the solid surfaces, with the differences being larger at the higher flow rates. The oblique geometry changed the flow direction at the glottal entrance and the established flow tended to have separation points further downstream in the glottis, thus making the surface pressures in the mid-glottal region reasonably equal.

Table II gives the dimensional and nondimensional pressure and flow values for the oblique condition. The transglottal pressure coefficient ranged from 1.096 to 1.263, being larger for lower flows and lower transglottal pressures. The pressure coefficient at glottal entry was different for the two sides, ranging from 0.97 to 1.14 for the convergent side (larger for lower flows), and 1.22 to 1.45 for the divergent side (larger for lower flows).

C. Computational results for the oblique glottis

Figure 9 shows the M5 data points as well as the computational pressure predictions for the 3-, 5-, 10- and 15-cm

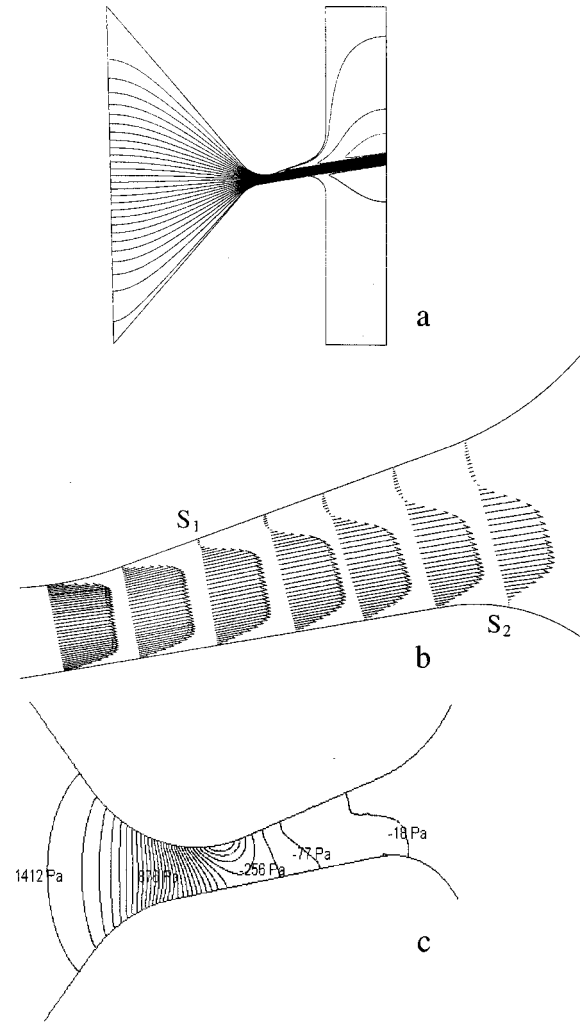


FIG. 10. (a) Predicted streamlines for the 5-cm H_2O condition in the asymmetric model. (b) Predicted velocity vectors for the 5-cm H_2O condition in the asymmetric model showing location of separation points on the divergent wall (S_1) and on the convergent wall (S_2). (c) Predicted pressure contours for the 5-cm H_2O condition in the asymmetric model ($1\ \text{Pa} = 1.02 \times 10^{-2}\ \text{cm}\ H_2O$).

H_2O condition. For the 3-cm H_2O condition, excellent agreement between these sets of values is apparent. The pressure dip near the glottal entrance was lower than the data might suggest, and the flow was 9.5% higher than for the experiment with M5. Comparison between the empirical and calculated results suggests similar findings for the 5-cm H_2O condition, although the fit to the data was not as good for the divergent side (the maximum difference (tap 7) was within 9% of the transglottal pressure). The flow was 9.6% higher than for M5.

Figures 10(a)–(c) show streamlines, velocity vectors, and pressure contours, respectively, for the 5-cm H_2O case (which is similar in nature to the 3-cm H_2O condition). The main flow exited approximately along the direction of the obliquity, creating a jet that moved toward the wall of the corresponding downstream duct. Flow separation took place on both walls upstream of the glottal exit. Although the main flow near the exit was closer to the divergent wall, the separation from that wall was upstream of the separation point on the convergent wall (0.09 cm human size from the glottal

TABLE III. Summary of computational information for symmetric and obliquity conditions. For the oblique case, S_1 and S_2 are the locations of the flow separation points on the divergent and convergent sides, respectively, measured from the glottal entrance [most medial location of the divergent side (tap 6)]. ''Pred.'' refers to the predicted flow values. ''Nodes'' are the total number of points defining the faces of the cells in the mesh. ''1st/2nd order'' stands for solution accuracy).[5] ''Resid.'' stands for the accepted residual value.[5] ''Iter'' is the number of iterations to satisfy the desired residual. The pressure, flow, and separation location values are in human values rather than model values.

Transglottal pressure (cm H_2O)	Geometry	Flow (cm^3/s) Pred./M5	Separation points S_1 (cm)	Separation points S_2 (cm)	Computational details Nodes (k)	(1st/2nd) order	Resid.	Iter. (k)
3	symmetric	108/97	···	···	15	2	10^{-5}	12
10	symmetric	198/185	···	···	50	2	10^{-5}	15
3	oblique	104/95	0.11	0.16	60	2	10^{-5}	10
5	oblique	137/125	0.09	0.15	60	2	10^{-5}	55
10	oblique	188/180	0.07	0.23	126	1	10^{-4}	12
15	oblique	232/225	0.07	0.23	126	1	10^{-4}	7

entrance for the divergent wall, 0.15 cm human size for the convergent wall). Recall that the glottal duct length was 0.30 cm human size. The separation zone near the convergent surface, however, formed over a larger region due to the high diverging angle opposite. The jet was pushed by the cross-channel pressure gradient [shown in Fig. 10(c)] closer to the diverging side in the downstream half of the glottal duct, but the jet did not ''hug'' the divergent surface. Table III provides the location of the separation points for all four transglottal pressure conditions.

For the 10-cm H_2O condition as shown in Fig. 9, the prediction for the divergent side was reasonable, with higher predicted values downstream of the minimal diameter. The prediction for the convergent side was excellent. The FLUENT code predicted the jet to ''hug'' the convergent surface creating a second dip near tap 11, whereas the data for M5 did not show this additional (but slight) dip. The jet was slightly compressed in width by the larger separated region above it; mass conservation then required the jet velocity to increase in this smaller width region, thereby decreasing the pressure and creating the dip. For the 15-cm H_2O condition, the comparison was similar to the 10-cm H_2O condition, with the divergent wall predictions being higher downstream than the empirical data and the predicted pressures for the convergent wall being excellent. The difference between the predicted and measured flow was 4.4% for 10-cm H_2O case and 3.1% for the 15-cm H_2O case.

Figures 11(a)–(c) show the predicted streamlines, velocity vectors, and pressure contours, respectively, for the 15-cm H_2O condition (which is similar in nature to the 10-cm H_2O condition). The higher inertia forced the jet toward the convergent wall in the glottal region. The larger separated region was associated with the divergent wall. The predicted pressure contours indicated that the pressures on the divergent wall were less than on the convergent wall near glottal entry, but then were higher in the downstream half of the glottis, creating forces to keep the flow near the convergent wall. The separation points, compared to the 3- and 5-cm H_2O conditions, were more upstream for the divergent side (S1), and more downstream for the convergent side (S2) for the 10- and 15-cm H_2O conditions (Table III).

D. Pressure comparisons between the symmetric and oblique conditions

Figure 12 is a composite of the empirical pressure profiles for the symmetric and oblique glottal configurations. The major difference between the two configurations was the pressure at the glottal entrance (tap 6). The pressures were lower for the symmetric configuration and higher for the two sides of the oblique glottis at that location. The increase in entrance area of 7.3% would ideally result in a smaller pressure drop (by 13%) (using the Bernoulli equation); this predicts the pressure on the *divergent* wall of the oblique configuration within 4% (average of 2.8%) across the four transglottal pressure conditions, a strong prediction. For the pressures at taps 8–11 within the glottis, the pressures on the nonflow wall of the symmetric configuration were higher than the pressures on either side of the oblique configuration. However, the pressures on the flow wall in the symmetric configuration were similar to those on the oblique walls. The symmetric flow wall pressures for taps 8–11 differed from the average of the pressures on the oblique walls by an average of only 1.5% (range 0.02% to 4.6%).

IV. DISCUSSION

This study relates directly only to steady-flow concepts because model M5, a nonvibrating laryngeal model, was used with constant flows. Although quasi-steady conditions of phonation have been assumed as justification for steady-flow experiments and the adoption of steady-flow results in aerodynamic models, it is recognized that the pressure profiles and glottal flow behavior may differ between steady-flow conditions as studied here and realistic phonation, and that it is too early to determine the extent of these differences (McGowan, 1993; Mongeau *et al.*, 1997; Jiang and Titze, 1994; Alipour and Scherer, 2000).

However, there is evidence that flows move off-axis through a time-varying glottis. Alipour *et al.* (1996a) used a

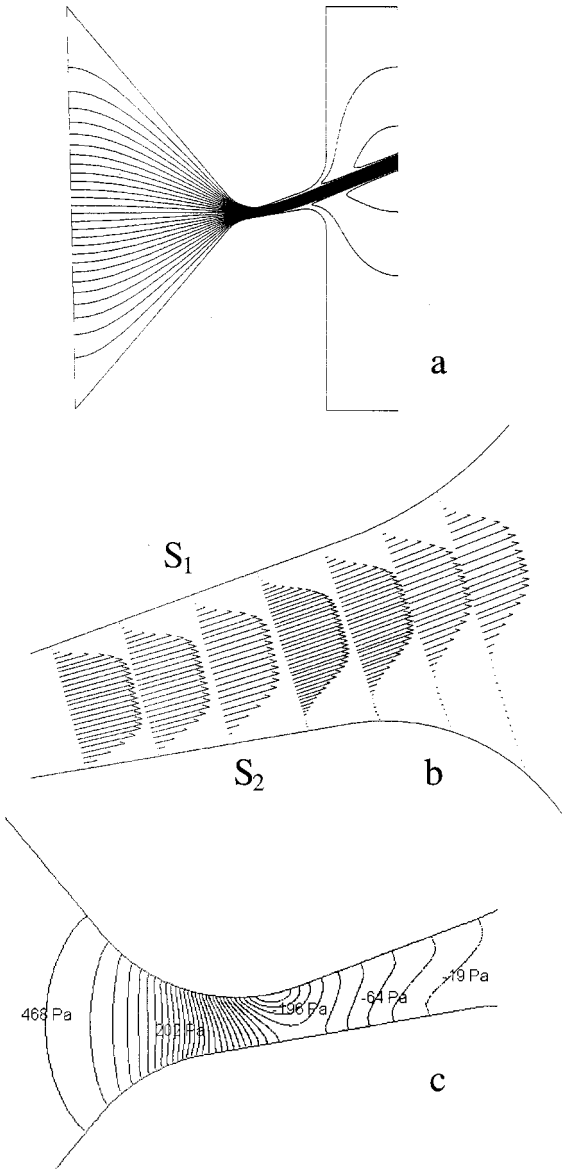


FIG. 11. (a) Predicted streamlines for the 15-cm H_2O condition in the asymmetric model. (b) Predicted velocity vectors for the 15-cm H_2O condition in the asymmetric model showing location of separation points on the divergent wall (S_1) and on the convergent wall (S_2). (c) Predicted pressure contours for the 15-cm H_2O condition in the asymmetric model ($1\text{Pa}=1.02\times10^{-2}$ cm H_2O).

finite volume computational method to simulate glottal flow with forced sinusoidal motion of the glottal walls (with the glottal entry location fixed). Asymmetric velocity distributions (as a result of flow separation) were a dominant feature of the simulation. Using excised larynxes, Alipour and Scherer (1995) and Alipour *et al.* (1995) determined peak and average velocity surfaces 1 cm above the glottis, which demonstrated greater velocities lateral to the glottal midline. These surfaces were highly three-dimensional, with greater flows emanating from the anterior portion of the glottis. Off-axis maximum velocities were detected in another study of an excised larynx in Alipour *et al.* (1996b). It was noted that the greatest velocities existed at a time just after maximum glottal opening, and dominated during that interval of vocal fold oscillation when the folds would produce a diverging glottal duct. Other research tends to suggest that there may be insufficient time to create asymmetric flow separation in the glottis during phonation (Hirschberg *et al.*, 1996), or that, if there were flow acceleration through the glottis, the separation would be symmetric (Pelorson *et al.*, 1995, who used the same divergent angle as used here, 10 degrees). Assuming that separation is symmetric, it may also be dynamic in that the location of the separation point might move, so that an assumption of a quasi-steady flow with the adoption of a moving separation point during glottal closure can be made. This was done with some success by Pelorson *et al.* (1994). There are additional studies of steady and developing unsteady flow that support the existence of asymmetric flow through symmetric diverging glottal shapes (Liljencrants, 1991; Hirschberg *et al.*, 1996; Alipour *et al.*, 1996b).

The finding that the highest air velocities come from the anterior glottis of excised canine larynxes, shown in the work by Alipour and Scherer (1995) and Alipour *et al.* (1995), was supported by Berke *et al.* (1989). The extended work for *in vivo* dog larynxes from the same research laboratory (Bielamowicz *et al.*, 1999) also indicated that the largest velocities again were through the anterior glottis; toward the middle of the glottis, however, the larger velocities were off center rather than midline. This strongly suggests an interaction between the dynamic three-dimensional shape of the glottis and the eccentricity of the glottis in the airway duct.

The most important areas of phonation that this study indirectly addresses are voicing perturbations, glottal jets, and bilateral vocal fold phasing. This study suggests that for a divergent glottal duct of 10 degrees, the flow in a symmetric glottis (that is, with an oblique angle of zero degrees) may become unstable, resulting in nonsymmetric flow separation regions near the vocal fold surfaces, unequal surface pressures, and flow exiting the glottis at an angle. The asymmetry of pressures would create slightly different pressure forces on the surfaces of the vocal folds. For example, for a transglottal pressure of 10 cm H_2O, the cross-channel differences from tap 7 to 12 (most of the vocal fold medial surface) ranged from 2.3% to 5.9%, with an average of 4.3% of the transglottal pressure. This may be a sufficient asymmetry of pressures to help create normal levels of voicing jitter (cycle to cycle variations in voicing periods), but not enough difference to disrupt phonation or cause significant phasing differences between the two vocal folds.

The oblique glottis studies with the same included angle suggest that the pressures are different across the glottal duct in such cases, and these differences are due to the obliquity *per se*, with pressures higher on the convergent side near glottal entry and near glottal exit. The differences at glottal entry were 27% of the transglottal pressure, a difference that might be considered high, and perhaps influential in maintaining the out-of-phase motion of the two vocal folds. The 3- and 5-cm H_2O conditions were qualitatively different from the 10- and 15-cm H_2O conditions in that the flow was near the divergent surface at glottal exit more for the former and near the convergent surface for the latter. The shift in the location of separation points was definitive.

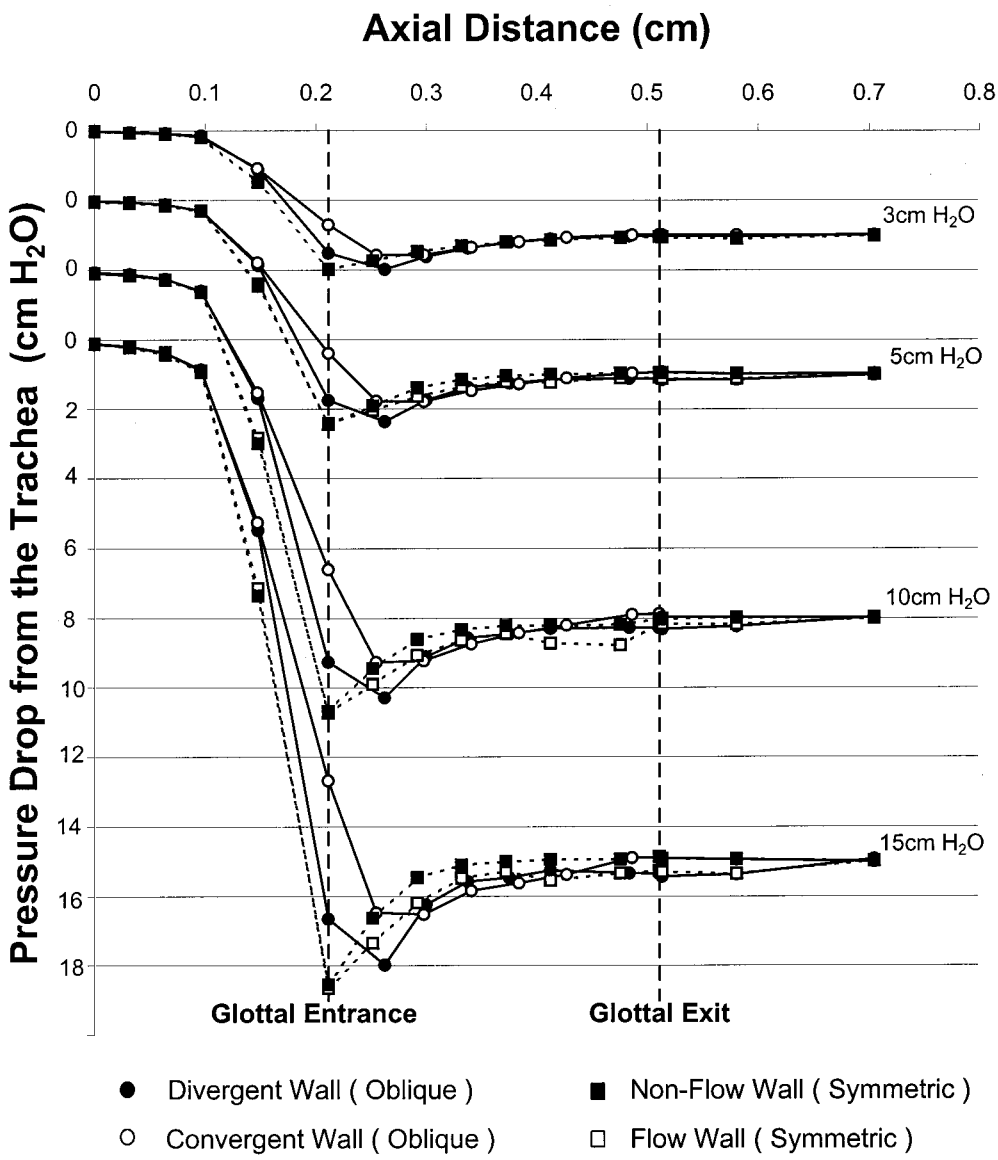


FIG. 12. Combined empirical pressure profiles for the symmetric and oblique glottal configurations for 10 degree divergence glottis for 3-, 5-, 10-, and 15-cm H_2O transglottal pressure drops. Pressures and distances are in human values rather than in model values. The pressures at the first pressure tap (zero axial distance) are offset in the figure for clarity in showing the pressure measurements.

This study suggests that attention should be placed on expanding empirical experimentation to determine the changes of flow separation points, flow directions, and bilateral pressures in the glottis for static and dynamic changes of glottal shape. Such studies will allow the refinement of current aerodynamic equations for glottal mechanics, which may need to include bilateral pressure expressions. The equations should be extended to oblique glottal shapes.

The glottal dimensions used in this study were limited to one minimal glottal diameter, one included glottal angle, and two oblique directions. This configurational limitation is severe, but a starting point for detailed pressure profiles in the glottis.

V. CONCLUSIONS

A Plexiglas model of the larynx was used to study intraglottal wall pressures for a symmetric and oblique glottal angle of 10 degrees divergence. Fourteen pressure taps on the vocal fold surfaces gave detailed pressure profiles. The computational program FLUENT was successful in matching the intraglottal pressures. Pressures that were unequal across the glottis were obtained for both the symmetric and oblique glottis. The results for the symmetric condition suggest that asymmetric pressures may be present during phonation, perhaps contributing to normal values of jitter. The results for the oblique (15 degree) glottis suggest that the asymmetric pressures may promote phasing differences between the two vocal folds.

This study suggests that current aerodynamic equations for glottal motion may need to be expanded to include bilateral expressions, for both symmetric and oblique glottal shapes. Further steady and unsteady flow studies should provide data to make practical decisions on modeling extensions.

ACKNOWLEDGMENTS

Support for this research has come from NIH Grant No. 1 R01 DC03577 from NIDCD.

[1]The model shaping is similar to that theoretically devised by Pelorson *et al.* (1994).

[2]The value of 0.04 cm human size for the minimal diameter is a value that most likely occurs during any phonation except for highly breathy qualities.

[3]The actual pressure drops in the model for human size transglottal pressure drops of 3, 5, 10, and 15 cm H_2O were $(7.5)^{-2}$ times those values and were therefore 0.0533, 0.0889, 0.1778, and 0.2667 cm H_2O, respectively. The actual volume flows in the model were 7.5 times human size, so that, for example, a flow of 100 cm^3/s human size would be 750 cm^3/s in model M5.

[4]Uncertainty analysis (Dieck, 1998) led to estimates of the accuracy of the pressure and flow measurements. The pressure tap measurement accuracy was approximately $U_{95}=\pm 0.01$ to 0.06 cm H_2O human size equivalent (the amount depending upon transducer gain). For the worst case, the systematic uncertainty was ± 0.042 cm H_2O human size, with random uncertainty ± 0.083 cm H_2O human size, at the 95% confidence level. When variation of pressure readings during data collection were taken into account, the estimated uncertainty of repeated pressure measurements was $U_{95}=\pm 0.064$ cm H_2O human size. Uncertainty for the flow measurements was estimated to be $U_{95}=\pm 4.24$ cm^3/s human size, with systematic uncertainty of ± 4.0 cm^3/s human size and random uncertainty of ± 1.56 cm^3/s human size. These uncertainties were considered to be small.

[5]A first-order accurate solution takes the value of a dependent variable at a cell surface to be equal to the value at the center of the upstream cell, whereas the second-order accurate solution calculates the cell surface value from the average of the upstream cell center value and the interpolated downstream cell center value. The accepted residual value is the calculated measure characterizing the degree to which conservation of mass and momentum for all the cells in the mesh are satisfied.

Alipour, F., and Scherer, R. C. (**1995**). ''Pulsatile airflow during phonation: An excised larynx model,'' J. Acoust. Soc. Am. **97**, 1241–1248.

Alipour, F., and Scherer, R. C. (**2000**). ''Dynamic glottal pressures in an excised hemilarynx model,'' J. Voice **14**(4), 443–454.

Alipour, F., Fan, C., and Scherer, R. C. (**1996a**). ''A numerical simulation of laryngeal flow in a forced-oscillation glottal model,'' Comput. Speech Lang. **10**, 75–93.

Alipour, F., Scherer, R., and Knowles, J. (**1996b**). ''Velocity distributions in glottal models,'' J. Voice **10**(1), 50–58.

Alipour, F., Scherer, R., and Patel, V. C. (**1995**). ''An experimental study of pulsatile flow in canine larynges,'' J. Fluids Eng. **117**, 577–581.

Ashjaee, J., and Johnston, J. P. (**1980**). ''Straight-walled, two-dimensional diffusers—transitory stall and peak pressure recovery,'' J. Fluids Eng. **102**, 275–282.

Berke, G. S., Moore, D. M., Monkewitz, P. A., Hanson, D. G., and Gerratt, B. R. (**1989**). ''A preliminary study of particle velocity during phonation in an in vivo canine model,'' J. Voice **3**(4), 306–313.

Bielamowicz, S., Berke, G. S., Kreiman, J., and Gerratt, B. R. (**1999**). ''Exit jet particle velocity in the *in vivo* canine laryngeal model with variable nerve stimulation,'' J. Voice **13**(2), 153–160.

Binh, N., and Gauffin, J. (**1983**). ''Aerodynamic measurements in an enlarged static laryngeal model,'' Speech Transmission Laboratory Quarterly Progress and Status Report STL-QPSR 2-3/1983, Royal Institute of Technology, Stockholm, pp. 36–60.

Carlson, J. J., Johnston, J. P., and Sagi, C. J. (**1967**). ''Effects of wall shape on flow regimes and performance in straight, two-dimensional diffusers,'' J. Basic Eng. **89**, 151–160.

Cherdron, W., Durst, F., and Whitelaw, J. H. (**1978**). ''Asymmetric flows and instabilities in symmetric ducts with sudden expansions,'' J. Fluid Mech. **84**(1), 13–31.

Dieck, R. H. (**1998**). ''Simple calibration uncertainty determination,'' Instrument Society of America Test Measurement Division Newsletter, **34**(1), 18–28.

Gauffin, J., Binh, N., Ananthapadmanabha, T. V., and Fant, G. (**1983**). ''Glottal geometry and volume velocity waveform,'' in *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, edited by D. M. Bless and J. H. Abbs (College-Hill, San Diego), pp. 194–201.

Guo, C. G., and Scherer, R. C. (**1993**). ''Finite element simulation of glottal flow and pressure,'' J. Acoust. Soc. Am. **94**(2), 688–700.

Hirano, M. (**1981**). *Clinical Examination of Voice* (Springer-Verlag, New York).

Hirano, M., and Bless, D. M. (**1993**). *Videostroboscopic Examination of the Larynx* (Singular, San Diego).

Hirano, M., Kiyokawa, K., and Kuria, S. (**1988**). ''Laryngeal muscles and glottis shaping,'' in *Vocal Physiology: Voice Production, Mechanisms and Functions*, edited by O. Fujimura (Raven, New York), pp. 49–65.

Hirschberg, A., Pelorson, X., Hofmans, G. C. J., van Hassel, R. R., and Wijnands, A. P. J. (**1996**). ''Starting transient of the flow through an *in-vitro* model of the vocal folds,'' in *Vocal Fold Physiology, Controlling Complexity and Chaos*, edited by P. J. Davis and N. H. Fletcher (Singular, San Diego), pp. 31–46.

Ishizaka, K., and Flanagan, J. L. (**1972**). ''Synthesis of voiced sounds from a two-mass model of the vocal cords,'' Bell Syst. Tech. J. **51**(6), 1233–1268.

Ishizaka, K., and Matsudaira, M. (**1972**). ''Fluid Mechanical Considerations of Vocal Cord Vibration,'' SCRL Monograph No. 8, Speech Communications Research Laboratory, Inc., Santa Barbara.

Jiang, J. J., and Titze, I. R. (**1994**). ''Measurement of vocal fold intraglottal pressure and impact stress,'' J. Voice **8**(2), 132–144.

Kline, S. J. (**1959**). ''On the nature of stall,'' J. Basic Eng. **81**, Series D, No. 3, 305–322.

Koike, Y., and Imaizumi, S. (**1988**). ''Objective evaluation of laryngostroboscopic findings,'' in *Vocal Physiology: Voice Production, Mechanisms and Functions*, edited by O. Fujimura (Raven, New York), pp. 433–442.

Liljencrants, J. (**1991**). ''Numerical simulations of glottal flow,'' in *Vocal Fold Physiology, Acoustics, Perceptual, and Physiological Aspects of Voice Mechanics*, edited by J. Gauffin and B. Hammarberg (Singular, San Diego), pp. 99–104.

Liljencrants, J. (**1996**). ''Analysis by synthesis of glottal airflow in a physical model,'' TMH-QPSR 2/1996 Royal Institute of Technology, Stockholm, pp. 139–142.

McGowan, R. S. (**1993**). ''The quasisteady approximation in speech production,'' J. Acoust. Soc. Am. **94**, 3011–3013.

Mongeau, L., Franchek, N., Coker, C. H., and Kubli, R. A. (**1997**). ''Characteristics of a pulsating jet through a small modulated orifice, with application to voice production,'' J. Acoust. Soc. Am. **102**, 1121–1134.

Parsons, D. J., and Hill, P. G. (**1973**). ''Effects of curvature on two-dimensional diffuser flow,'' J. Fluids Eng. **95**(3), 349–360.

Pelorson, X., Hirschberg, A., van Hassel, R. R., Wijnands, A. P. J., and Auregan, Y. (**1994**). ''Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model,'' J. Acoust. Soc. Am. **96**, 3416–3431.

Pelorson, X., Hirschberg, A., Wijnands, A. P. J., and Bailliet, H. (**1995**). ''Description of the flow through in-vitro models of the glottis during phonation,'' Acta Acustica **3**, 191–202.

Reneau, L. R., Johnston, J. P., and Kline, S. J. (**1967**). ''Performance and design of straight, two-dimensional diffusers,'' J. Basic Eng. **89**, 141–150.

Scherer, R. C. (**1991**). ''Physiology of phonation: A review of basic mechanics,'' in *Phonosurgery: Assessment and Surgical Management of Voice Disorders*, edited by C. N. Ford and D. M. Bless (Raven, New York), pp. 77–93.

Scherer, R. C. (**1995**). ''Laryngeal function during phonation,'' in *Diagnosis and Treatment of Voice Disorders*, edited by J. S. Rubin, R. T. Sataloff, G. S. Korovin, and W. J. Gould (Igaku-Shoin Medical, New York), pp. 86–104.

Scherer, R. C., and Guo, C. G. (**1990**). ''Laryngeal modeling: Translaryngeal pressure for a model with many glottal shapes,'' in *ICSLP Proceedings, 1990 International Conference on Spoken Language Processing*, Vol. 1 (Acoustical Society of Japan), pp. 3.1.1–3.1.4.

Scherer, R. C., and Titze, I. R. (**1983**). ''Pressure-flow relationships in a model of the laryngeal airway with a diverging glottis,'' in *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, edited by D. M. Bless and J. H. Abbs (College-Hill, San Diego), pp. 179–193.

Scherer, R. C., Titze, I. R., and Curtis, J. F. (**1983a**). ''Pressure-flow relationships in two models of the larynx having rectangular glottal shapes,'' J. Acoust. Soc. Am. **73**, 668–676.

Scherer, R. C., Vail, V. J., and Rockwell, B. (**1995**). ''Examination of the laryngeal adduction measure EGGW,'' in *Producing Speech: Contemporary Issues: for Katherine Safford Harris*, edited by F. Bell Berti and L. J. Raphael (American Institute of Physics, Woodbury, NY), pp. 269–290.

Scherer, R. C., Titze, I. R., Linville, R., Hueffner, D., and Shaw, K. (**1983b**). ''The effects of vocal fold growths on pressure-flow relationships in the larynx,'' in *Transcripts of the Eleventh Symposium: Care of the Profes-*

sional Voice, edited by V. Lawrence (The Voice Foundation, New York), pp. 25–36.
Schönhärl, E. (**1960**). *Die Stroboskopie in der praktischen Laryngologie* (Thieme, Stuttgart).
Schutte, H. K., Svec, J. G., and Sram, F. (**1998**). ''First results of clinical application of videokymography,'' Laryngoscope **108**, 1206–1210.
Shadle, C. H., Barney, A. M., and Thomas, D. W. (**1991**). ''An investigation into the acoustics and aerodynamics of the larynx,'' in *Vocal Fold Physiology, Acoustics, Perceptual, and Physiological Aspects of Voice Mechanics*, edited by J. Gauffin and B. Hammarberg (Singular, San Diego), pp. 73–82.
So, R. M. C. (**1976**). ''Entry flow in curved channels,'' J. Fluids Eng. **98**, 305–310.
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.
Streeter, V. L., and Wylie, E. B. (**1975**). *Fluid Mechanics*, 6th ed. (McGraw-Hill, New York).
Svec, J. G., and Schutte, H. K. (**1996**). ''Videokymography: High-speed line scanning of vocal fold vibration,'' J. Voice **10**(2), 201–205.
Svec, J. G., Schutte, H. K., and Sram, F. (**1999a**). ''Variability of vibration of normal vocal folds as seen in videokymography,'' in *Communication and Its Disorders: A Science In Progress*, Proceedings 24th Congress International Association of Logopedics and Phoniatrics, Amsterdam, The Netherlands, 23–27 August 1998, edited by P. H. Dejonckere and H. F. M. Peters (International Association of Logopedics and Phoniatrics), Vol. 1, pp. 122–125.
Svec, J. G., Sram, F., and Schutte, H. K. (**1999b**). ''Videokymography: a new high-speed method for the examination of vocal-fold vibrations,'' Otorinolaryngologie a foniatrie (Prague) **48**(3), 155–162.
Titze, I. R. (**1974**). ''The human vocal cords: A mathematical model. Part II,'' Phonetica **29**, 1–21.
Titze, I. R. (**1986**). ''Mean intraglottal pressure in vocal fold oscillation,'' J. Phonetics **14**, 359–364.
Titze, I. R. (**1988**). ''The physics of small-amplitude oscillation of the vocal folds,'' J. Acoust. Soc. Am. **83**, 1536–1552.
Titze, I. R., and Talkin, D. T. (**1979**). ''A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation,'' J. Acoust. Soc. Am. **66**, 60–74.
Tsui, Y.-Y., and Wang, C.-K. (**1995**). ''Calculation of laminar separated flow in symmetric two-dimensional diffusers,'' J. Fluids Eng. **117**, 612–616.
van den Berg, J., Zantema, J. T., and Doornenbal, P. (**1957**). ''On the air resistance and the Bernoulli effect of the human larynx,'' J. Acoust. Soc. Am. **29**, 626–631.
von Leden, H., Moore, P., and Timcke, R. (**1960**). ''Laryngeal vibrations: Measurements of the glottic wave. Part III: The pathologic larynx,'' Arch. Otolaryngol. **71**, 16–35.
Ward Smith, A. J. (**1971**). *Pressure Losses in Ducted Flows* (Davey, Harford, England).
Zemlin, W. R. (**1998**). *Speech and Hearing Science, Anatomy and Physiology*, 4th ed. (Allyn & Bacon, Needham Heights, MA), pp. 150–152.

Stochastic models of jitter[a)]

Jean Schoentgen[b)]
Laboratoire de Phonétique Expérimentale, Université Libre de Bruxelles, Brussels, Belgium

(Received 6 December 1999; accepted for publication 29 December 2000)

This study presents stochastic models of jitter. Jitter designates small, random, involuntary perturbations of the glottal cycle lengths. Jitter is a base-line phenomenon that may be observed in all voiced speech sounds. Knowledge of its properties is therefore relevant to the acoustic modeling, analysis, and synthesis of voice quality. Also, models of jitter are conceptual frameworks that enable experimenters and clinicians to distinguish jitter in particular from aperiodic cycle length patterns in general. Vocal jitter is modeled by means of the ribbon model of the glottal vibration combined with stochastic models of the disturbances of the instantaneous frequency. The disturbance model comprises correlation-free noise and vocal microtremor. Properties of jitter that are simulated are the stochasticity, stationarity, and normality of the decorrelated cycle length perturbations, the size of decorrelated jitter, the correlation between the perturbations of neighboring glottal cycles, the modulation level and modulation frequency owing to microtremor, the asynchrony between external disturbances and glottal cycles, the dependence of the size of jitter on the average glottal cycle length, and the relation between jitter and laryngeal pathologies. Modeled jitter is discussed in the light of measured jitter, as well as the physiological and statistical plausibility of the model parameters. [DOI: 10.1121/1.1350557]

PACS numbers: 43.70.Bk, 43.70.Gr, 43.72.Ja [AL]

I. INTRODUCTION

Jitter designates feeble random cycle-to-cycle perturbations of the glottal cycle lengths (Lieberman, 1963). The objective of this article is to present data-driven models of vocal jitter, which simulate how disturbances of the instantaneous frequencies of vibration of the glottal walls cause perturbations of the glottal cycle lengths. The modeling combines a parsimonious stochastic model of the disturbances with the so-called ribbon model of vocal fold vibration (Titze, 1989). The disturbance model takes into account vocal microtremor and correlation-free noise.

The modeling enables discerning between jitter-related phenomena that have not always been clearly held apart. A first distinction is the one between disturbances of the instantaneous frequencies of vibration of the glottal walls on the one hand, and perturbations of the glottal cycle lengths on the other. The first causes the second, but they do not, therefore, share the same properties. One goal of this article is therefore to examine within the framework of a model the relation between the properties of the perturbations of the glottal walls and the glottal cycles.

A second distinction is the one between jitter in particular and aperiodicities of the glottal cycle length in general. Indeed, many mechanisms have been documented that may cause the movements of the vocal folds or glottis to be aperiodic. Labelling the acoustic consequences of any of these mechanisms as acoustic jitter or shimmer would be problematic, because the causes of glottal aperiodicities are multiple and not active at the same time. The present model of jitter is therefore also an attempt to improve on the conventional definitions of jitter. The point of view that we adopt is that jitter is a base-line phenomenon that can be understood in terms of a few statistical markers that are present all the time, even in the voices of healthy subjects, and from which other types of aperiodicities must be distinguished, either in size or kind.

Other motivations for developing models of jitter are the following. First, jitter has been described by means of heuristic markers, especially within a clinical framework. The reason is that it has been empirically demonstrated that vocal jitter increases for some dysphonic voices. Models of jitter could suggest or confirm the mathematical form of markers that would characterize perturbed cycle lengths statistically rather than heuristically (Pinto and Titze, 1990; Schoentgen and De Guchteneere, 1995).

Second, models of jitter are part of many speech synthesizers. The models are designed to improve naturalness or mimic hoarse voices (Rozsypal and Millar, 1979; Childers and Wu, 1990; Bangayan *et al.* 1997). Most existing models are only loosely based on data and do not simulate the whole gamut of known properties of jitter.

Third, one utilitarian application of models of jitter is the generation of synthetic signals to test or calibrate signal processing algorithms that detect glottal cycles or measure their lengths (Muta *et al.*, 1988).

To model jitter, one must know its statistical properties. To discover these, one must conceptually isolate speech signal markers that report properties of jitter from markers of other glottal phenomena. One possible criterion is the temporal resolution that is required to observe changes in the glottal cycle lengths. We now turn to a discussion of acoustic excitation phenomena that can be detected at sampling frequencies that are of the orders of magnitude of 1, 10, and 100 kHz, respectively.

[a)]Part of this article is based on a talk given at the 2nd International Conference on Voice Physiology and Biomechanics, Berlin, March 1999.
[b)]Electronic mail: jschoent@ulb.ac.be

Sampling the speech signal at a frequency up to a few kHz would be enough to assess glottal cycle lengths while studying intonation, vibrato, micro-prosody, or pathological voice tremor, which is a quavering intonation in speech (Lebrun *et al.*, 1982; Gamboa *et al.*, 1998). Micro-prosody refers to the modification of the fundamental frequency by phonetic segments; that is, high vowels tend to have a higher fundamental frequency than low vowels and voiced plosive segments tend to have lower fundamental frequencies than the surrounding vowel segments (Petersen, 1978; Ladefoged and Maddieson, 1996). These changes temporally extend over one phonetic segment or more.

Sampling the speech signal at a frequency of the order of magnitude of 10 kHz is adequate for tracking transients, gauging additive noise, and signal typing (Behrman *et al.*, 1998). Transients occur when the laryngeal oscillator switches from nonvibration to vibration or from one mode of vibration to another. Transients must indeed be distinguished from stationary or mildly nonstationary signals. Additive noise owing to turbulent airflow in the glottis or elsewhere in the vocal tract is one source of disturbance of the speech signal that is unrelated to the glottal vibrations per se (Isshiki *et al.*, 1978).

Signal typing, finally, categorizes speech signals according to whether they are single-cycle periodic, multi-cycle periodic, almost periodic, or random (Peterson and Schoup, 1966; Dejonckere, 1985; Titze, 1994b; Behrman *et al.*, 1998).

Normally, signal typing is performed by means of speech signal spectra without referring to vocal fold movement. Formally, magnitude spectra of steady speech signals are either continuous or discrete. Discrete spectra can be made up of several distinct harmonic series whose fundamental frequencies are or are not commensurable.

Type I signals are characterized by a single harmonic series. The signal's cycle lengths are identical. That is, disregarding intonation, vibrato, pathological tremor, or jitter, the glottal signal is periodic with the period being equal to the length of one glottal cycle.

Type II signals are characterized by several harmonic series whose fundamental frequencies are commensurable. These signals are multi-cycle periodic, that is, several unequal glottal pulses are grouped into a larger period (Dejonckere, 1983; Kiritani *et al.*, 1993).

When the spectra are made of partials or harmonic series whose fundamental frequencies are not commensurable, the signals are called quasi-periodic or almost periodic (Moon, 1987). In practice, signal typing does not appear to attempt to distinguish periodic from quasi-periodic signals and to assign the latter to a distinct category. For a discussion of glottal phenomena, which possibly belong to the quasi-periodic category, see Tigges *et al.* (1997).

Finally, type III signals are signals whose spectra are continuous. Continuous spectra are the signature of signals that are random. This would mean that the excitation pulses are all different concerning their shape, length, or amplitude and that the properties of the present excitation pulse cannot be exactly predicted from past pulses. Whether better than chance short-term predictions can be made depends on whether the cycle-to-cycle changes are correlated or not.

One issue, which is relevant to type III signals, is the distinction between chaotic and stochastic signals. Often, random signals are called chaotic when they are known to be solutions of a small system of ordinary differential equations. Deciding whether measured data are chaotic or stochastic can, however, be of considerable practical and theoretical difficulty (Abarbanel, 1996). The application of analysis methods, inspired by nonlinear system dynamics, to the speech signal has so far produced few incontrovertible results (Herzel *et al.*, 1994; Kumar and Mullick, 1996; Behrman and Baken, 1997; Behrman, 1999).

Finally, sampling the speech signal at a frequency of the order of magnitude of 100 kHz and measuring the cycle lengths of type I signals reveals that the individual cycle lengths vary randomly. These random variations have been called jitter. They are typically less than one percent of the average cycle length.

Properties of jitter can only be directly measured in type I (and a few type II) signals that are stationary or mildly nonstationary. Indeed, the recovery of jitter during transients, in speech signals that are severely contaminated by additive noise or in signals that are not type I, are still unsolved problems.

We now discuss properties of jitter with a view to simulating these properties by means of model of the glottal vibration and a model of the stochastic disturbances of the instantaneous frequency of vibration.

(i) The probability distribution of the cycle length perturbations is Gaussian (Horii, 1979; Laver, 1980; Heiberger and Horii, 1982; Pinto and Titze, 1990).

(ii) The size of relative jitter is typically in the range from 0.1% to 1% of the average cycle length. Many authors have confirmed this order of magnitude (Hollien *et al.*, 1973; Horii, 1979; Heiberger and Horii, 1982).

(iii) The perturbations of the lengths of adjacent cycles are positively correlated. This means that the present perturbation depends on past perturbations (Kasuya *et al.*, 1983; Imaizumi, 1986; Schoentgen and De Guchteneere, 1995). We have confirmed these observations by fitting linear autoregressive models to time series of cycle length perturbations. The orders of the models indicate the number of past perturbations that must be taken into account to explain the present perturbation. The order is determined experimentally. In the case of healthy speakers the most frequent order equalled two (De Guchteneere, 1996). This would hint at the existence of a single peak in the spectral contour of the perturbation time series. In a corpus of several hundred glottal length time series that had been obtained for healthy and dysphonic speakers, the range of the model orders was between one and nine, however. A possible reason is that the orders of the models had been estimated by means of the partial autocorrelation coefficients, which may overestimate the number of relevant past perturbations.

(iv) The most likely source of the observed correlation between the perturbations of the lengths of neighboring cycles is microtremor. Vocal microtremor has been defined by some as variations of the fundamental frequency of speech sounds, whose modulation frequency is between 7

and 10 Hz, but which is often less or more (Schultz-Coulon *et al.*, 1979). Quantitative data on microtremor are scarce. We hereafter focus on microtremor data published by Winholtz and Ramig for 12 healthy speakers (1992). According to these authors, the modulation frequency ranged from 4 to 10 Hz with median values at 6 Hz for male speakers and 5 Hz for female speakers. The modulation level ranged from 0.5% to 2.8%. Loosely speaking, the modulation frequency determines the temporal extent of the correlation between the perturbations, and the modulation level the size of the perturbations.

(v) Jitter appears to be a genuine stochastic phenomenon. We have indeed shown that no nonlinear relations could be found in time series of perturbations, from which any linear correlation had been removed beforehand (Schoentgen and De Guchteneere, 1996). This suggests that linear time series models are adequate for representing jitter. It follows that jitter must be modeled stochastically, because linear systems only output random signals when they are driven by a stochastic input. This is indirectly confirmed by a physiological model that causally links the statistical properties of muscle twitches to some of the disturbances of the instantaneous frequency of vibration (Titze, 1991, 2000).

(vi) On average, cycle lengths perturbations increase with the cycle lengths, that is, the longer the cycles the greater the jitter (Hollien *et al.*, 1973; Koike *et al.*, 1977; Horii, 1979).

(vii) It has been repeatedly claimed that jitter increases in the presence of at least some laryngeal pathologies. Published results have substantiated this claim. However, in earlier studies, the distinction between cycle length variability in type I, type II, and type III signals has not always been made clear enough. Results we have obtained by means of decorrelated cycle lengths perturbations appear to confirm that jitter tends to increase in at least some speakers suffering from laryngeal pathologies (Koike *et al.*, 1977; Davis, 1979; Laver *et al.*, 1986; Kasuya *et al.*, 1986; Schoentgen and De Guchteneere, 1997).

(viii) One assumption, which is made implicitly by most authors, is that meaningful statistics of jitter can be obtained from one sustained voiced speech waveform (Atkinson, 1976). Formally, this would suggest that the time series of the cycle lengths is ergodic. This means that the perturbation time series must be stationary and not decomposable into different stochastic processes that do not share similar statistical properties (Ventsel, 1973).

(ix) Raw and decorrelated perturbations are of the same order of magnitude. This means that the standard deviation of the decorrelated perturbations is typically greater than 0.1% and less than the standard deviation of the raw perturbations. Empirical comparisons suggest that the standard deviation of decorrelated jitter is roughly two-thirds the standard deviation of raw jitter (De Guchteneere, 1996).

(x) Finally, the disturbances that cause the glottal cycle length perturbations do not occur in synchrony with the glottal cycles. This is an issue only because most existing models of jitter do synchronously reset the fundamental frequency contour or glottal cycle length once per cycle. A possible reason is that most of the models have been developed to drive speech synthesizers whose control parameters are the so-called fundamental frequency or glottal cycle length (Wendahl, 1966; Coleman, 1971). The glottal cycle length is indeed a control parameter that must be updated synchronously. The reason is that the glottal period or fundamental frequency is only meaningfully defined for periodic signals. For aperiodic signals, the so-called fundamental frequency must be replaced by the instantaneous frequency.

Existing models of the glottal cycle perturbations are reviewed hereafter. Most of the models reproduce a small subset of the properties that have been listed previously. Some of the models ignore what is known about jitter and generate synthetic perturbations with characteristics that have not been observed. None of the models attempts to explain properties of jitter, with one possible exception (Titze, 1991, 2000). This model, however, does not appear to concern perturbations of the glottal cycle lengths as such. Instead, it attempts to explain some properties of the disturbances of the instantaneous frequency of vibration. The author concluded that, in steady vowels, disturbances of the order of 0.2% to 1.2% may be caused by unfused tetani of the thyroarytenoid or cricothyroid muscles.

Synthetic perturbations have been either deterministic or random. Deterministic perturbations alternate between constant negative and positive values (Milenkovic, 1987). Random perturbations have been drawn from a uniform probability distribution (Qi and Shipp, 1992) or a Gaussian one (Rozsypal and Millar, 1979; Bangayan *et al.*, 1997). The size of the perturbations has been controlled by scaling the standard deviation of the probability distributions or, in one instance, by filtering (Lalwani and Childers, 1991).

A few authors have included a relation between the perturbations of adjacent cycles. These simulations have been carried out by means of one of the following:

(i) copy synthesis, that is, the insertion of measured length perturbations into the synthesizer (Bangayan *et al.*, 1997; Childers and Wu, 1990; Hillenbrand, 1987);
(ii) one-over-*f* noise, that is, noise whose power spectrum falls as the inverse of the frequency (Hillenbrand, 1988); and
(iii) a linear auto-regressive model of the normalized samples of the fundamental frequency contour. The model had been fitted to measured cycle lengths beforehand (Imaizumi and Gauffin, 1992).

In this article, we propose stochastic models that simulate jitter via the disturbances of the instantaneous frequencies of selected points of the right and left glottal walls. Each model involves two submodels. The first is the model of the stochastic disturbances. The second is a model of the vibrating glottis, which is an instance of the so-called ribbon model (Titze, 1984, 1989; Childers *et al.*, 1986; Childers, 2000).

In Sec. II A, we derive particular instances of the general ribbon model; in Sec. II B, we introduce the models of the stochastic disturbances of the instantaneous frequencies of vibration; in Sec. II C, we combine the ribbon models with the disturbance models into four intermediate models of in-

creasing complexity. The purpose is to take advantage of the closed-form solutions of simple models to provide insight into the behavior of more complicated models. Finally, we present a model that combines the characteristics of the four submodels that have been introduced previously with a model of vocal microtremor. The results section presents the simulation results obtained by means of this model and the discussion section addresses the interpretation of the simulation results in terms of known properties of jitter.

II. MODELS

A. The ribbon model of vocal fold vibration

Titze (1984, 1989) has proposed a model of the undisturbed periodic timing of the glottal width. This model is known by some as the ribbon model (Childers, 2000). Titze and Childers have shown by means of simulations that the timing of the glottal width agrees with the timing of the glottal area, vocal fold contact area, and electroglottogram, which records the electrical impedance of the neck. The glottal cycle lengths and their perturbations are indeed frequently measured by means of the electroglottogram (Abberton and Fourcin, 1997). We have confirmed that the cycle length disturbances obtained by means of the speech signal and electroglottographic signal share the same statistical properties and demonstrate the same temporal patterns (De Guchteneere, 1996; Schoentgen and De Guchteneere, 1991, 1995, 1997)

The ribbon model consists of a static prephonatory and a kinematic sinusoidal component that together describe the normalized glottal hemi-width *h*. Normalization implies that the width has been divided by the constant amplitude of vibration of one glottal wall:

$$h(y,z,t)=h_0+\sin(\pi y/L)\sin[2\pi f_0(t-z/c)]. \tag{1}$$

Symbols *h*, *y*, and *z* designate the spatial coordinates in the lateral, anterior–posterior, and inferior–superior directions. Symbol *t* is the continuous time coordinate, symbol *c* is the speed of the surface wave in the cover of the vocal fold, symbol f_0 designates the frequency of vibration of one glottal wall or vocal fold, and symbols L and h_0 designate the vibrating vocal fold length and normalized prephonatory glottal hemi-width.

In the framework of the ribbon model, the movements of the interpenetrating left and right glottal walls are periodic and mirror images of each other. This means that the movements of the left and right walls are assumed to be coupled. Otherwise, any minute dissimilarity would cause the phases of the left and right folds to drift apart.

The rightmost sine function in model (1) takes into account that the most common vocal fold movement is the superposition of a left–right movement in the lateral direction and a diverging–converging movement in the inferior–superior direction. The phase delay Φ between the movements of the lower and upper margins of the glottis is the following (Titze, 1989):

$$\Phi=2\pi f_0 T/c. \tag{2}$$

Symbol T is the vibrating vocal fold thickness. Vibrating vocal fold length and thickness depend on the frequency of vibration f_0 via the following expressions:

$$L=0.038L_0^{1.6}f_0^{0.5}, \quad T=T_0L_0/L.$$

The values of constants L_0 and T_0 are 1.6 cm and 0.45 cm for males and 1.0 cm and 0.35 cm for females (Titze, 1989). Model (1) may be interpreted as a parametrized solution of a many degrees-of-freedom model of the vibrating vocal folds. Apart from two anatomical constants (T_0 and L_0), the model is characterized by three parameters, which are frequency f_0, phase delay Φ, and, implicitly, the coupling between the movements of the left and right glottal walls.

Hereafter, we use four simplified instances of the ribbon model. From the first to the fourth they relate the glottal width to the following quantities:

(a) the phase of the movement of one rigid glottal wall;
(b) the phases of the lower and upper margins of one glottal wall;
(c) the phases of the left and right rigid glottal walls; and
(d) the phases of the lower and upper margins of the left and right glottal walls.

(a) The simplest model is obtained by assuming that the glottis is a right prism, the glottal walls are rigid, and the movements of the left and right walls mirror each other. This means that phase delay Φ is zero and the glottal hemi-width does not depend on glottal depth *y*. Consequently, the timing of the normalized glottal width *h* is described by a single phase coordinate θ_h. The constant 2 enables taking into account that the total width is twice the hemi-width:

$$h(\theta)=2h_0+2\sin(\theta), \quad \theta_h=\theta. \tag{3}$$

(b) Models and data confirm that the most common movements of the vocal folds consist of a superposition of a left–right movement in the lateral direction and converging–diverging movement in the inferior-superior direction. This means that the movements of the inferior (index *i*) and superior (index *s*) margins of the glottis must be described by two different phases:

$$h_i(\theta_i)=2h_0+2\sin(\theta_i), \quad h_s(\theta_s)=2h_0+2\sin(\theta_s), \quad \theta_{hi}=\theta_i, \quad \theta_{hs}=\theta_s. \tag{4}$$

(c) Because it is unlikely that the disturbances of the left and right glottal walls are exact mirror reflections of each other, glottal widths (3) or (4) must be rewritten considering different disturbances to the left (index *l*) and right (index *r*). The constant phase π takes into account that the undisturbed left and right are reflections of each other in the mid-sagittal plane. Expression (5b) is obtained from expression (5a) by means of elementary trigonometric relations. Expression (5c) follows from expression (5b). It establishes the link between the left and right phases and the phase of the glottal width:

$$h(\theta_h)=2h_0+\sin(\theta_r)-\sin(\theta_l+\pi), \tag{5a}$$

$$h(\theta_h)=2h_0+2\cos\left(\frac{\theta_r-\theta_l}{2}\right)\sin\left(\frac{\theta_r+\theta_l}{2}\right), \tag{5b}$$

$$\theta_h = \frac{\theta_r + \theta_l}{2}. \tag{5c}$$

(d) When the differences between the movements of the left and right walls and lower and upper margins of the glottis are taken into account, then models (4) and (5) are combined into model (6):

$$\begin{aligned} h_i(\theta_{hi}) &= 2h_0 + \sin(\theta_{i,r}) - \sin(\theta_{i,l} + \pi), \\ h_s(\theta_{hs}) &= 2h_0 + \sin(\theta_{s,r}) - \sin(\theta_{s,l} + \pi), \\ \theta_{hi} &= \frac{\theta_{i,r} + \theta_{i,l}}{2}, \quad \theta_{hs} = \frac{\theta_{s,r} + \theta_{s,l}}{2}. \end{aligned} \tag{6}$$

Instances (3)–(6) of the ribbon model relate the phases of the vibrations of selected points of the glottal walls to the phase of vibration of the glottal width. By definition, the instantaneous frequency is the rate of change of the phase and the glottal cycle length is the time interval over which the phase increases by 2π. Consequently, the relations between instantaneous frequency f, glottal cycle length τ_0, and fundamental frequency f_0 are the following:

$$\frac{d\theta}{dt} = 2\pi f, \quad \int_{\theta_{\text{start}}}^{\theta_{\text{start}} + 2\pi} d\theta = 2\pi = \int_{t_{\text{start}}}^{t_{\text{start}} + \tau_0} 2\pi f dt, \quad f_0 = \frac{1}{\tau_0}.$$

B. Models of the stochastic disturbances of the instantaneous frequency

A cyclic movement must be described by means of the instantaneous frequency, which agrees with the fundamental frequency only when the movement is periodic. We here discuss parsimonious hypotheses regarding the stochastic disturbances of the instantaneous frequencies of vibration of the glottal walls, owing to correlation-free noise and microtremor. Correlation-free or white noise processes are sequences of uncorrelated random variables e_n that are identically distributed. One of the simplest white noise processes is the two-point process (7). Symbol p designates the probability and symbol Δ a constant, which is later interpreted as a time step.

$$e_n = \sqrt{\Delta} \begin{cases} +1, & p = 0.5, \\ -1, & p = 0.5. \end{cases} \tag{7}$$

One can easily check that the mean M of the two-point process is equal to 0 and the variance equal to Δ.

The autocovariance and autocorrelation functions of the two-point process involve the mean of the product $(e_n - M)(e_{n+k} - M)$. The autocovariance function of the two-point process therefore equals 0 when index $k > 0$ and Δ when $k = 0$. Indeed when indexes m and n differ, the product $e_n e_m$ is also distributed according to a two-point process whose mean and variance are equal to 0 and Δ^2. When the indices are identical, the product's average and spread are Δ and 0.

In the Introduction we pointed out that in vowels sustained by healthy speakers without vibrato the properties of raw jitter are conditioned by the so-called microtremor. Microtremor must be distinguished from pathological tremor that marks the voices of subjects suffering from neurological diseases, such as Parkinson's disease or essential tremor.

Winholtz and Ramig (1992) have published spectra of microtremor signals. For healthy speakers, these spectra display a single spectral peak. The position of the peak is the microtremor frequency and its width the microtremor bandwidth. We have inferred a value of 4 Hz for the bandwidth from two published microtremor spectra. The bandwidth controls the shape of the synthetic microtremor signal. The smaller the bandwidth is, the smoother and the more sinusoidallike is the microtremor.

Filtering white noise by means of a linear second-order autoregressive model is a direct method for generating signals that, as the microtremor signal, are characterized by a single-peak power spectrum. Symbol y designates the synthetic microtremor signal, a_i the constant filter parameters, e the input white noise, and b the peak value of the white noise. Index n is the time step. Symbols Ψ and B, finally, designate the microtremor frequency and bandwidth in radians per sample. Coefficient a_0 normalizes the gain of the filter so that it does not depend on the frequency or bandwidth. The filter coefficients are the following (Steiglitz, 1996):

$$\begin{aligned} y_n &= a_1 y_{n-1} + a_2 y_{n-2} + a_0 b e_n, \\ R &= 1 - \frac{B}{2}, \quad \cos\varphi = \frac{2R}{1 + R^2} \cos\Psi, \\ a_0 &= (1 - R^2)\sin\varphi, \quad a_1 = \frac{4R^2}{1 + R^2}\cos\Psi, \quad a_2 = -R^2. \end{aligned} \tag{8}$$

An alternative to the linear second-order autoregressive model (8) is the linear first-order model (9), for which the spectral peak is always positioned at 0 Hz. Coefficient a_1 fixes the decay of the spectral contour with increasing frequency. Parameter b is the crest value of the input white noise process:

$$y_n = a_1 y_{n-1} + b e_n. \tag{9}$$

Interpretations of expressions (8) and (9) are the following. When the white noise input to filter (9) is a two-point process (7), the output is a sum of decaying exponential curves whose initial values stochastically switch between $\pm b$. Similarly, when the white two-point process is input to the second-order linear filter, the output is a stochastic sum of cosine functions that decay exponentially.

C. Stochastic models of jitter

We here combine instances of the ribbon model and models of the stochastic disturbances that have been discussed in Secs. II A and B. The goal is to make explicit relations that exist between the stochastic disturbances of the phases of vibration of the glottal walls on the one hand and the stochastic perturbations of the glottal cycle lengths on the other.

We examine hereafter five models of jitter. The first four are models for which closed-form solutions exist. The fifth can only be solved numerically. The objectives of the study

of the five models are the following. The Roman numerals refer to the observed properties of jitter that are listed in the introduction section.

(1) Model I comprises a correlation-free disturbance of the instantaneous frequency of vibration of one rigid glottal wall. The goals of the modeling are as follows. First, show that disturbances of the instantaneous frequency cause perturbations of the glottal cycle lengths (properties *i*, *ii*). Second, relate the size of the length perturbations to the cycle length (property *vi*). Third, demonstrate that the cycle length perturbations are stationary and ergodic in practice (property *viii*).
(2) In model II the correlation-free disturbance is replaced by artificial microtremor (9). The purpose is to demonstrate that microtremor correlates the perturbations of neighboring glottal cycles (properties *iii* and *iv*).
(3) Model III simulates the coupling between the left and right glottal walls, which are not disturbed identically. The goals are the following. First, show that the coupling synchronizes the movements of the left and right glottal walls and prevents their phases from drifting apart. Second, demonstrate that asymmetric coupling boosts the phase perturbations. Asymmetric coupling is therefore one possibility via which laryngeal pathologies might influence the size of jitter (property *vii*).
(4) Model IV includes a delay between the lower and upper glottal boundaries. The goal is to show that the delay correlates the perturbations of adjacent cycle lengths (property *iii*).
(5) Model V combines the properties of the previous four models with microtremor model (8). The purpose is to simulate properties (*i*)–(*x*) of observed jitter.

1. Model I: a random walk model of jitter

A simple model of jitter is obtained by assuming that the glottis is a right prism that evolves owing to the movement of one glottal wall. This means that the time-dependent glottal width (3) evolves sinusoidally. The phase of the glottal width and the phase of the glottal wall agree. The temporal evolution of the disturbed phase must be written as follows. Symbol f designates the instantaneous frequency and symbol θ the phase:

$$\frac{d\theta}{dt}=2\pi f. \tag{10}$$

Expression (10) can be rewritten because jitter is small (property *ii*):

$$\frac{d\theta}{dt}=2\pi(f_0+b\xi). \tag{11}$$

Symbol f_0 designates the unperturbed instantaneous frequency of vibration, b is a small constant, and ξ is white noise that simulates stochastic disturbances.

Expression (11) is a heuristic only because no ordinary function exists whose derivative is a white noise process. Expression (11) must therefore be transformed into a stochastic differential equation (12). Symbol dt is an ordinary differential and symbol dW a stochastic differential, which has mathematical properties different from an ordinary differential (Gardiner, 1983):

$$d\theta=2\pi(f_0dt+bdW). \tag{12}$$

Equation (12) can be turned into a stochastic difference equation that can be solved numerically or algebraically. Symbol Δ designates the time step and n the index. Symbol Δ_W is the stochastic difference:

$$\begin{aligned}&\theta_{n+1}=\theta_n+2\pi(f_0\Delta+b\Delta_W),\\&\theta_{n+1}=\theta_n+2\pi f_0\Delta+2\pi b\sqrt{\Delta}\begin{cases}+1, & p=0.5,\\-1, & p=0.5.\end{cases}\end{aligned} \tag{13}$$

Kloeden and Platen (1999) have shown that the stochastic difference can be realized by means of a two-point process (7). The condition is that the statistical properties of the solution of discrete-time equation (13) are required to approximate the statistical properties of the solution of continuous-time equation (12). Equation (13) can be solved algebraically by repeated substitutions:

$$\theta_n=2\pi f_0n\Delta+2\pi b\sum_{i=0}^{n}e_n,\quad \theta_N=2\pi,\quad \tau=N\Delta, \tag{14}$$

$$\begin{aligned}&\tau=\tau_0-\tau_0b\sum_{i=0}^{N}e_n,\quad M(\tau)=\tau_0,\\&V(\tau)=(\tau_0b)^2N\Delta\approx\tau_0^3b^2.\end{aligned} \tag{15}$$

By definition, glottal cycle length τ equals the number N of time steps that are needed to raise the phase by 2π. Length τ is therefore obtained by inserting N for n and 2π for θ in solution (14). Mean $M(\tau)$ and variance $V(\tau)$ are obtained by means of the probability distribution (7) and the rules that obtain the mean and variance of a linear function of a random variable (Ventsel, 1973).

The variance is a measure of absolute jitter squared and the coefficient of variation is a measure of relative jitter. The coefficient of variation is the standard deviation divided by the mean. The average cycle length is equal to the unperturbed instantaneous cycle length τ_0. The variance of the cycle lengths is proportional to the cube of the average cycle length. The reason is that in practice the perturbed length differs from the average length by 1% at most. The coefficient of variation therefore depends on the square root of the average cycle length when peak value b is a constant.

Further, the mean and variance of the cycle lengths are constants when the small disparities between perturbed and average cycle lengths are disregarded. Consequently, the time series of the cycle lengths perturbations is stationary (property *viii*). This means that within the framework of model I, the cycle length time series is ergodic because the only source of disturbance is two-point process (7) (Breiman, 1986).

Finally, the autocovariance equals the average of the product of two adjacent perturbations, which is zero. The reason is that each length is perturbed by sum of correlation-free disturbances and the ranges of the sums of two neighboring lengths do not overlap.

2. Model II: A filtered noise model of jitter

In this section we calculate the perturbations of the cycle lengths when the instantaneous frequency is disturbed by an instance of artificial microtremor instead of correlation-free noise. For simplicity's sake, we consider synthetic microtremor (9), whose autocorrelation is a decaying exponential and whose spectral peak is always positioned at 0 Hz. Synthetic microtremor (8), whose autocorrelation function is a decaying cosine function and whose main spectral peak is positioned at the microtremor frequency, is studied by means of numerical simulations in the framework of model V.

$$\theta_{n+1}=\theta_n+2\pi f_0\Delta+2\pi y_n,$$
$$y_{n+1}=ay_n+be_n, \quad 0<a<1,$$
$$\theta_n=\theta_0+2\pi f_0\Delta n+2\pi\sum_{i=1}^{n} y_{i-1}, \tag{16}$$
$$\tau=\tau_0-\tau_0\sum_{i=1}^{N} y_{i-1}=\tau_0-\tau_0 b\sum_{i=1}^{N}\sum_{s=1}^{i} a^{i-s}e_{s-1},$$
$$M(\tau)=\tau_0, \quad V(\tau)\approx\frac{\tau_0^3 b^2}{1-a^2}.$$

Equations (16) can be solved by repeated substitutions. As for model I, the perturbed cycle length is obtained by inserting 2π for phase θ, N for index n, and zeroing θ_0 in solution θ_n. Mean $M(\tau)$ and variance $V(\tau)$ are obtained by means of the rules that relate the mean and variance of a random variable to the mean and variance of a linear function of the same random variable (Ventsel, 1973). Variance $V(\tau)$ is an approximation because the terms that were not proportional to the cycle length have been discarded:

$$C(\tau_1,\tau_2)=\frac{\tau_0^2 b^2\Delta a(1-a^{N1})(1-a^{N2})(1-a^{N1+1})}{(1-a)^3(1+a)}. \tag{17}$$

Autocovariance (17) of the perturbations of adjacent lengths $N1$ and $N2$ has been obtained by means of the expression for cycle length τ. The approximate lag one autocorrelation is obtained by dividing autocovariance C by variance V. The autocorrelation therefore increases with coefficient a and decreases with length N. That is, the autocorrelation is the stronger the slower the decay of the external disturbances and the shorter the cycle.

3. Model III: A disturbed coupled oscillator model of jitter

Most of the *in vitro*, mechanical, and numerical models of the vocal folds simulate the movement of one fold only. The hypothesis that is implicitly made is that the movement of the first fold exactly mirrors the movement of the second. This hypothesis is plausible only when small disparities between left and right are dampened. Expressions (5) suggest that otherwise the smallest difference would cause an amplitude modulation of the glottal width even in the absence of external disturbances.

Since one cannot assume that stochastic disturbances of the left and right instantaneous frequencies mirror each other, we include a coupling between the phases of the left and right glottal walls. Symbol K is the coupling constant. The coupling term is the sine function of the subtraction of the left from the right phase. One reason for including the sine function is that a shift by 2π of one phase has no observable consequences.

Taking into account that $\sin x\sim x$ when x is small, one sees that when the left phase is greater than the right phase the coupling term decreases the left and increases the right phase and vice versa:

$$\theta_{l,n+1}=\theta_{l,n}+2\pi f_0\Delta-K\sin(\theta_{l,n}-\theta_{r,n})+2\pi be_{l,n},$$
$$\theta_{r,n+1}=\theta_{r,n}+2\pi f_0\Delta-K\sin(\theta_{r,n}-\theta_{l,n})+2\pi be_{r,n}. \tag{18}$$

Within the framework of model III, we assume that the disturbances are neither spatially nor temporally correlated. Symbols e_l and e_r therefore designate uncorrelated random variables that are distributed according to probability law (7).

Expression (5b) shows that the timing of the glottal width depends on the product of a cosine and a sine function. The argument of the cosine function is the difference and the argument of the sine function is the sum of the left and right phases. Equations (18) are therefore rewritten as a sum and a difference:

$$h_n=2h_0+2\cos\left(\frac{\theta_{r,n}-\theta_{l,n}}{2}\right)\sin\left(\frac{\theta_{r,n}+\theta_{l,n}}{2}\right), \tag{5b}$$
$$\frac{\theta_{r,n+1}+\theta_{l,n+1}}{2}=\frac{\theta_{r,n}+\theta_{l,n}}{2}+2\pi f_0\Delta+2\pi b\frac{e_{l,n}+e_{r,n}}{2}, \tag{19a}$$
$$\frac{\theta_{r,n+1}-\theta_{l,n+1}}{2}=\frac{\theta_{r,n}-\theta_{l,n}}{2}-K\sin(\theta_{r,n}-\theta_{l,n})+2\pi b\frac{e_{r,n}-e_{l,n}}{2}. \tag{19b}$$

The sum of the phases is the argument of the sine function in glottal width (5b). Equation (19a) shows that the sum does not depend on the coupling when the coupling is symmetric. Equation (19a) is formally identical to the coupling-free model I.

The difference of the phases is the argument of the cosine function in glottal width (5b). Equation (19b) shows that the difference depends on the coupling between the left and right. Any contrasts between the coupled and uncoupled models are therefore due to the cosine function. We examine hereafter under what conditions the influence of the cosine function is small. Formally, Eq. (19b) and its solution are the following, taking into account that $\sin x\sim x$ when x is small:

$$\alpha_n=\frac{\theta_{r,n}-\theta_{l,n}}{2}, \quad \alpha_{n+1}=\alpha_n(1-2K)+2\pi b\frac{e_{l,n}-e_{r,n}}{2},$$
$$\alpha_n=\alpha_0(1-2K)^n+\left[\sum_{i=1}^{n}(1-2K)^{n-i}2\pi b\frac{e_{r,i}-e_{l,i}}{2}\right].$$

Solution α_n shows that when the model is disturbance-free, the steady-state solution is $\theta_l=\theta_r$ provided that the coupling constant is positive and less than 0.5. When the disturbances are finite, the right-hand side of the solution must be taken into account, which is a weighted sum of the past and present

disturbances. The sum decreases when the coupling constant increases. In the framework of the model, the argument of the cosine function is therefore smaller the stronger the coupling and the smaller the disturbances.

A generalization of the previous model is one for which the left and right coupling constants are not identical. The sum and difference of the left and right phases are then as follows, assuming that $\sin x \sim x$. Equations (20) show that $\theta_l = \theta_r$ is a stable solution when the external disturbances are zero. The condition is that the sum of the coupling constants is less than one:

$$\begin{aligned}\frac{\theta_{l,n+1}+\theta_{r,n+1}}{2} &= \frac{\theta_{l,n}+\theta_{r,n}}{2} + 2\pi f_0\Delta - (K_l - K_r) \\ &\quad \times \frac{(\theta_{l,n}-\theta_{r,n})}{2} + 2\pi b\,\frac{e_{l,n}+e_{r,n}}{2}, \\ \frac{\theta_{l,n+1}-\theta_{r,n+1}}{2} &= \frac{\theta_{l,n}-\theta_{r,n}}{2} - (K_l+K_r)\,\frac{(\theta_{l,n}-\theta_{r,n})}{2} \\ &\quad + 2\pi b\,\frac{e_{l,n}-e_{r,n}}{2}.\end{aligned} \tag{20}$$

The sum and difference of the phases determine the timing of the glottal width via relation (5b). For asymmetric coupling, the sum contains an additional term that depends on the difference between the coupling constants. Since the cosine term in the glottal width is nearly a constant when the coupling is strong and the disturbances weak, the coupling is expected to have a substantial influence on jitter in the asymmetric case only.

4. Model IV: A disturbed vocal fold wave model of jitter

Within the framework of the ribbon model the movements of the upper and lower margins of the glottis may be out of phase. Model IV simulates the consequences of the out-of-phase movements, assuming that the external disturbances of the upper and lower margins are identical and correlation-free. Symbol D designates the delay in number of time steps between the phases of the inferior and superior margins that are indexed i and s:

$$\begin{aligned}\theta_{i,n+1} &= \theta_{i,n} + 2\pi f_0\Delta + 2\pi b e_n, \\ \theta_{s,n+1} &= \theta_{i,n-D} + 2\pi f_0\Delta + 2\pi b e_n.\end{aligned} \tag{21}$$

The solutions (22) of Eqs. (21) are the following:

$$\begin{aligned}\theta_{i,n} &= 2\pi f_0 n\Delta + 2\pi b\sum_{j=1}^{n} e_{j-1}, \\ \theta_{s,n} &= 2\pi f_0 (n-D)\Delta + 2\pi b\sum_{k=1}^{n-D} e_{k-1} + 2\pi b\sum_{l=1}^{n} e_{l-1}.\end{aligned} \tag{22}$$

Conventionally, the cycle lengths are measured by monitoring markers in the speech signal of the closing of the glottis. Most of the time, the upper margin of the glottis opens and closes after the lower margin. We therefore define the cycle length as the number of time steps required to raise by 2π the phase of the upper margin. Cycle length τ is therefore obtained by inserting $N = n - D$ and $\theta = 2\pi$ in the equation that describes the movement of the upper margin:

$$\begin{aligned}&\tau = \tau_0 - b\tau_0\sum_{k=0}^{N} e_k - b\tau_0\sum_{l=D}^{N+D} e_l, \quad M(\tau) = \tau_0, \\ &V(\tau) \approx 2b^2\tau_0^3, \quad C(\tau_1,\tau_2) \approx b^2\tau_0^2 D\Delta.\end{aligned} \tag{23}$$

Length τ shows that the cycle lengths are perturbed by two random walks. The first is the direct perturbation that is the same for the lower and upper margins. The second is the disturbance of the lower margin, which is delayed. Mean M and variance V are obtained by means of the conventional rules (Ventsel, 1973). Compared to the rigid-wall model, the approximate variance of the length perturbations doubles owing to the coexistence of direct and delayed disturbances of the glottal wall. Autocovariance C is different from zero because it involves the product of the sum of two random walks that partially overlap. One sees that the autocovariance increases with delay D. The correlation owing to the overlap cannot extend further than one cycle, however, because the delay cannot exceed the length of a cycle when the disturbances are white noise.

5. Model V: A stochastic model of jitter

The following model combines the properties of the models that have been discussed so far with model (8) of the vocal microtremor. From top to bottom, Eqs. (24) describe the phases of the lower and upper, left and right glottal margins. The two last equations are the left and right synthetic microtremor, which is modeled by means of linear second-order autoregressive model (8). The complete disturbances are the sum of synthetic microtremor and correlation-free noise that simulates stochastic disturbances other than microtremor. The peak values of the synthetic microtremor and correlation-free noise are controlled independently:

$$\begin{aligned}\theta_{i,l,n+1} &= \theta_{i,l,n} + 2\pi f_0\Delta - K_l \sin(\theta_{i,l,n} - \theta_{i,r,n}) \\ &\quad + y_{l,n} + 2\pi b_2 e_{2,l,n}, \\ \theta_{i,r,n+1} &= \theta_{i,r,n} + 2\pi f_0\Delta - K_r \sin(\theta_{i,r,n} - \theta_{i,l,n}) \\ &\quad + y_{r,n} + 2\pi b_2 e_{2,r,n}, \\ \theta_{s,l,n+1} &= \theta_{i,l,n-D} + 2\pi f_0\Delta - K_l \sin(\theta_{s,l,n} - \theta_{s,r,n}) \\ &\quad + y_{l,n} + 2\pi b_2 e_{2,l,n}, \\ \theta_{s,r,n+1} &= \theta_{i,r,n-D} + 2\pi f_0\Delta - K_r \sin(\theta_{s,r,n} - \theta_{i,l,n}) \\ &\quad + y_{r,n} + 2\pi b_2 e_{2,r,n}, \\ y_{l,n+1} &= a_1 y_{l,n} + a_2 y_{l,n-1} + 2\pi a_0 b_1 e_{1,l,n+1}, \\ y_{r,n+1} &= a_1 y_{r,n} + a_2 y_{r,n-1} + 2\pi a_0 b_1 e_{1,r,n+1}.\end{aligned} \tag{24}$$

Equations (24) must be solved numerically. The existence of stable solutions of the undisturbed equations can, however, be ascertained algebraically by means of two auxiliary functions (Kelley and Peterson, 2001). The existence of stable undisturbed solutions is enough to infer the existence of stable disturbed solutions when the disturbances are small. The auxiliary functions are the following. For simplicity's

sake, the coupling constants are assumed to be identical:

$$V_n=(\theta_{i,l,n}-\theta_{i,r,n})^2+(\theta_{s,l,n}-\theta_{s,r,n})^2,$$

$$W_n=(\theta_{i,l,n}+\theta_{i,r,n}-4\pi f_0 n\Delta)^2 + (\theta_{s,l,n}+\theta_{s,r,n}-4\pi f_0 n\Delta)^2.$$

When the coupling constants are less than $\frac{1}{2}$, one can show by direct substitution that initial phases exist for which $V_{n+1}-V_n\leqslant 0$ for all time steps n. Function V therefore evolves towards zero, which means that the left and right phases evolve towards the same values.

Similarly, one can show by direct substitution that $W_{n+1}-W_n=0$. This means that function W is a constant. From the ancillary equations $W=\text{const}$ and $V=0$ one can easily deduce that the undisturbed instantaneous frequencies of the glottal margins eventually become identical and equal to f_0.

To discover the admissible range of values of the parameters K, a_1, and a_2, the linearized difference equations (24) are recast into an algebraic form by means of a one-sided z-transform. The linearization is legitimate when the disturbances are small and $\sin x$ can be replaced by x. The admissible range of parameters is the range for which the phases do neither exponentially diverge to infinity nor converge to a constant, which would mean that the oscillator spontaneously accelerates or decelerates. The corresponding mathematical conditions are that no poles of the z-transformed phases lie outside the unit circle and that at least one pole lies on the unit circle (Hamilton, 1994). The solutions of the linear system of algebraic equations are given by the cofactors of the matrix of the system of equations divided by determinant d. The poles of the solutions are therefore the values of the variable z that zero the determinant d, which is given hereafter:

$$d=\frac{(z^2-a_1z-a_2)^2(z+2K)(z-1)(z-1+2K)}{z},$$

$$0\leqslant K<\frac{1}{2},\quad \left|\frac{a_1\pm\sqrt{a_1^2+4a_2}}{2}\right|<1.$$

As expected, one root is equal to 1. The inequalities involving the coupling constant and the model coefficients of the microtremor arise because of the condition that the other roots must lie outside the unit circle. When the left and right coupling constants are not identical, the condition is that their sum must be less than one.

III. METHODS

A. Numerical simulations

Model V was solved numerically. The cycle lengths were determined by monitoring the sum of the left and right phases of the upper margin of the glottis. Every 2π, the count of the time steps giving the cycle length was reset, and the precision of the count was improved by interpolating linearly the glottal width to determine the instant of the width becoming zero.

To gauge the noise floor of the simulations, the microtremor was zeroed and the crest value of the correlation-free disturbances was fixed at 5×10^{-6}. Then 50 simulations were carried out with a reseed of the random number generator for each. The average relative jitter was equal to $0.017\%\pm0.008\%$. This means that the noise floor was at least ten times smaller than typical decorrelated jitter.

TABLE I. Default values (third column) and range (fourth column) of the values of the modeling parameters in the framework of the linear multiple regression analyses.

Average cycle length (s)	τ_0	0.01	0.005–0.01
Coupling constant (right)	K_r	0.1	0–0.1
Difference (right-left)	abs(K_r-K_l)	0	0–0.1
Wave velocity (cm/s)	c	200	200–400
Microtremor frequency (Hz)	Ψ	6	1–8
Microtremor bandwidth (Hz)	B	4	2–6
Peak value of driving noise	b_1	0.25	0.1–0.5
Peak value of white disturbances	b_2	0.05	0.001–0.1
Time step (s)	Δ	10^{-5}	

The sizes of time step Δ, driving peak value b_1 of the microtremor model, microtremor frequency Ψ and bandwidth B, coupling constants K, peak value b_2 of the correlation-free disturbances, surface wave speed c, and undisturbed cycle length τ_0 must be fixed for each simulation. When not stated otherwise, these parameters assumed the default values that are given in Table I.

1. Time step, Δ

The size of jitter observed in sustained speech typically ranges from 0.1 to 1% of the average cycle length. Therefore, typical sampling frequencies of 50 kHz or better are needed when measuring or simulating jitter (Titze *et al.*, 1987; Schoentgen and De Guchteneere, 1991). Accordingly, we fixed the time step for the numerical simulations at 10^{-5} s.

2. Correlation-free disturbance amplitude, b_2

The peak value of the correlation-free stochastic disturbances was fixed so that the coefficient of variation of the decorrelated cycle lengths was between 0.1% and 1%. Informal comparisons of natural raw and decorrelated jitter we performed earlier suggest that the standard deviation of the raw jitter was very roughly one and a half times the standard deviation of decorrelated jitter (De Guchteneere, 1996). Accordingly, we empirically fixed the default value of b_2 at 0.05.

3. Microtremor frequency, Ψ

The microtremor frequency was fixed taking into account data published by Winholtz and Ramig (1992). For healthy speakers, the authors reported microtremor frequencies ranging from 4.7 to 10.6 Hz. The disparity between male and female speakers was not statistically significant. We therefore fixed the default value of the microtremor frequency at 6 Hz, which was the median value of the male speakers.

4. Microtremor bandwidth, B

The bandwidth fixed the narrowness of the peak in the spectral contour of the microtremor. We inferred a value of 4 Hz from two microtremor spectra published by Winholtz and Ramig (1992). We confirmed this value by means of a perturbation spectrum published by Titze (1994a). The width of the main spectral peak 3 dB below the summit was between 3 and 4 Hz, taking into account the precision of the graph. Accordingly, we fixed the default bandwidth at 4 Hz.

5. Amplitude, b_1

Amplitude b_1 codetermined the modulation level of the cycle length. Winholtz and Ramig reported a modulation level range from 1.1% to 2.8% for healthy speakers. The gap between male and female speakers was not statistically significant. Accordingly, we fixed the default value of amplitude b_1 at 0.25, which produced synthetic modulation levels in the range measured by Winholtz and Ramig. The amplitude of the driving noise also depended on the tremor frequency and bandwidth via the gain of the microtremor model (8). The gain was calculated so as to keep the height of the microtremor peak constant when microtremor frequency and bandwidth changed.

6. Coupling constants, K

The algebraic solutions we have obtained for model IV and the algebraic criteria for the existence of stable solutions we have established for model V show that the sum of the left and right coupling constants must be less than one. Accordingly, we fixed the default values for both at 0.1.

7. Velocity, c

The default velocity of the surface wave was fixed at 200 cm/s (Titze, 1989). The delay in radians between the movements of the upper and lower margins of the glottis was determined via relation (2).

8. Undisturbed glottal cycle length, τ_0

The undisturbed cycle length was fixed to agree with values typical of male speakers. The default value was 0.01 s.

B. Statistical analyses and markers

1. Linear autoregressive analysis

Because the length perturbations are correlated owing to microtremor the standard deviation of the raw cycle lengths was not adequate for statistically describing raw perturbations. Fitting an autoregressive linear model (25) to the time series and calculating the standard deviation of the correlation-free residual was done instead (Schoentgen and De Guchteneere, 1995):

$$(\tau_i-\tau_0)=\alpha_1(\tau_{i-1}-\tau_0)+\alpha_2(\tau_{i-2}-\tau_0)+\cdots+\alpha_M(\tau_{i-M}-\tau_0)+\alpha_0 e_i\,. \quad (25)$$

Symbol τ designates the cycle lengths and τ_0 their average. The calculation of AR (autoregression) coefficients α_i was carried out by means of conventional linear methods (Box and Jenkins, 1970; Markel and Gray, 1976). The order M of the AR model was fixed as follows. The analysis was started with the order set to 10. When the value of the Mth AR coefficient was less than $\pm 2/\sqrt{N}$, the order was decremented by one and the analysis was started anew. Parameter N was equal to 100, that is, the number of cycles. The decrementing of the order stopped when the Mth AR coefficient exceeded the critical threshold or the order was equal to zero. This conventional method tended to overestimate the model order because typically one AR coefficient out of 20 may exceed the critical threshold for statistical reasons alone (Chatfield, 1984).

The spectrum of the AR coefficients was a model of the spectral contour of the raw perturbations. The bandwidths and frequencies of the poles of AR model (25) were therefore estimates of the bandwidths and frequencies of the spectral peaks. They were obtained by numerically calculating the roots of the polynomial formed by means of the AR coefficients (Markel and Gray, 1976). The pole or pole-pair with the narrowest bandwidth was kept and the others were discarded, because the pole with the narrowest bandwidth had the greatest influence on the shape of the spectral contour model. One pole was selected because within the framework of model V only the frequency of one real pole or one complex-conjugate pole pair could be interpreted in terms of the microtremor frequency.

The estimates of the microtremor frequency obtained by means of the autoregressive model are in radian per sample. To turn radians per sample into Hz, a value must be assigned to the sampling frequency, which here was not a constant because of the cycle length perturbations. Since these were small, the sampling frequency was assigned the value of the average frequency. The expected error was equal to the standard deviation of the raw jitter, that is, less than 1%.

The so-called residue, that is, the decorrelated perturbations, was obtained by subtracting the right-hand side of Eq. (25) from the left-hand side. The coefficient of variation of the residue was a measure of the relative decorrelated jitter.

2. Modulation level

The modulation level was a measure of the magnitude of the cycle length perturbations, owing to microtremor. The modulation level was the difference between the maximum deviation and the average divided by the average cycle length times 100 (Winholtz and Ramig, 1992). Since the modulation level measured the effect of the slower multicycle perturbations, we smoothed the raw perturbations by means of a second-order linear smoother to reduce the influence of the faster cycle-to-cycle perturbations owing to correlation-free disturbances (Geckinly and Yavuz, 1981).

3. Variogram analysis

The stationarity and whiteness of the perturbation time series was checked graphically and numerically by means of the variogram, which is a ratio of two variances. The dividend was the variance of the differences between two per-

turbations *m* cycles apart. The divisor was the variance of the difference between perturbations one cycle apart (Box and Luceno, 1997). Shift *m* grew from 0 to 30.

Theoretically speaking, the variogram of white noise was flat and equal to 1; the variogram of correlated stationary noise initially grew and subsequently flattened; the variogram of a nonstationary time series was never flat. In the case of stationary time series, variogram and correlogram were equivalent (Box and Luceno, 1997). We preferred the variogram to the correlogram because it gave rise to graphs that were less visually cluttered when more than one time series was involved.

4. Multiple linear regression analysis

We performed multiple linear regression analyses to study the magnitude and sign of the dependence of the statistical markers of jitter on the model parameters. In multiple linear regression, the value of one dependent variable depends linearly on several independent variables (Ross, 1987). The independent variables were divided into two groups according to whether they were parameters of the external disturbances or parameters of the ribbon model. The dependent variables were successively the coefficient of variation of the residue of the AR model (that is, the relative decorrelated jitter), the order of the AR model, the frequency of the main spectral peak of the AR model, the modulation level, and the lag one autocorrelation coefficient of the raw perturbations.

The parameters were varied in five steps between extreme values that are given in Table I. Accordingly, each regression analysis involved one dependent variable, and 625 cycle length time series that were obtained by varying four independent variables in five steps between their extremal values. The results section reports the standardized regression coefficients because they can be compared directly in the framework of the same regression analysis.

IV. RESULTS

The objective was to simulate perturbed glottal cycle lengths whose statistical properties agreed with the observed properties. Models I–IV that are described in the modeling section were based on different assumptions. Models I and II respectively involved disturbances by correlation-free and correlated noise of the instantaneous frequency of vibration of one rigid glottal wall. Model III involved correlation-free stochastic disturbances of the coupled left and right glottal walls. Finally, model IV involved correlation-free stochastic disturbances of the upper and lower glottal margins, which moved out of phase.

Model I generated length perturbations that were stationary, uncorrelated, and depended on the average cycle length. Model II produced stationary length perturbations that depended on the average cycle length and were correlated. Model III generated stationary perturbations that depended on the average cycle length and on the coupling strength between the glottal walls. Model IV produced stationary perturbations that depended on the average length and were correlated due to the out-of-phase movement of the upper and lower margins of the glottis. The purpose of these four models was to obtain closed-form expressions of the effects of some external disturbances on the ribbon model of the glottis.

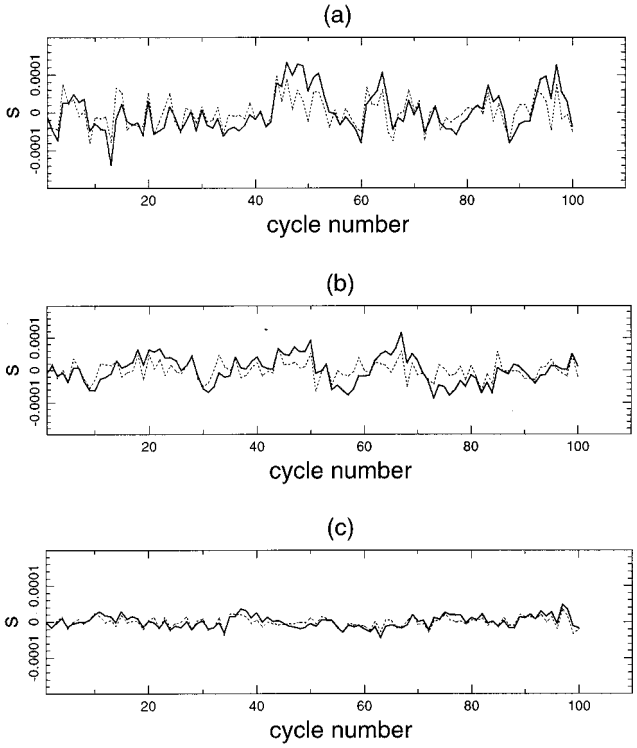


FIG. 1. Three glottal cycle length time series simulated via stochastic model V. The average cycle lengths were from top to bottom equal to 0.01, 0.0075, and 0.005 s. The continuous lines are the unprocessed time series, the dotted lines are the decorrelated series.

The rest of the results section is devoted to model V, which must be solved numerically. Model V involved disturbances by correlation-free noise and simulated microtremor of the left and right glottal walls, whose upper and lower margins moved out of phase. Model V was shown to generate cycle length perturbations whose statistical properties agreed with the specifications listed in the Introduction. Table I lists the default values of the parameters of model V, together with the ranges of the values of the parameters when linear multiple regression analyses were performed.

Figure 1(a) shows a raw cycle length time series overlaid on the corresponding decorrelated time series. The decorrelated time series was obtained by fitting a linear autore-

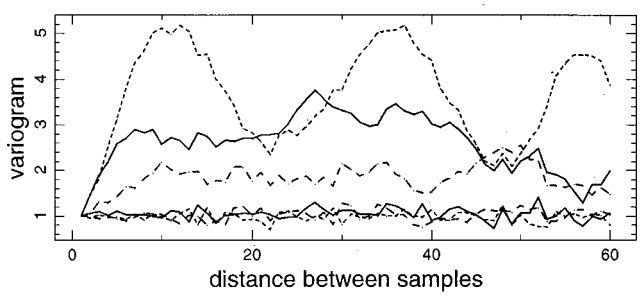


FIG. 2. Variograms of the unprocessed and decorrelated time series reproduced in Fig. 1. The continuous, dotted, and dashed lines respectively correspond to the top, middle, and bottom series in Fig. 1.

TABLE II. Average variogram values of unprocessed and decorrelated glottal cycle lengths for lags equal to 10, 20, and 30 cycles. The number of time series equalled 50. The model parameters assumed their default values. The averages in the last row attest to the whiteness, and the independence of the averages from the lags attests to the stationarity of the decorrelated cycle length perturbations.

Intercycle distance	10	20	30
Variogram raw length perturbations	3.7 ± 1.2	2.9 ± 0.8	2.9 ± 0.9
Variogram decorrelated length perturbations	1.0 ± 0.1	1.0 ± 0.1	1.0 ± 0.1

gressive model and filtering the raw time series. Figure 1(b) shows a time series for an average cycle length of 0.0075 s and Fig. 1(c) for an average of 0.005 s. The coefficient of variations of the decorrelated time series respectively equalled 0.37%, 0.36%, and 0.27%. Time series shown in Fig. 1 were typical of model V and their statistical properties agreed with the properties of observed perturbations.

Plots of the histograms of the decorrelated cycle perturbations suggested that they were distributed normally. This was confirmed by plotting the observed cumulative relative frequency versus the expected cumulative frequency. The data then clustered tightly around a straight line bisecting the graph.

Figure 2 shows the variograms of the raw and decorrelated cycle length series of Fig. 1. The decorrelated time series gave rise to variograms whose trends were flat and tightly clustered around one. The slight growth of the variability of the variograms towards the right is a consequence of the decreasing number of cycles that were involved in the computation of the variance. The variograms of the raw length series tended towards or oscillated about a value different from one.

This was confirmed by simulating a 50 cycle length time series and calculating the variogram values for lags equal to 10, 20, and 30 cycles. Table II shows that the decorrelated series gave rise to variograms that did not evolve with the intercycle distance and which were equal to one with a coefficient of variation equal to or less than 10%. The variogram values of the raw lengths tended towards a value of 2.9 with a coefficient of variation equal to or less than 31%.

Table III reports the results of five multiple linear regression analyses. The independent variables were the average cycle length, the right coupling constant, the difference between the left and right coupling constants, and the speed of the surface wave. The independent variables that assumed their default values were the peak value of the correlation-free stochastic disturbances, the microtremor bandwidth, the microtremor frequency, and the peak value of the driving white noise of the vocal microtremor model.

The dependent variables were the decorrelated jitter in percent (that is, the coefficient of variation of the AR model residue), the modulation level in percent, the lag one autocorrelation coefficient, the order of the AR model, and the main pole frequency of the AR model (that is, the frequency of the real pole or complex-conjugate pole pair with the smallest bandwidth).

Table III is organized as follows. Each column corresponds to one dependent variable and summarizes 625 simulations. The second to fourth rows give the first quartile, median, and third quartile of the histogram of each dependent variable.

Rows 5–8 give for each dependent and independent variable the statistically significant standardized regression coefficients. The $(+)$ and $(-)$ signs indicate that the dependent variable respectively rose or fell with an increasing independent variable. The null hypothesis was that the regression coefficients were equal to zero. The probability that the null hypothesis was rejected while true was always less than 0.1%. The regression coefficients that were not statistically significant are omitted from Table III.

The last row gives the multiple correlation coefficient values. The null hypothesis was that the correlation coefficient was equal to zero. The probability that the null hypothesis was rejected while true was always less than 0.1%. Correlation coefficients that were not statistically significant were omitted from the table.

Table III shows the following. First, the median main pole frequency (5.8 Hz) was an estimate of the vocal microtremor frequency, which equalled 6 Hz. Second, decorrelated jitter and modulation level depended on the average cycle length, the coupling constants, and their difference. Third, the lag one correlation between the perturbations increased with the average cycle length. Fourth, the AR model

TABLE III. Linear multiple regression analyses of the statistical markers (first row) as a function of the ribbon model parameters (first column). The ranges of the parameters are given in Table I. The quartiles of the histograms of the statistical markers are reported in rows 2–4. The statistically significant standardized regression coefficients are listed in rows 5–8 and the multiple correlation coefficient in the last row. All coefficients are statistically significant at the 0.1% level at least.

	Decorrelated jitter (%)	Modulation level (%)	Lag one autocorrelation coefficient	AR model order	Pole frequency (Hz)
First quartile	0.34	0.88	0.53	2	0
Median	0.40	1.13	0.62	5	5.87
Third quartile	0.46	1.42	0.69	8	6.84
τ_0	$+0.74$	$+0.61$	$+0.44$		
K_r	-0.37	-0.23			
$\text{abs}(K_r-K_l)$	$+0.23$	$+0.15$			
c					
R	0.86	0.67	0.44		

TABLE IV. Linear multiple regression analyses of the statistical markers (first row) as a function of the ribbon model parameters (first column). The ranges of the parameters are given in Table I. The quartiles of the histograms of the statistical markers are reported in rows 2–4. The statistically significant standardized regression coefficients are listed in rows 5–7 and the multiple correlation coefficient in the last row. All coefficients are statistically significant at the 0.1% level at least, with one exception (marked with an asterisk). The standardized coefficient in row 7 was significant at the 1% level only. The difference between Tables III and IV is that for the latter, the left and right coupling constants are identical.

	Decorrelated jitter (%)	Modulation level (%)	Lag one autocorrelation coefficient	AR model order	Pole frequency (Hz)
First quartile	0.31	0.86	0.51	2	0
Median	0.36	1.07	0.61	5	5.71
Third quartile	0.43	1.27	0.68	7	6.77
τ_0	+0.90	+0.60	+0.44		
K					
c			−0.20(*)		
R	0.90	0.60	0.49		

order and the main pole frequency did not depend significantly on any ribbon model parameter.

Table IV is organized as Table III. The difference between Tables III and IV is that for Table IV the simulations were performed with equal left and right coupling constants. Consequently, each column summarizes 125 simulations. The observed ranges of the dependent variables were very similar to those in Table III. Two differences were observed regarding the statistical significance of the regression coefficients. First, the decorrelated jitter did not depend on the coupling constants. Second, the lag one correlation depended on the speed of the surface wave.

Table V reports the results of five linear multiple regression analyses. The independent variables were the microtremor frequency, the microtremor bandwidth, the peak value of the driving noise of the vocal microtremor model, and the peak value of the correlation-free disturbances. The independent variables that assumed their default values were the average cycle length, the left and right coupling constants, and the speed of the surface wave.

The dependent variables were, as previously, the decorrelated jitter in percent, the modulation level in percent, the lag one autocorrelation coefficient, the order of the AR model, and the frequency of the main pole of the AR model.

Table V is organized as follows. Each column corresponds to one dependent variable and summarizes 625 simulations. Rows 2–4 report the first quartile, median, and third quartile of the histograms of the dependent variables. Rows 5–8 give for each dependent and independent variable the statistically significant standardized regression coefficients. The null hypothesis was that the regression coefficients were zero. The probability that the null hypothesis was rejected while true was less than 0.1%. One exception was the AR model order, whose regression coefficient was significant at the 1% level only. The regression coefficients that were not statistically significant were omitted from Table V.

The last row gives the multiple correlation coefficient values. The null hypothesis was that the correlation coefficient was equal to zero. The probability that the null hypothesis was rejected when true was less than 0.1%. One exception was the AR model order, whose correlation coefficient was significant at the 1% level only. Correlation coefficients that were not statistically significant were omitted from the table.

Table V shows that decorrelated jitter and frequency modulation level increased with the microtremor peak value, bandwidth and frequency, and the correlation-free stochastic disturbances.

The lag one correlation increased with the microtremor

TABLE V. Linear multiple regression analyses of the statistical markers (first row) as a function of the parameters of the microtremor model and correlation-free stochastic disturbances (first column). The ranges of the parameters are given in Table I. The quartiles of the histograms of the statistical markers are reported in rows 2–4. The statistically significant standardized regression coefficients are listed in rows 5–8 and the multiple correlation coefficient in the last row. All coefficients were statistically significant at the 0.1% level at least, with two exceptions (marked with asterisks). The standardized regression coefficients in column 5 were significant at the 1% level only.

	Decorrelated jitter (%)	Modulation level (%)	Lag one autocorrelation coefficient	AR model order	Pole frequency (Hz)
First quartile	0.22	0.81	0.25	2	0.00
Median	0.42	1.25	0.60	4	3.49
Third quartile	0.63	1.84	0.94	6	6.53
Ψ	+0.1	+0.18			+0.32
B	+0.1	+0.19	+0.09		
b_1	+0.26	+0.76	+0.69	+0.12(*)	+0.13
b_2	+0.93	+0.27	−0.55		
R	0.98	0.85	0.89	0.15(*)	0.35

parameters and decreased with increasing correlation-free stochastic disturbances.

Also, the main pole frequency increased with the microtremor frequency and the peak value of the correlation-free stochastic disturbances. Finally, the order of the autoregressive model increased slightly with the size of the microtremor.

The previous simulations were repeated with a fixed microtremor bandwidth. Each regression analysis therefore involved 125 simulations. The ranges of the dependent variables were very similar to those in Table V. The results are therefore not reported here. Regarding the statistical significance of the regression coefficients, the only difference was that the main pole frequency did not depend on the peak value of the noise driving the vocal microtremor model.

V. DISCUSSION

The discussion section is devoted to a comparison of simulated and observed jitter and an examination of the effects on simulated jitter of the parameters of the ribbon model and model of the external disturbances.

According to the list of attributes reported in the Introduction, the salient properties of jitter that must be addressed are the stochasticity of the cycle length perturbations, the stationarity, the asynchrony, the sizes of decorrelated jitter and modulation level, the Gaussian distribution of the decorrelated perturbations, the cycle-to-cycle correlation of the perturbations, the microtremor frequency, and the increase of the perturbations owing to laryngeal pathologies.

A. Stochasticity

We have posited that jitter is a stochastic phenomenon. Indeed, synthetic cycle perturbations are the collective outcome of instantaneous disturbances owing to correlation-free noise and vocal microtremor simulated by a linear AR model. The deterministic alternative would have consisted in modeling jitter as the outcome of the interaction of a few ordinary variables.

The statistical scenario is the most likely for the following reason. We have shown elsewhere that linear autoregressive models can be successfully fitted to observed perturbation series. Filtering the perturbations by means of the autoregressive model decorrelates the perturbations and obtains the so-called residual. Analyses of the autocorrelation have confirmed that the residuals were free of linear cycle-to-cycle dependencies (Schoentgen and De Guchteneere, 1995). Later, we processed the residuals to detect any remaining nonlinear relations. Six residuals, out of a total of 265, displayed nonlinear relations (Schoentgen and De Guchteneere, 1996). In five, the nonlinear relations could be traced to quantization noise; one case remained unexplained. The failure to detect any nonlinear dependencies in the decorrelated length perturbations suggested that the perturbations might be simulated within the framework of linear models. In linear models, however, randomness can only arise stochastically. The stochastic scenario also agrees with the muscle twitch model (Titze, 1991).

Obviously, the stochastic framework only applies to the minute cycle-to-cycle perturbations observed in stationary type I signals, for instance. It does not inevitably apply to the macroscopic random perturbations of acoustic excitation pulse size and length that characterize type III signals.

B. Stationarity

Stationarity means that the joint statistical distribution of the perturbations does not evolve in time (Breiman, 1986). The solution of model I suggests that asking whether the cycle perturbations are stationary is distinct from asking whether the disturbances of the glottal walls are. Indeed, the synthetic length perturbations may be stationary, even if the phase disturbances are not. The explanation involves the definition of the glottal cycle length, which equates the cycle length to the temporal interval required to increase the oscillator's phase by 2π. After one cycle, the sum of the cumulative increments of the phase is reset, with the consequence that in successive cycles the statistical properties of the perturbations stay practically the same, because all the cycles are almost identical.

This dichotomy between phase and length can be illustrated explicitly by means of formulas (14) and (15), which refer to model I. The rightmost term in formula (14) is the collective disturbance of the glottal phase. The variance of this term is proportional to temporal index n, which increases monotonically. In formulas (15), the rightmost term in the expression of the cycle length τ are the perturbations whose variance V is proportional to the number N of temporal steps within one cycle. The number of steps N was approximately constant because the cycle lengths perturbations were less than 1%. Consequently, jitter was constant in practice even though the cumulative disturbances of the glottal phase increased monotonically.

Lack of stationarity of the phase disturbances and stationarity of the cycle length perturbations was a dichotomy typical of model I. In model III, on the contrary, the stationarity of the disturbances of the phase difference arose due to the coupling between left and right that dampened the effect of past disturbances.

The stationarity of the synthetic length perturbations generated by model V were mathematically tested by means of variograms. Figure 2 and Table II show that the variograms of the decorrelated perturbations had a flat trend and an average equal to one. The flat trend demonstrates that the length series were stationary and the value one that they were decorrelated.

The variograms of the raw perturbation series were, however, more variable. In the raw perturbation series, the effects of synthetic microtremor and correlation-free stochastic disturbances coexisted. The increased variability of the variograms of raw series appeared to be due to the shortness of the records rather than to lack of stationarity. The reason is that time series that can be represented by linear autoregressive models are stationary when the roots of the autoregressive models are outside the unit circle (Kelley and Peterson, 2001).

Length of record, stationarity, and ergodicity are issues because in practice jitter is measured by means of one or a feeble number of vowels that are sustained for typically 1 s. Ergodicity is the property that statistical markers of an en-

semble of records can be inferred by means of a single record that is long enough. Obviously, one does not expect the property of ergodicity to hold when the properties of the vocal folds change noticeably between recordings. However, even if the properties did not change one could not automatically assume ergodicity to be true. For instance, Titze (1994a) reported the perturbation series of a speaker sustaining a vowel for 10 s. The corresponding correlogram had not dropped to zero after a lag of almost 600 cycles. This would suggest that at least that record was the realization of a stochastic process that was not ergodic (Hayes, 1996; Breiman, 1986). Since determining whether a statistical process is ergodic is not practical in most circumstances, one cannot decide at present whether the intra-recording-session variability that has been reported in the literature is a consequence of the lack of ergodicity, the shortness of the records, or a lack of robustness of the signal processing. Figure 2 and Table II suggest that decorrelating the perturbation time series has a beneficial effect on the interrecord variability of decorrelated jitter anyhow.

C. Asynchrony

Asynchrony was not a serious issue because no evidence existed that the cycle perturbations were triggered by recurring glottal events. The most parsimonious assumption was therefore that the phase disturbances did not occur in synchrony with the glottal cycles.

The reason synchrony was a point that must be raised is that most existing simulations of jitter updated the length once for each glottal cycle. A possible explanation is that in most speech synthesizers the relevant control parameters were the so-called fundamental frequency or cycle length, instead of the instantaneous frequency or instantaneous length. Indeed, if the average fundamental frequency was continually perturbed, the perturbations would grow over time and generate nonstationary speech signals, even if the average fundamental frequency was a constant. To avoid this, the cycle length and fundamental frequency are updated in most speech synthesizers in synchrony with the glottal cycle.

D. Correlation

Correlation numerically expressed the strengths of linear relations between the perturbations of neighboring glottal cycle lengths. Such relations have been studied only occasionally. Imaizumi (1986) reported correlograms of cycle lengths, which suggest that oscillatory patterns exist in some cycle length sequences. It was not reported, however, whether the related speech signals were type I or type II. Winholtz and Ramig (1992) measured microtremor by means of purpose-built hardware. The microtremor frequency and modulation level in healthy speakers were typically equal to 6 Hz and 1.2%. Schoentgen and De Guchteneere (1995, 1997) have fitted a linear autoregressive model to observed cycle lengths to mathematically represent and remove correlation. The purpose was to decorrelate perturbations to isolate the random component of jitter. The observed time series and the AR model orders indirectly confirmed the findings of Winholtz and Ramig.

In this article, the correlation of the perturbations of abutting cycles has been studied in a closed form by means of model II and numerically by means of model V. The main difference between models II and V was that model II was disturbed by random decaying exponential curves and model V by random exponentially decaying oscillations (microtremor) and correlation-free noise. Formula (17), which describes model II, shows that the autocovariance increased with an increasing coefficient a. The bigger this coefficient was the slower was the decay of the exponentials. Consequently, the slower the decay and the shorter the cycles were the further extended the correlation.

Within the framework of model V, vocal microtremor was simulated by means of a linear second-order AR model. The characteristics of the model were chosen so as to reproduce published microtremor data (Winholtz and Ramig, 1992). Synthetic and observed vocal microtremor therefore shared the same spectral properties. Vocal microtremor was the only perturbation of the cycle lengths that was reported in the literature and that was known to correlate the perturbations over many cycles.

Tables III–V report the results of multiple regression analyses performed on time series obtained by means of model V. One dependent variable that is relevant to the discussion of correlation is the lag one autocorrelation coefficient of the raw perturbations (column 4). The model parameters that significantly influenced lag one autocorrelation were, in decreasing order of importance, the peak value of the microtremor driving noise, the peak value of the correlation-free stochastic disturbances, the average cycle length, the speed of the surface wave, and the microtremor bandwidth.

The lag one autocorrelation coefficient increased with the increasing peak value of the driving noise of the microtremor model. The reason is that the random microtremor oscillations biased over consecutive cycles the external disturbances towards slightly positive or negative values. This bias increased with the peak value of the driving noise. Therefore, the lag one correlation increased with the size of the microtremor.

The lag one correlation predictably declined when the peak value of the correlation-free stochastic disturbances increased. The correlation-free disturbances did not produce any correlation in the cycle perturbations and their increase therefore diminished the existing correlation owing to vocal microtremor.

A possible explanation for the dependence of the lag one correlation on the average cycle length is the following. Long cycles favored the accumulation of disturbances owing to correlation-free disturbances and vocal microtremor. The correlation between neighboring cycles increased as a corollary because vocal microtremor slightly biased the perturbations towards positive or negative values, a bias that typically extended over several cycles. The longer the cycles were, the stronger was the accumulated bias in the observed perturbations and the stronger was the correlation. This ef-

fect can be seen when comparing Figs. 1(a) and (c), which show the perturbation series for average lengths equal to 0.01 and 0.005 s.

Table IV suggests that an increasing speed of the surface wave reduced the lag one autocorrelation coefficient. According to formula (2), an increasing speed reduced the delay between the upper and lower glottal margins. The simulation therefore agreed with autocovariance (23) that describes model IV. Formula (23) predicts that the lag one autocovariance drops with a decreasing delay between the upper and lower margins of the glottis.

The regression of the lag one autocorrelation coefficient on the speed is barely statistically significant in Table IV and not statistically significant in Table III. We have confirmed the feeble statistical significance of the influence on the correlation of the surface wave speed by performing simulations during which the speed was increased in 30 steps from 200 to 400 m/s and in 50 steps from 100 to 400 m/s. The regression of the lag one autocorrelation coefficient on the speed of the surface wave was not statistically significant. This suggests that in the framework of model V, the speed of the surface wave was not an important parameter as far as the statistical properties of the cycle perturbations were concerned. A possible explanation is that past disturbances owing to temporal overlap were masked by present disturbances with which they coexisted within the framework of model V.

Table V indicates that the lag one correlation of the raw perturbations grew slightly with the microtremor bandwidth. Since the modulation level also grew with the bandwidth, the explanation is the same as the one given previously in relation with the average cycle length. The reason the modulation level increased with the bandwidth is explained in a subsequent section.

E. Vocal microtremor frequency

The sixth columns in Tables III–V report quartiles and regression coefficients of the main pole frequency. The AR model is a model of the contour of the magnitude spectrum of the cycle length perturbations and the main pole frequency was the frequency of the spectral peak with the smallest bandwidth.

Table III shows that the main pole frequency was an estimate of the microtremor frequency. Indeed, the first quartile was equal to zero, the median equal to 5.87 and the third quartile equal to 6.47 Hz, for a synthetic microtremor frequency of 6 Hz. Table IV, which reports the same simulations with fewer independent variables, confirms this.

Table V (sixth column) shows that the pole frequency depended, in decreasing order of importance, on the vocal microtremor frequency and the peak value of the driving noise of the microtremor model.

The first quartile of the main pole frequency was zero. This means that at least 25% of the perturbation sequences were modeled by an AR model whose most prominent spectral peak was positioned at zero frequency (the origin of the frequency axis), while the synthetic microtremor frequency equalled 6 Hz. Possible explanations are the following. First, the AR models were fitted by means of a least squares method. The positions of the poles in the z-plane depended on the sampling frequency, which here was equal to the average fundamental frequency. As a rule, the higher the sampling frequency was the closer the poles were to the real axis of the z-plane (that is, to zero frequency). The ratio of the microtremor frequency to the sampling frequency was indeed 6% when the fundamental frequency was 100 Hz and 3% when the fundamental frequency was 200 Hz. One should also note that with a fundamental frequency equal to 200 Hz and a microtremor frequency equal to 6 Hz, the positioning of the spectral peak of the modeled contour at zero instead of 6 Hz corresponded to an error of 0.06π radians only.

Second, the nefarious effect of a high fundamental frequency on the possibility of estimating the microtremor frequency by means of an AR model was compounded by the spectral peak owing to microtremor being less salient when the fundamental frequency increased, for reasons that have been explained in the previous paragraphs. The less prominent the vocal microtremor spectral peak was the more difficult it was to fit it by AR models. This also explained the dependence of the pole frequency on the driving noise of the microtremor model (Table V, sixth column). An increasing driving noise increased the modulation level and therefore the prominence of the spectral peak at the microtremor frequency. This made it easier for the AR model to fit the spectral contour accurately and therefore less likely that the main AR model peak was assigned to 0 Hz.

Third, in Tables III–V only the frequency of the most prominent peak of the AR model is reported. The fifth columns show that the AR model orders were typically between two and eight. This means that the AR models were characterized by more than one spectral peak. Consequently, it happened that a pole pair that was not reported in the tables modeled the microtremor spectral peak, because the corresponding peak was not the narrowest one. Selecting the relevant poles of a fitted AR model to estimate the microtremor frequency was a problem similar to selecting the relevant poles of a fitted linear predictive model to estimate the formant frequencies of the speech signal (Markel and Gray, 1976).

When the purpose of AR model fitting was to decorrelate the length perturbations, selecting a model order that was higher than necessary had no nefarious consequences. The reason was that excess poles were characterized by bandwidths that were large and the spectral peaks therefore contributed little to the model of the spectral contour.

F. AR model order

Tables III–V show that the AR model order did not depend on the parameters of the ribbon model or external disturbances, except a feeble dependence on the peak value of the driving white noise of the microtremor model. A possible explanation of this exception is the following. Increasing the driving noise increased the modulation level (Table V, third column), which was a measure of the amplitude of the random oscillations owing to synthetic microtremor. An increased modulation level eased the fitting of AR models with pole pairs that agreed with the spectral peaks owing to

vocal microtremor. The AR model order therefore increased for reasons that have been explained previously.

A visual inspection of the observed raw perturbation series suggests that a similar phenomenon might explain the increase of the AR model order in the case of dysphonic voices. We do not claim, however, that the increase of the order was always a direct consequence of increased tremor. Instead, we observed that the perturbation time series of dysphonic voices were occasionally mildly nonstationary, with a weak trend towards longer cycles at the end of the utterance (De Guchteneere, 1996). Such a feeble lack of stationarity led in practice to AR models with somewhat increased orders.

G. Size of decorrelated jitter

Relative decorrelated jitter was quantitated by means of the coefficient of variation in percent of the decorrelated cycle perturbations. Tables III–V (second columns) attest that the size of the coefficient of variation depended on the following model parameters in increasing order of importance: the vocal microtremor bandwidth and frequency, the peak value of the driving noise of the microtremor model, the coupling constants between the left and right glottal walls, the average cycle length, and the peak value of the correlation-free stochastic disturbances.

The increase of the size of decorrelated jitter with the peak value of the correlation-free stochastic disturbances was a direct consequence of the accumulation of many disturbances over one cycle. The stronger the external disturbances were, the greater was the decorrelated jitter.

Most authors agree that absolute perturbations increase with the average cycle length (Horii, 1979). Yet, the mathematical form of the relation between decorrelated perturbations and average cycle length is unknown. Possible reasons are that, first, different heuristic markers have been used in different studies; second, correlation owing to microtremor as well as to trends owing to declination or pathological tremor have often been taken into account in a cursory fashion only; and third, observed perturbations decline for numerical reasons with the cycle length when the sampling frequency is low. It is therefore not known whether relative jitter depends on the average cycle length. Experiments reported by Horii (1979) showed that over 12 increasing fundamental frequency levels, the relative perturbations increased four times and decreased six times. The overall trend was weakly rising, possibly owing to a sampling frequency that was too small to capture the perturbations of high-pitched voices.

Tables III and IV and formulas (15) and (16) show that within the framework of models I, II and V, the absolute perturbations increased with the average cycle length and therefore decreased with an increasing fundamental frequency. The reason was that the longer the cycles were, the more stochastic disturbances contributed to one cycle perturbation.

Tables III and IV show that the synthetic relative perturbations moderately increased with the average cycle length and therefore moderately decreased with a raising fundamental frequency. The difference between the first and third quartiles of the coefficients of variation was -0.12% when the fundamental frequency increased from 100 to 200 Hz. The smallness of this reduction and the results reported by Horii (1979) suggest that the assumption that was made in models I–V of the peak value of the correlation-free stochastic disturbances as independent of the instantaneous frequency was adequate.

If future observations disclosed that absolute perturbations fall and relative perturbations rise with increasing fundamental frequency, one would have to conclude that correlation-free stochastic disturbances feebly increase with the instantaneous frequency of vibration.

Table III (sixth row) shows within the framework of model V that decorrelated jitter decreased when the coupling between the left and right glottal walls was strengthened. A possible explanation is suggested by model III. The solution of model III shows that the coupling constant dampened the contribution of the past disturbances to the perturbation of the interglottal distance. The same solution attests that if the ribbon model was disturbance-free the coupling constant would warrant that the left and right glottal phases were identical.

Table III (seventh row) also discloses that when the coupling between the left and right glottal walls was not symmetrical the decorrelated jitter increased with the difference. Lack of symmetry means that the influence of the left on the right was not identical to the influence of the right on the left.

When comparing Tables III and IV one sees that the statistically significant influence of the coupling constants was possibly the indirect consequence of the inequality between the left and right coupling constants. Table IV indeed reports the same simulation experiments as Table III. The difference was that the left and right coupling constants were equal. Then the influence of the coupling on the perturbations was statistically not significant. The reason for this may be discussed in the framework of model III. Mainly, the symmetric coupling had a negligible influence owing to the smallness of the disturbances and the perfect left–right symmetry of the ribbon model and its initial conditions. In real larynxes, however, the properties of the glottal walls and their initial conditions of vibration are never perfectly identical. Therefore, the coupling between the left and right cannot be neglected.

Table V (second column) shows that the influence of the peak value of the driving noise of the microtremor influenced the decorrelated jitter much less than the correlation-free stochastic disturbances. This was because the random cyclic disturbances owing to microtremor were removed from the decorrelated jitter. Decorrelated jitter was therefore mainly influenced by the disturbances that were random and correlation-free.

Finally, Table V suggests that decorrelated jitter was feebly influenced by the microtremor bandwidth and frequency. For the bandwidth, a possible explanation is the following. The gain of the microtremor model (8) was fixed so that the spectral maximum was equal to one, whatever the shape of the spectral contour (Steiglitz, 1996). This means that when the bandwidth increased the spectral peak broadened, but the height of the summit stayed at the same value.

A consequence is that the microtremor spectral energy increased, which explains the increase of the perturbations with the bandwidth.

A possible explanation for the feeble influence of the microtremor frequency on the size of decorrelated jitter is the following. When the microtremor frequency decreased, the spectral peaks owing to the two complex-conjugate poles moved towards each other and boosted each other's peak amplitude. Since the model gain was calculated so as to keep the peak height constant, the gain was corrected downwards and consequently the microtremor energy fell with falling microtremor frequency. Therefore, the small influence of bandwidth and frequency on the sizes of the perturbations might have been the consequence of a modeling option of the microtremor. This option could, however, not be omitted since the omission would have replaced a controlled variation of the gain by an uncontrolled one.

H. Size of the modulation level

The modulation level was the maximal deviation of the smoothed perturbations from the average divided by the average cycle length (Winholtz and Ramig, 1992). The modulation level was a measure of the strength of the vocal microtremor.

According to Tables III–V (third columns) the modulation level depended, in increasing order of importance, on the following model parameters: the microtremor bandwidth and frequency, the coupling constants, the peak value of the random correlation-free disturbances, the average glottal cycle length, and the peak value of the driving white noise of the microtremor model.

Possible explanations for the dependence of the modulation level on microtremor bandwidth, microtremor frequency, coupling constants, and average glottal cycle length are the same as for decorrelated jitter. These explanations have been given previously.

One quantitative difference between modulation level and decorrelated jitter was that the modulation depended strongly on the peak value of the driving noise of the microtremor and feebly on the peak value of the random correlation-free disturbances. The decorrelated jitter, on the contrary, depended strongly on the random disturbances and feebly on the peak value of the driving noise of the microtremor. The reason was that an increase of the size of the microtremor meant bigger cyclic disturbances that built up over several glottal cycles and collectively resulted in bigger raw length perturbations. These were, however, removed in the decorrelated perturbations and therefore did influence decorrelated jitter only feebly.

One explanation for the weak dependence of the modulation level on the correlation-free stochastic disturbances is that the smoothing that was performed before computing the modulation level did not completely remove the influence of the random noise, because the smoothing window was short.

I. Normality

Several authors have reported that the histograms of the cycle lengths perturbations are Gaussian (Laver, 1980; Heiberger and Horii, 1982; Pinto and Titze, 1990). Pinto and Titze (1990) have expressed some reservations. The doubts that have been raised could be related to a correlation that had not been removed before calculating histograms. The histogram shape might indeed have been warped when the correlation was long term and the record length short (Woods *et al.*, 1986).

Statistical tests have confirmed that the decorrelated perturbations simulated by means of model V were normally distributed. This is a direct consequence of the central limit theorem, which warrants that the sum of many small, independent, identically distributed disturbances with zero means are normally distributed. Neither the assumption that the means are zero nor the assumption that the disturbances are identically distributed are essential (Breiman, 1986). Also, central limit theorems exist that apply even in situations when the random variables are not independent (Hamilton, 1994).

In other words, the position that jitter is a collective temporal effect implies that observed jitter could agree with many hypotheses regarding the disturbances of the instantaneous frequencies of vibration of the glottal walls. Similarly, the position that the disturbances of the instantaneous frequencies are the collective effect of many spatially coexisting weak perturbances implies that the statistical properties of the instantaneous frequencies would agree with many hypotheses regarding the properties of the individual spatial disturbances. Because of the central limit theorems, it is therefore unlikely that data on vocal jitter would enable inferring the physiological properties of its temporally and spatially distributed perturbative constituents. This would suggest that models of acoustic jitter should be parsimonious to avoid including hypotheses that are not relevant or counterfactual.

J. Parsimony

One may wonder to what extent model V is a parsimonious representation of jitter. Criteria of parsimony are the number of variables and the model structure. Simplicity in structure relates to using standard distributions, having independence in the random variables, and having linearity in variables and parameters (Gilchrist, 1984).

Model V was quasi-linear because of the smallness of the perturbations involved. The number of stochastic variables was between one and four (that is, the phases of the margins of the glottal walls) and the number of parameters was eight: average cycle length, coupling constants, speed of the surface wave, peak value of the stochastic disturbances of the instantaneous frequencies of vibration, peak value of the driving noise of the microtremor model, microtremor frequency, and microtremor bandwidth. One possible redundant parameter was the speed of the surface wave, which appeared to have a weak influence only on the statistical properties of the cycle perturbations.

A statistical description of measured glottal cycle lengths involved the following: the coefficient of variation (that is, decorrelated jitter in percent), the modulation level, the average cycle length, the order of the AR model, and the

values of the AR coefficients. The order of the AR model varied between 0 and 10, that is, the total number of statistical markers varied between 3 and 13. In most cases, AR model coefficients, although statistically significant, did not contribute equally to the representation of the spectral contour of the cycle perturbations.

The issue of parsimony did not only involve the number of model parameters versus the number of statistical markers, but also the structure of the disturbance model, which included microtremor and correlation-free stochastic disturbances. The hypotheses regarding the disturbance model agreed with the criteria of simplicity (Gilchrist, 1984). The microtremor model was linear and reproduced the spectral contour of published microtremor data. The correlation-free stochastic disturbances and the stochastic driving signal of the microtremor model was a two-point process, which is the simplest possible stochastic process and which agreed with the result of Kloeden and Platen (1999) that the exact choice of the stochastic disturbances is not relevant when the aim is to simulate the statistical properties of the variables involved in a stochastic model.

K. Laryngeal pathologies

One item that remains to be discussed is the claim that at least some laryngeal pathologies change the size of jitter. This claim has sparked the interest of many experimenters. One should not forget, however, that in the framework of earlier experiments the distinction between type I, II, and III signals has not always been made explicit. In signals that were not type I, observed cycle-to-cycle alterations were possibly the consequence of the kinematics of the vocal folds rather than increased perturbation. Similarly, correlation and trends owing to microtremor, pathological tremor, and declination have often been processed cursorily only before quantifying jitter.

Consequently, whether jitter in the narrow sense of the word increases with laryngeal pathologies remains an open question. The models that have been discussed here suggest that jitter is the response of the vibrating glottis to outside disturbances. This means that increased jitter discloses that the external disturbances have grown or that the vibrator has changed and responds differently. Within the framework of models III and V, the parameters that changed the response were the average cycle length and the coupling constants. Table V and model III show that increasing coupling constants lowered decorrelated jitter and increasing differences between the left and right coupling constants raised decorrelated jitter.

This phenomenon was distinct, however, from the one that has been occasionally mentioned in the literature in relation to left–right asymmetries. Strong asymmetries may switch the laryngeal oscillator from a type I regime to a type II regime, that is, alter the kinematics of the vocal folds (Ishizaka and Isshiki, 1976).

One may therefore provisionally conclude that at least for some diseases, the reported increase of jitter was genuine insofar as the increase referred to an amplification of the perturbations within the same kinematic regime.

VI. CONCLUSIONS

This article proposed stochastic models of jitter. The model is parsimonious. That is, it is linear, it involves a small number of parameters, and the stochastic disturbances are based on a two-point process. Properties of jitter that could be simulated were stochasticity, ergodicity and stationarity of the cycle length perturbations, the correlation between perturbations of neighboring cycles, the Gaussian distribution of the decorrelated perturbations, the lack of synchrony between the glottal cycles and the intracycle disturbances that collectively cause the observed perturbations, the frequency modulation owing to microtremor, the size of the decorrelated perturbations, the relation between size and average glottal cycle length, and the relation between size and left–right asymmetries. The model also suggests that laryngeal pathologies condition jitter indirectly via an alteration of the response of the glottal vibrations to external disturbances. Finally, the model evokes experimental methods that would enable estimating the microtremor frequency from the glottal length time series.

ACKNOWLEDGMENTS

The author is grateful to Professor Guy Mélard for suggestions concerning time series modeling. This work was supported by the National Fund for Scientific Research, Belgium.

Abarbanel, H. D. I. (**1996**). *Analysis of Observed Chaotic Data* (Springer, New York).

Abberton, E., and Fourcin, A. (**1997**). ''Electrolaryngography,'' in *Clinical Phonetics*, edited by M. J. Ball and C. Code (Whurr, London), pp. 119–148.

Atkinson, J. E. (**1976**). ''Inter- and intraspeaker variability in fundamental voice frequency,'' J. Acoust. Soc. Am. **60**, 440–445.

Bangayan, P., Long, C., Alwan, A., Kreiman, J., and Gerrat, B. (**1997**). ''Analysis by synthesis of pathological voices using the Klatt synthesiser,'' Speech Commun. **22**, 343–368.

Behrman, A. (**1999**). ''Global and local dimensions of vocal dynamics,'' J. Acoust. Soc. Am. **105**, 432–443.

Behrman, A., and Baken, R. J. (**1997**). ''Correlation dimension of electroglottographic data from healthy and pathologic subjects,'' J. Acoust. Soc. Am. **102**, 2371–2379.

Behrman, A., Agresti, C., Blumstein, E., and Lee, N. (**1998**). ''Microphone and electroglottographic data from dysphonic patients: type 1, 2 and 3 signals,'' J. Voice **12**, 249–260.

Box, G., and Jenkins, G. (**1970**). *Time Series Analysis, Forecasting and Control* (Holden-Day, San Francisco).

Box, G., and Luceno, A. (**1997**). *Statistical Control by Monitoring and Feedback Adjustment* (Wiley, New York), pp. 113–115.

Breiman, L. (**1986**). *Probability and Stochastic Processes* (Scientific, Palo Alto, CA).

Chatfield, C. (**1984**). *The Analysis of Time Series* (Chapman and Hall, London), pp. 25 and 69.

Childers, D. G. (**2000**). *Speech Processing and Synthesis Toolboxes* (Wiley, New York), pp. 390–393.

Childers, D. G., and Wu, K. (**1990**). ''Quality of speech produced by analysis-synthesis,'' Speech Commun. **9**, 97–117.

Childers, D. G., Hicks, D. M., Moore, G. P., and Alsaka, Y. A. (**1986**). ''A model for vocal fold vibratory motion, contact area, and the electroglottogram,'' J. Acoust. Soc. Am. **80**, 1309–1320.

Coleman, R. F. (**1971**). ''Effect of waveform changes upon roughness perception,'' Folia Phoniatr. **23**, 314–322.

Davis, St. B. (**1979**). ''Acoustic characteristics of normal and pathologic voices,'' in *Speech and Language*, Advances in Basic Research and Practice, Vol. 1, edited by N. J. Lass (Academic, New York), pp. 273–331.

De Guchteneere, R. (**1996**). ''Mesure et analyse chronologique des fluctuations de la durée des cycles glottiques,'' unpublished Doctoral Thesis, Université Libre de Bruxelles, Brussels, Belgium.
Dejonckere, Ph. (**1983**). ''An analysis of the diplophonia phenomenon,'' Speech Commun. **2**, 47–56.
Dejonckere, Ph. (**1985**). ''Les troubles du rythme glottique,'' Acta Oto-Rhino-Laryngol. Belg. **39(5)**, 837–843.
Gamboa, J., Jimenez, F., Nieto, A., Cobeta, I., Vegas, A., Orti-Pareja, M., Gasalla, T., Molina, J. A., and Garcia-Albea, E. (**1998**). ''Acoustic voice analysis in Patients with essential tremor,'' J. Voice **12**(4), 444–452.
Gardiner, C. W. (**1983**). Handbook of Stochastic Methods (Springer Verlag, Berlin).
Geckinly, N., and Yavuz, D. (**1981**). ''A set of optimal discrete linear smoothers,'' Signal Process. **3**, 49–62.
Gilchrist, W. (**1984**). *Statistical Modelling* (Wiley, Chichester), pp. 104–107.
Hamilton, J. D. (**1994**). *Time Series Analysis* (Princeton U. P., Princeton, NJ), pp. 180–200.
Hayes, M. H. (**1996**). *Statistical Digital Signal Processing and Modeling* (Wiley, New York), pp. 88–93.
Heiberger, V. L., and Horii, Y. (**1982**). ''Jitter and shimmer in sustained phonation,'' in *Speech and Language*, Advances in Basic Research and Practice, Vol. 7, edited by N. J. Lass (Academic, New York), pp. 299–332.
Herzel, H., Berry, D., Titze, I. R., and Saleh, M. (**1994**). ''Analysis of vocal disorders with methods from nonlinear dynamics,'' J. Speech, Lang. Hear. Res. **37**, 1008–1019.
Hillenbrand, J. (**1987**). ''A methodological study of perturbation and additive noise in synthetically generated voice signals,'' J. Speech Hear. Res. **30**, 448–461.
Hillenbrand, J. (**1988**). ''Perception of aperiodicities in synthetically generated voices,'' J. Acoust. Soc. Am. **83**, 2361–2371.
Hollien, H., Michel, J., and Doherty, E. T. (**1973**). ''A method for analysing vocal jitter in sustained phonation,'' J. Phonetics **1**, 85–91.
Horii, Y. (**1979**). ''Fundamental frequency perturbation observed in sustained phonation,'' J. Speech Hear. Res. **22**, 5–19.
Imaizumi, S. (**1986**). ''Acoustic measures of roughness in pathological voice,'' J. Phonetics **14**, 457–462.
Imaizumi, S., and Gauffin, J. (**1992**). ''Acoustic and perceptual modelling of the voice quality caused by fundamental frequency perturbation,'' in Proceedings International Conference on Spoken Language Processing, Banff, pp. 133–136.
Ishizaka, K., and Isshiki, N. (**1976**). ''Computer simulation of pathological vocal-cord vibration,'' J. Acoust. Soc. Am. **60**, 1193–1198.
Isshiki, N., Kitajima, K., Kojima, H., and Harita, Y. (**1978**). ''Turbulent noise in dysphonia,'' Folia Phoniatr. **30**, 214–224.
Kasuaya, H., Kobayashi, Y., and Kobayashi, T. (**1983**). ''Characteristics of pitch period and amplitude perturbations in pathologic voice,'' in Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing, Boston, pp. 1372–1375.
Kasuya, H., Masubichi, K., Ebihara, S., and Yoshida, H. (**1986**). ''Preliminary results on voice screening,'' J. Phonetics **3–4**, 445–456.
Kelley, W. G., and Peterson, A. C. (**2001**). *Difference Equations* (Academic, San Diego), pp. 125–150 and 164.
Kiritani, S., Imagawa, H., and Hirose, H. (**1993**). ''High-speed digital image analysis of vocal cord vibration in diplophonia,'' Speech Commun. **13**, 23–32.
Kloeden, P. E., and Platen, E. (**1999**). *Numerical Solution of Stochastic Differential Equations* (Springer, Berlin), pp. 305–338.
Koike, Y., Takahashi, H., and Calcaterra, T. C. (**1977**). ''Acoustic measures for detecting laryngeal pathology,'' Acta Oto-Laryngol. **84**, 105–117.
Kumar, A., and Mullick, S. K. (**1996**). ''Nonlinear dynamical analysis of speech,'' J. Acoust. Soc. Am. **100**, 615–629.
Ladefoged, P., and Maddieson, I. (**1996**). *The Sounds of the World's Languages* (Blackwell, Oxford), pp. 47–99.
Lalwani, A. L., and Childers, D. G. (1991). ''Modeling vocal disorders via formant synthesis,'' Proc. IEEE International Conference Acoustics, Speech and Signal Processing Vol. 1, pp. 505–508.
Laver, J. (**1980**). *The Phonetic Description of Voice Quality* (Cambridge U. P., Cambridge).
Laver, J., Hiller, S., Mackenzie, J., and Rooney, E. (**1986**). ''An acoustic screening system for the detection of laryngeal pathology,'' J. Phonetics **14**, 517–524.
Lebrun, Y., Devreux, F., Rousseau, J.-J., and Darimont, Ph. (**1982**). '' Tremolous speech,'' Folia Phoniatr. **34**, 134–142.
Lieberman, P. (**1963**). ''Some acoustic measures of the fundamental periodicity of normal and pathologic larynges,'' J. Acoust. Soc. Am. **35**, 344–353.
Markel, J. D., and Gray A. H. (**1976**). *Linear Prediction of Speech* (Springer-Verlag, Berlin).
Milenkovic, P. (**1987**). ''Least mean square measures of voice perturbation,'' J. Speech Hear. Res. **30**, 529–538.
Moon, F. C. (**1987**). *Chaotic Vibrations* (Wiley, New York).
Muta, H., Baer, T., Wagatsuma, K., Muraoka, T., and Fukuda, H. (**1988**). ''A pitch-synchronous analysis of hoarseness in running speech,'' J. Acoust. Soc. Am. **84**, 1292–1301.
Petersen, N. R. (**1978**). ''Intrinsic fundamental frequency of Danish vowels,'' J. Phonetics **6**, 177–189.
Peterson, G. E., and Schoup, J. E. (**1966**). ''The elements of an acoustic phonetic theory,'' J. Speech Hear. Res. **9(1)**, 68–99.
Pinto, N. B., and Titze, I. R. (**1990**). ''Unification of perturbation measures in speech signals,'' J. Acoust. Soc. Am. **87**, 1278–1289.
Qi, Y., and Shipp, T. (**1992**). ''An adaptive method for tracking voicing irregularities,'' J. Acoust. Soc. Am. **91**, 3471–3477.
Ross, S. M. (**1987**). *Introduction to Probability and Statistics for Engineers and Scientists* (Wiley, New York), pp. 245–290.
Rozsypal, A., and Millar, B. (**1979**). ''Perception of jitter and shimmer in synthetic vowels,'' J. Phonetics **7**, 343–355.
Schoentgen, J., and De Guchteneere, R. (**1991**). ''An algorithm for the measurement of jitter,'' Speech Commun. **10**, 533–538.
Schoentgen, J., and De Guchteneere, R. (**1995**). ''Time series analysis of jitter,'' J. Phonetics **23**, 189–201.
Schoentgen, J., and De Guchteneere, R. (**1996**). ''Searching for nonlinear relations in whitened jitter time series,'' Proceedings International Conference on Spoken Language Processing, Philadelphia, pp. 753–756.
Schoentgen, J., and De Guchteneere, R. (**1997**). ''Predictable and random components of jitter,'' Speech Commun. **21**, 255–272.
Schultz-Coulon, H.-J., Battmer, R.-D., and Fedders, B. (**1979**). ''Zur quantitativen Bewertung der Tonhöheschwankungen im Rahmen der Stimmfunktionsprüfung,'' Folia Phoniatr. **31**, 56–69.
Steiglitz, K. (**1996**). *A Digital Signal Processing Primer* (Addison–Wesley, Menlo Park, CA), pp. 86–90.
Tigges, M., Mergell, P., Herzel, H.-P., Wittenberg, T., and Eysholdt, U. (**1997**). ''Observation and modelling of glottal biphonation,'' Acust. Acta Acust. **83**, 707–714.
Titze, I. R. (**1984**). ''Parametrization of the glottal area, glottal flow, and vocal fold contact area,'' J. Acoust. Soc. Am. **75**, 570–580.
Titze, I. R. (**1989**). ''A four-parameter model of the glottis and vocal fold contact area,'' Speech Commun. **8**, 191–201.
Titze, I. R. (**1991**). ''A model for neurological sources of aperiodicity in vocal fold vibration,'' J. Speech Hear. Res. **34**, 460–472.
Titze, I. R. (**1994a**). *Principles of Voice Production* (Prentice–Hall, Englewood Cliffs, NJ).
Titze, I. R. (**1994b**). ''Summary statement: Workshop on acoustic voice analysis,'' National Center for Voice and Speech, Denver.
Titze, I. R. (**2000**). ''On the simulation of neurologic jitter and tremor with a biomechanical model of laryngeal muscle activation,'' in Proceedings 4th International Workshop on Advances in Quantitative Laryngoscopy, Voice, Speech Research, Jena, pp. 132–138.
Titze, I. R., Horii, Y., and Scherer, R. C. (**1987**). ''Some technical considerations in voice perturbation measurements,'' J. Speech Hear. Res. **30**, 252–260.
Ventsel, H. (**1973**). *Théorie des probabilités* (Editions MIR, Moscow).
Wendahl, R. W. (**1966**). ''Laryngeal analog synthesis of jitter and shimmer auditory parameters of harshness,'' Folia Phoniatr. **18**, 98–108.
Winholtz, W. S., and Ramig, L. O. (**1992**). ''Vocal tremor analysis with the vocal demodulator,'' J. Speech Hear. Res. **35**, 562–573.
Woods, A., Fletcher, P., and Hughes A. (**1986**). *Statistics in Language Studies* (Cambridge U.P., London), pp. 48–58.

The relationship of vocal tract shape to three voice qualities

Brad H. Story
Department of Speech and Hearing Sciences, University of Arizona, Tucson, Arizona 85721-0071

Ingo R. Titze
National Center for Voice and Speech, Wilbur James Gould Voice Research Center, Denver Center for the Performing Arts, Denver, Colorado 80204 and Department of Speech Pathology and Audiology, University of Iowa, Iowa City, Iowa 52242

Eric A. Hoffman
Division of Physiologic Imaging, Department of Radiology, University of Iowa College of Medicine, Iowa City, Iowa 52242

(Received 5 May 2000; accepted for publication 3 January 2001)

Three-dimensional vocal tract shapes and consequent area functions representing the vowels [i, æ, ɑ, u] have been obtained from one male and one female speaker using magnetic resonance imaging (MRI). The two speakers were trained vocal performers and both were adept at manipulation of vocal tract shape to alter voice quality. Each vowel was performed three times, each with one of the three voice qualities: normal, yawny, and twangy. The purpose of the study was to determine some ways in which the vocal tract shape can be manipulated to alter voice quality while retaining a desired phonetic quality. To summarize any overall tract shaping tendencies mean area functions were subsequently computed across the four vowels produced within each specific voice quality. Relative to normal speech, both the vowel area functions and mean area functions showed, in general, that the oral cavity is widened and tract length increased for the yawny productions. The twangy vowels were characterized by shortened tract length, widened lip opening, and a slightly constricted oral cavity. The resulting acoustic characteristics of these articulatory alterations consisted of the first two formants ($F1$ and $F2$) being close together for all yawny vowels and far apart for all the twangy vowels.
[DOI: 10.1121/1.1352085]

PACS numbers: 43.70.Fq [AL]

I. INTRODUCTION

Voice quality is a broad term that refers to the extralinguistic aspects of a speaker's voice with regard to identity, personality, health, and emotional state. Quoting Abercrombie (1967, p. 91), voice quality is ''those characteristics which are present more or less all the time that a person is talking.'' A broad description of voice quality would include features contributed by all the subsystems of speech production; i.e., respiratory, phonatory, and articulatory systems.

Some components of voice quality arise from individualized organic considerations of the speech organs. Vocal fold mass, vocal tract length, tracheal length, jaw and tongue size, and nasal cavity volume all fall into this category and may indicate information about age, sex, physique, and health. Other aspects of voice quality are brought forth from the way in which a speaker habitually uses the vocal organs for speaking. These may include socio-linguistic qualities acquired from the influence of a particular speaking community (i.e., regional dialect, accent, familial tendencies, etc.), emotional and psychological effects, or other purely idiosyncratic speech patterns. Laver (1980) outlined a formal system to describe this second category of voice qualities based on a concept of ''settings'' of the speech organs. These so-called ''settings'' represent habitual muscle tensions throughout the speech production system that impose a specific pattern of use during speech and consequently a specific voice quality.

With the focus of this paper limited to the contributions of the *vocal tract* to voice quality, it is noted that Laver (1980) proposed that long-term ''settings'' of the vocal tract bias the resulting formant structure toward a particular type of global timbre. He defined two categories of vocal tract settings as longitudinal and latitudinal. Longitudinal settings describe the state of the long axis of the vocal tract such as larynx height and protrusion/retraction of the lips. The latitudinal settings are ''quasi-permanent tendencies to maintain a particular constrictive (or expansive) effect'' within some region located along the length of the vocal tract (Laver, 1980, p. 35). Included here would be labial, lingual, faucal, pharyngeal and mandibular settings. In a similar vein, Estill *et al.* (1996) have proposed a system for voice quality control that includes a set of six elements describing the state of the vocal tract: (1) soft palate control; (2) anchoring (use of large muscles in the head, neck and torso to facilitate control of smaller muscles in the larynx); (3) pharyngeal width; (4) pharyngeal length; (5) tongue control; and (6) aryepiglottic control (tightening of the aryepiglottic sphincter). Selection of the ''position'' of each element along with a level of effort can produce a wide variety of voice qualities. Regardless of the system used, the necessity to produce appropriate phonetic sounds for intelligible speech means that vocal tract settings or positions cannot be regarded as being rigidly imposed on the vocal tract at every instant in time, but will exert an influence on the tract shape whenever conditions allow. Hence, a long-term ''quality'' is imposed on the speech signal.

Studies of the vocal tract are most often concerned with understanding and reproducing articulatory configurations that generate appropriate phonetic sounds (e.g., Stevens and House, 1955; Fant, 1960; Baer *et al.*, 1991; Narayanan *et al.*, 1995; Story *et al.*, 1996, 1998) or investigating gender and age differences of the vocal tract shape (e.g., Goldstein, 1980; Yang and Kasuya, 1994; Fitch and Giedd, 1999). The obvious importance of phonetic structure to transmission of the linguistic message makes this focus quite understandable. However, voice quality can also have a significant effect on the linguistic message as was exemplified in the classic study of Ladefoged and Broadbent (1957) who found that a listener's identification of a test word was greatly influenced by the voice quality of a phrase preceeding the test word. Thus voice quality altered the phonetic identification of vowels. This was confirmation of Joos (1948) proposal that the relationship between formant frequencies of a particular vowel and those present in other words spoken by the same speaker will determine phonetic quality rather than absolute formant frequencies. Laver (1980) and Traunmüller (1994) have both urged that increased attention be given to voice quality in studies of speech communication because it plays a vital role in communicating information to a listener that may be highly relevant to the message. Speaker-specific differences may account for much of the acoustic variability encountered in speech signals.

This paper is concerned with measurement and acoustic modeling of the vocal tract shape (specifically vowel area functions) with regard to variations in voice quality. Specifically, vocal tract shapes produced by two professional vocal performers under the articulatory conditions for a *yawny* and *twangy* voice quality are compared to their normal speaking vocal tract. A yawny quality was speaking as if initiating a yawn while the twangy quality is best described as that often used by Country and Western singers as well as the voice of former United States presidential candidate Ross Perot (Estill *et al.*, 1996).

Magnetic Resonance Imaging (MRI) was used to gather volumetric image sets of four vowel shapes ([i, æ, ɑ, u]) under the normal, yawny, and twangy conditions. The image sets were subsequently processed to yield three-dimensional reconstructions of the vocal tract and finally area functions. It must be emphasized that the use of vocal performers limits the scope of this study to the performers' own interpretation of these different voice qualities. Thus the results are not necessarily characteristic of normal, yawny, and twangy voice qualities in general. In addition, there is nothing inherently special about the yawny and twangy voice qualities, but they were chosen for comparison to the normal because they were hypothesized to have significantly enlarged and reduced cavity volumes, respectively, relative to the normal tract shape. Additionally, these qualities are ''global'' in the sense that they represent potential changes along the entire vocal tract rather than distinct local modifications such as Laver's (1980) latitudinal settings (e.g., palatized, pharyngealized, etc.). With a limited amount of scanning time available for this study, large and global modifications were desired.

The first aim of the paper is to report twelve area functions (four vowels, three voice qualities) obtained from each of the two subjects (one male, one female). In addition, formant frequencies determined from acoustic recordings of the subjects and those computed from the area functions are compared. A method is also described that guides the modification of the area functions so their computed formants match those extracted from recorded speech. The second aim is to investigate the mean area function across the four vowels for each quality in an attempt to understand the possible vocal tract ''settings'' employed to produce each quality.

II. METHODS

A. Image collection and analysis

The methods for image collection and analysis are identical to those presented in the authors' previous publications (Story *et al.*, 1996, 1998) and will not be repeated here. Only specific information regarding the subjects and the protocol will be given.

Volumetric imaging (using MRI) of the vocal tract was used to collect 12 vocal tract shapes from one male and one female subject. These consisted of the four vowels [i, æ, ɑ, u] produced with three distinct voice qualities, *normal*, *yawny*, and *twangy*. Electron beam computed tomography (EBCT) was also used to collect one image set of the vowel [ɑ] for each subject. For the present study, this image set was used only to estimate the dimensions of the teeth in order to make a correction to the MR images during analysis (because of the low amount of hydrogen in the teeth they are effectively imaged as airspace by MRI).

At the time of scanning, the male subject (M1) was 37 years old with no history of speech or voice disorders and is native to the state of New York (Rochester). He was 6 ft. 1 in. tall and weighed approximately 190 pounds. The female subject (W1) was 42 years old and also had no history of speech or voice disorders. She was 5 ft. 4 in. tall, weighed approximately 110 pounds, and is native to the state of Colorado. Both subjects have had extensive training in the vocal arts and both are professional performers. M1 holds the degrees of Bachelor of Fine Arts (BFA), Master of Music (MM), and Doctor of Musical Arts (DMA) in vocal music and has been active in teaching voice for many years. W1 holds a Bachelor of Arts (BA) in vocal music and has taught voice for 20 years. More importantly, both subjects demonstrated an ease of producing a wide variety of vocal qualities.

Prior to the imaging sessions, each subject participated in two practice/training sessions in which they lay supine on a comfortable cushion and practiced sustaining each vowel spoken with each voice quality; concentration on maintaining a steady vocal tract shape was emphasized. The subjects were, for the most part, allowed to self-interpret each voice quality. However, during the training session some description of each quality was given. Normal speech was simply a normal speaking quality, yawny quality was speaking as if yawning, and twangy was described with examples of Country and Western singers and also the voice of former U.S. presidential candidate Ross Perot (Estill *et al.*, 1996). The subjects were allowed to phonate at a comfortable pitch of their choice. M1 maintained the same pitch ($B_2^{\flat} = 116$ Hz)

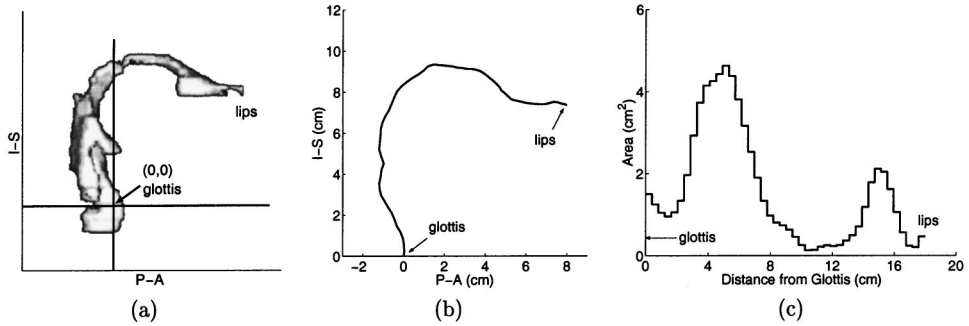


FIG. 1. Demonstration of determining the area function from a 3D reconstruction of the vocal tract for M1's vowel [u], (a) 3D reconstruction where the (0,0) point is located just above the glottis and represents the first element of the area function, (b) sagittal projection of the vocal tract centerline determined for the 3D reconstruction, and (c) the area function shown as a collection of concatenated tubelets.

for all vowel productions and within each voice quality. W1's level of comfort with each voice quality was pitch dependent. She chose to use $F_3^\sharp = 185\,\text{Hz}$ for normal, $B_2 = 175\,\text{Hz}$ for yawny, and $B_3^\flat = 233\,\text{Hz}$ for twangy. By the end of the second practice session each speaker was adept at reproducing the three voice qualities. During image collection typical modal phonation was used for all vowel productions and all productions were verified by the experimenters (two of whom are speech scientists) to be representative of both a given vowel and voice quality.

The MR images were acquired using a General Electric Signa 1.5 Tesla scanner at the University of Iowa Hospitals and Clinics. The image acquisition mode and pulse sequence parameters were identical to those used in Story *et al.* (1996, 1998). A 24–28 slice series of 5 mm thick contiguous, parallel, axial sections extended from just superior to the hard palate down to about the first tracheal ring. The field of view (slice dimensions) for each slice was $24\,\text{cm} \times 24\,\text{cm}$ which, with a pixel matrix of 256×256, gives a pixel dimension of 0.938 mm/pixel. Acquisition of full volume of slices for a given vowel required approximately 5 min of actual scanning time (about 30 repetitions of the vowel). However, with pauses for respiration the total acquisition time was on the order of 10–15 min.

B. Image analysis

The image analysis proceeded with an airway segmentation technique followed by shape-based interpolation to generate a 3D reconstruction of each vocal tract shape (Story *et al.*, 1996, 1998). As an example, the 3D reconstruction of M1's normal [u] is shown as a sagittal projection in Fig. 1(a). The point indicated as (0,0) is located just above the glottis and represents the inlet to the vocal tract. Cross-sectional areas between this point and the lip termination were determined by first finding the centerline through the 3D reconstruction with an iterative bisection algorithm (see Story *et al.*, 1996, p. 542). A sagittal projection (2D) of the centerline is shown in Fig. 1(b); note that the origin of the centerline corresponds to the (0, 0) point in Fig. 1(a). Next, areas were measured from oblique sections calculated to be locally perpendicular to the centerline. The collection of these areas extending from just above the glottis to the lips comprises the area function. Each area function was subsequently resampled with a cubic spline to contain 44 area sections. The distance Δl between each cross-sectional area was dependent on the measured total length (L_t) of each vocal tract shape (i.e., $\Delta l = L_t/44$). A smoothing filter was also applied to each area function to remove small discontinuities assumed to be imaging artifacts. The filter function was implemented as

$$a_f(i) = 0.029a_o(i) + 0.471a_o(i-1) + 0.471a_o(i-2) + 0.029a_o(i-3), \quad (1)$$

where a_0 and a_f are the original and filtered areas, respectively, and i is the section number (section 1 is just above the glottis and section 44 is at the lips). Figure 1(c) shows the 44-section filtered area function, again for M1's normal [u], as a concatenated series of discrete area sections. Henceforth, however, all area functions will be shown with a smooth curve so that figures don't become unnecessarily complicated.

The piriform sinuses were not well represented in many of the image sets for the two subjects and hence no attempt was made to measure their cross-sectional areas. As is known from previous studies, the presence of the piriform sinuses can have an effect on the formant locations (Dang and Honda, 1997; Story *et al.*, 1998) and the absence of them in the present study is noted as a limitation.

While the area function is most often used for quantitatively representing a vocal tract shape, it is often difficult to put this view into the anatomical perspective of a real speaker. In the authors' previous publications a sagittal projection of each 3D vocal tract reconstruction [like the one in Fig. 1(a)] was shown along with the area function. However, since this paper is of a comparative nature (i.e., comparing different voice qualities) a simpler representation seemed more useful. Toward this end, each area function is also presented in Sec. III as a ''pseudo-midsagittal'' projection. This was done by plotting each element of an area function as a line of equivalent diameter drawn perpendicular to the vocal tract centerline. Figure 2(a) demonstrates this process for M1's normal [u] vowel, where the curved dotted line is the sagittal (x-y) projection of the centerline (determined with the previously described iterative bisection algorithm). The thick solid lines are the equivalent diameters of successive area elements from glottis to lips. The endpoints of each equivalent diameter line are then connected to form the inner and outer profiles of the tract shape [Fig. 2(b)]. For presentation purposes in Sec. III only these profiles will be shown as demonstrated in Fig. 2(c).

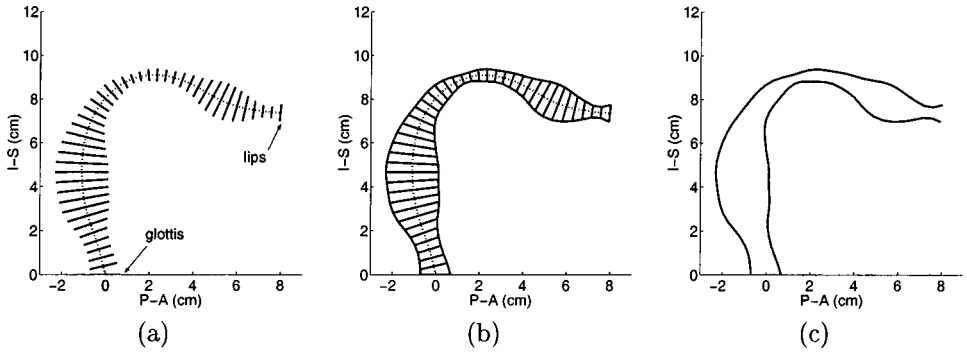


FIG. 2. Demonstration of creating a ''pseudo-midsagittal'' representation of an area function (M1's [u]), (a) vocal tract centerline (dotted line) and equivalent diameters of cross-sectional areas plotted perpendicular to the centerline (solid lines), (b) same as (a) except consecutive endpoints of each equivalent diameter have been connected with solid lines, and (c) same as (b) but with the equivalent diameter lines and the centerline removed.

C. Acoustic considerations

Audio recordings of each speaker were not made during the scanning procedure but later during a separate recording session. Some researchers have successfully recorded acoustic signals during image acquisition with custom microphones (Baer *et al.*, 1991). However, in general the MR examination room is magnetically and acoustically hostile to audio recording systems and was not attempted for this study. It is recognized that this is a limitation in comparing the acoustic characteristics of the vocal tract area functions with recordings of natural speech. But, one might also argue that since the image acquisition requires on the order of 30 or more repetitions of each vowel (Story *et al.*, 1998), there is no short segment of acoustic signal that would truly correspond to the measured vocal tract shape.

The audio recording was acquired in a fully anechoic chamber at the University of Iowa. The subjects lay supine and wore ear plugs to partially simulate the conditions in the MR scanner. An AKG-C410 head-mounted microphone was positioned 4 cm from the speaker at 45 deg off-axis of the mouth. The speakers sustained each vowel within the three voice qualities for approximately 5 s. The microphone signal was recorded onto digital audio tape (DAT) and later transferred to digital audio files (44.1 kHz sampling frequency and 16 bits resolution) via a Signalogic SIG31 data acquisition board installed in a computer workstation.

To find the first three formants of each vowel production, a 46 coefficient LPC algorithm (auto-correlation method) was first used to estimate the frequency response function. Then a peak-picking technique with parabolic interpolation (Titze *et al.*, 1987) determined the formant locations. A window size of 25 ms was used and the reported formants are the mean values over approximately 4 s of the vowel production.

For each area function, a corresponding frequency-response function was computed using a transmission-line type model (Sondhi and Schroeter, 1987). Losses due to yielding walls, viscosity, heat conduction, and radiation were included in the calculations. In all cases, for both speakers, the computed frequency response functions showed formant peaks that were at least slightly different than those determined from recorded speech; in some cases the differences were large. Since the image collection and audio recording sessions were conducted at different times and places and under different conditions, some discrepancy between computed formants and natural formants is expected. However, it is of interest to know what differences may have existed between the imaged vocal tract shape and the shape the speaker used during the audio recording. To do this, modifications were made to each area function so that the computed formant locations matched reasonably well with the formant values determined from the recorded speech.

Beautemps *et al.* (1995) have reported an optimization method for adjusting vocal tract area functions so that the error between the computed and measured formants is minimized. In the present study, a more interactive approach was taken. Area function modifications were carried out manually but were guided by the use of acoustic sensitivity functions calculated for $F1$, $F2$, and $F3$. The sensitivity of a particular formant is defined as the difference between the kinetic energy (*KE*) and potential energy (*PE*) divided by the total energy in the system (Fant and Pauli, 1975)

$$S_n(i)=\frac{KE_n(i)-PE_n(i)}{TE_n} \quad n=1,2,3 \text{ and } i=1,...,44, \tag{2}$$

where i is the section number (section 1 is just above the glottis and section 44 is at the lips), n is the formant number and $TE_n=\Sigma_{i=1}^{44}[KE_n(i)+PE_n(i)]$. The kinetic and potential energies for each formant frequency are based on the pressures $P_n(i)$ and flows $U_n(i)$ computed for each section of an area function with the transmission-line type model mentioned previously. They are calculated as

$$KE_n(i)=\frac{1}{2}\frac{\rho l(i)}{a(i)}|U_n(i)|^2, \tag{3}$$

and

$$PE_n(i)=\frac{1}{2}\frac{a(i)l(i)}{\rho c^2}|P_n(i)|^2 \tag{4}$$

where $a(i)$ and $l(i)$ are the cross-sectional area and length of section i within an area function, respectively. ρ is the density of air and c is the speed of sound.

The sensitivity function can then be used to compute the change in a particular formant frequency (F_n) due to perturbation of the area function (Δa) with the relation,

$$\frac{\Delta F_n}{F_n}=\sum_{i=1}^{44} S_n(i)\frac{\Delta a(i)}{a(i)}. \tag{5}$$

This equation says that if the sensitivity function is positively valued and the area perturbation is also positive (area is in-

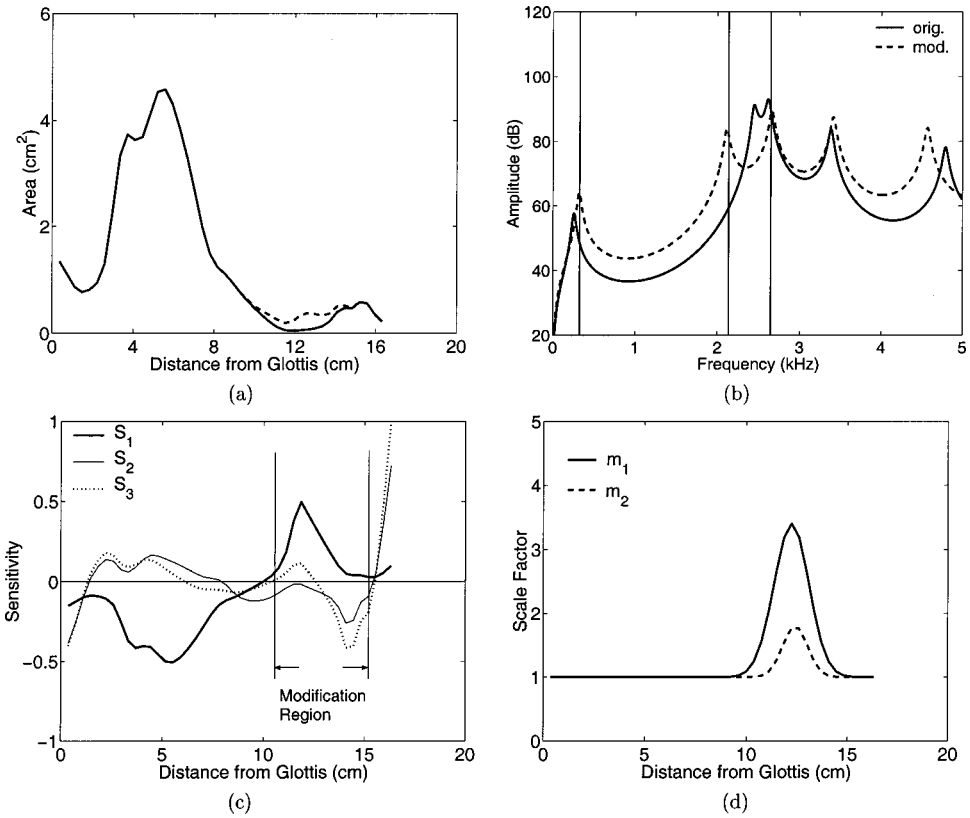


FIG. 3. Demonstration of using acoustic sensitivity functions to guide a modification of M1's [i] area function such that the first three formants coincide with those measured from audio recordings of the speaker's [i] vowel: (a) original (solid) and modified (dashed) [i] area function, (b) frequency response functions calculated for the original (solid) and modified (dashed) [i] area function (solid vertical lines represent formant frequencies determined from the speaker's audio recording), (c) sensitivity functions of $F1$, $F2$, and $F3$ for the original [i] area function, and (d) modification functions which, when multiplied by the original area function in (a), will produce the modified area function in (a).

creased) the change in formant frequency will be upward (positive). If the area change is negative (area decreased) the formant frequency will decrease. When the sensitivity function is negatively valued, the opposite effect occurs for positive or negative area perturbations. Because the sensitivity function is strictly valid for only small area perturbations it is difficult to efficiently use Eq. (5) in a quantitative fashion. Instead the the sensitivity functions are used as a qualitative guide to making manual vocal tract shape changes.

As an example of using sensitivity functions to perform an area function modification, Fig. 3(a) shows original (solid line) and modified (dashed line) versions of M1's normal [i] area function. The corresponding frequency response functions are given in Fig. 3(b) where the three vertical lines indicate the formant frequencies obtained from LPC analysis of the original audio recording. This figure shows that the area function modification has indeed moved the first three formants to locations nearly coincident with those from the recorded speech.

The necessary modification was determined by first noting that for the computed formant peaks to coincide with the formant locations of the recorded speech, $F1$ and $F3$ must both increase while a decrease is required of $F2$; however, only a modest increase in $F3$ is needed. Using the calculated sensitivity functions (for $F1$, $F2$, and $F3$) shown in Fig. 3(c) as a guide, a region extending from about 10.5 cm to 15.3 cm (distance from glottis) is observed in which S_1 remains positive and S_2 remains negative. This means that an area expansion in this specific region will increase $F1$ and decrease $F2$, as desired. Also within this length region, S_3 is both positively and negatively valued so that an increase (or decrease) in area should have little effect; i.e., the effect on this formant would be largely cancelled if the area modification spans equalivalent positive and negative portions of the sensitivity function. The goal is to expand the area in this region such that $F3$ is moved only slightly upward but $F1$ and $F2$ are moved more significantly.

The original area function, when multiplied by the two functions shown in Fig. 3(d) do move $F1$, $F2$, and $F3$ in the desired directions. Trial and error was used to create these functions with the equation,

$$m_k(i) = 1 + q_k e^{-18(i - l_k/w_k)^2}, \tag{6}$$

where q_k is an amplitude scaling factor, l_k is the location of the center of the modification (specified as distance from the glottis), w_k is the length of the region over which the modification will have an effect, and i is again the section number. (The ''18'' in the exponential is a scaling factor that allows w_k to be specified as a region length.) A modified area function can then be specified as

$$a_m(i) = a(i) \prod_{k=1}^{K} m_k(i), \tag{7}$$

where K is the required number of modification functions. In the case shown in Fig. 3, $K=2$. For the first modification function m_1 [Fig. 3(d), solid line], the parameters were set to be $q_1 = 2.4$, $l_1 = 12.2$ cm, and $w_1 = 5.2$ cm; for the second function m_2 they were set to $q_2 = 0.8$, $l_2 = 12.4$ cm, and $w_2 = 3.7$ cm.

The resulting modified area function [Fig. 3(a)] now provides a means for understanding the discrepancy between the formant locations determined from the recorded speech versus those computed for the measured area function. The

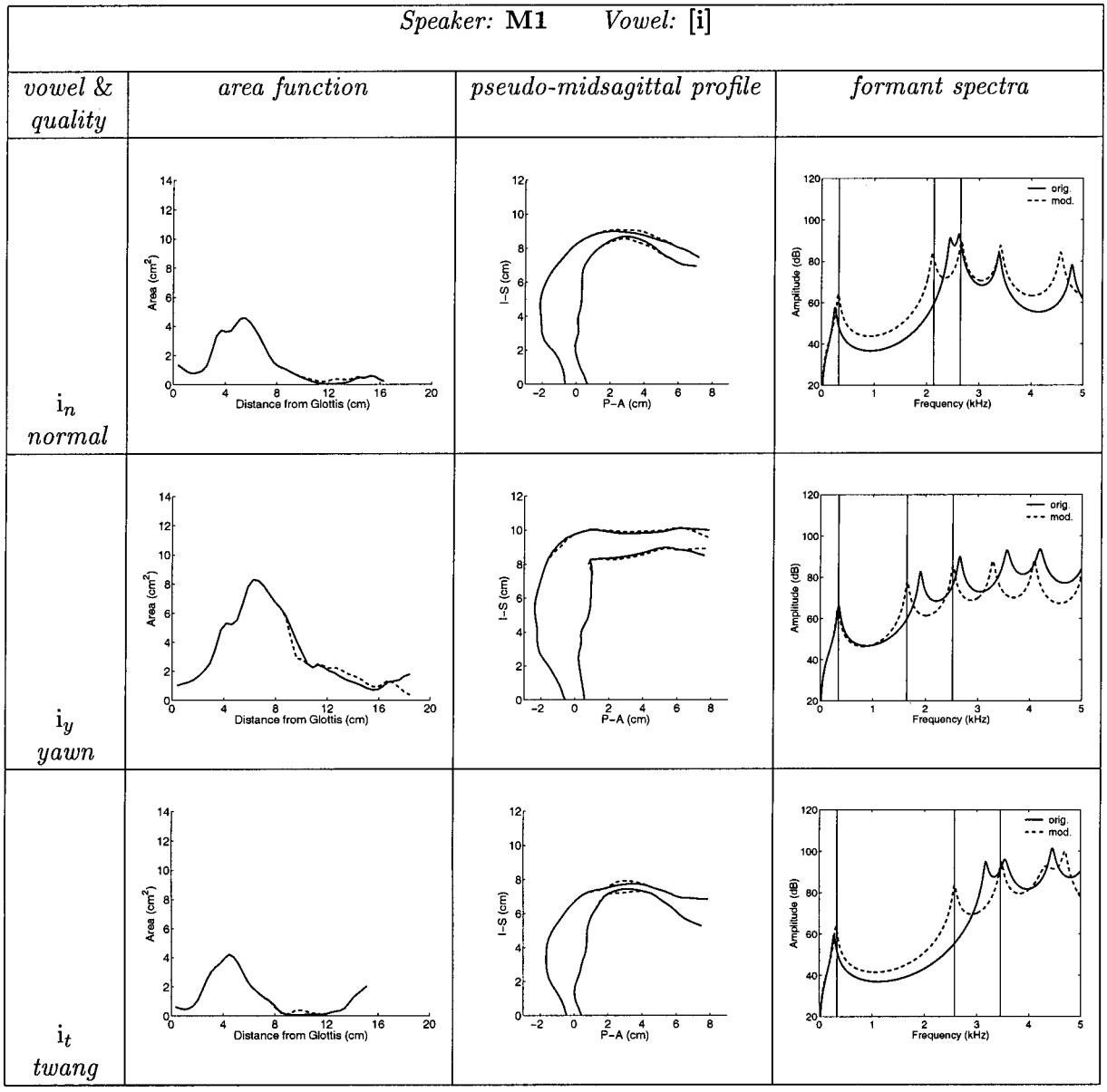


FIG. 4. Area functions, pseudo-midsagittal profiles, and formant spectra for M1's [i] vowel spoken in the *normal*, *yawny*, and *twangy* qualities. In each of the nine graphs the solid lines correspond to the original area function and the dashed lines to the modified area function (based on the method in Fig. 3). In each pseudo-midsagittal profile, the point (0,0) is located just above the glottis and the open end represents the lip termination. In addition, the horizontal axis represents the posterior-to-anterior dimension (P-A) while the vertical axis represents the inferior-to-superior dimension (I-S). The formant spectra plots include frequency response functions computed for the original (solid) and modified (dashed) area functions as well as the three vertical lines that indicate formant frequencies obtained from LPC analysis of the original audio recordings. This figure is shown in tabular form with each row representing one specific voice quality and each column giving a particular representation of the vowel (i.e., area function, pseudo-midsagittal profile, and formant spectra).

expansion that was needed in the highly constricted portion of the [i] vowel suggests two possibilities: (1) the speaker produced the vowel in the scanner differently (i.e., with a tighter constriction) than in the recording session, or (2) the area in this region was underestimated because of imaging in only the axial plane [the slice thickness may have been comparable to the size of the airspace—see Story *et al.* (1998) pp. 475–476 for a discussion of this problem]. More likely, the discrepancy probably comes about as a combination of both performance differences and imaging artifacts. As will be seen in a later section, some of the modifications seem to be obvious performance differences while others lend themselves more to an explanation of imaging artifact. The value of showing each modification is that it indicates the possible magnitude of change that may have occurred between the imaging session and the recording session.

It must be clarified, however, that it is not possible to know that the modified area functions derived by this method are actually the true vocal tract shapes used by the speaker. In fact, Mermelstein (1966) showed that three formants (the number of formants used in the present analysis) do not provide enough information to uniquely specify a vocal tract area function. But, the modifications do provide, in the least, a realistic possibility of how the tract may have been shaped during the recording of a particular sample. Note that only modifications in the form of area expansions or contractions have been used; it is also recognized that length modifications could move formant locations significantly.

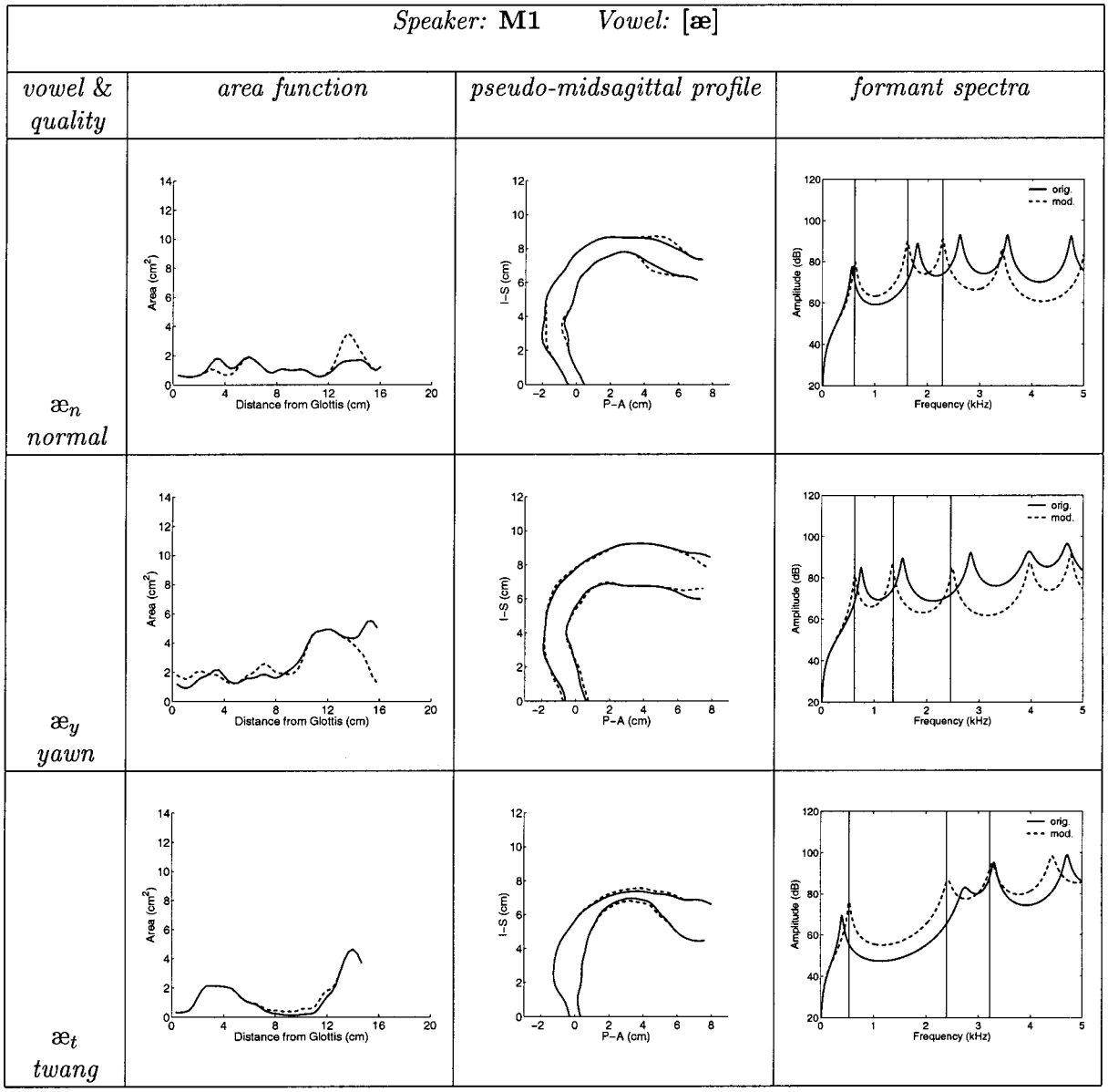


FIG. 5. Area functions, pseudo-midsagittal profiles, and formant spectra for M1's [æ] vowel spoken in the *normal*, *yawny*, and *twangy* qualities. The arrangement and description of this figure is identical to Fig. 4.

III. RESULTS AND DISCUSSION

Figures 4–7 and Figs. 9–12 show the area functions, pseudo-midsagittal profiles, and formant spectra for each of the two speakers' four vowels and three voice qualities. Each figure contains information corresponding to only one vowel and is presented in tabular form with the three qualities (normal, yawn, and twang) listed vertically and the type of representation listed horizontally. Thus figure labels of (a), (b), ... etc. have not been used. In all figures the original measurements are shown with solid lines while modifications are shown with dashed lines. In each plot containing formant spectra, the three vertical lines indicate the formant frequencies obtained from LPC analysis of the original recordings. Tables I and II present total vocal tract length and volume for each speaker, respectively. The vowels produced with a given voice quality are referred to in each figure with their appropriate IPA symbols, but augmented with the subscript of either *n*, *y*, or *t* to denote normal, yawn, or twang, respectively (e.g., the yawny *a* would be shown as a_y).

A. Male speaker (M1)

1. Vocal tract shape

The three voice quality variations of each vowel for speaker M1 are shown in Figs. 4–7. The first point to note about these figures concerns the ''accuracy'' of the area functions as assessed by the size of the modifications required to align the computed formants with those extracted from real speech (i.e., shown as three vertical lines in the formant spectra plots). In most cases, the area functions were altered only slightly, often requiring just a mild expansion of a constricted region. The source of this type of error could likely be due to an imaging artifact and subsequent underestimation of cross-sectional area. Of course the possibility does exist that, during image collection, the subject did ac-

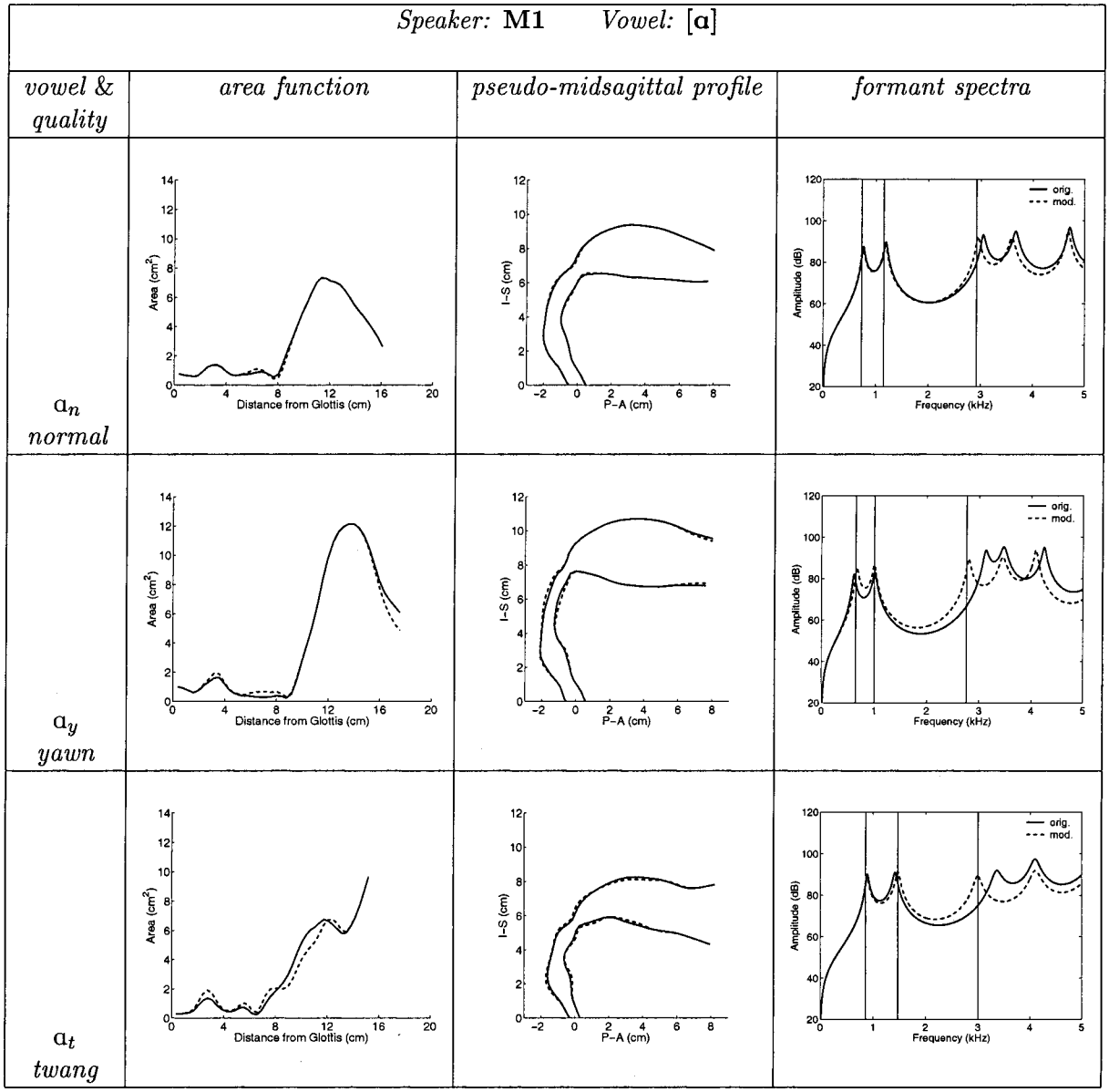


FIG. 6. Area functions, pseudo-midsagittal profiles, and formant spectra for M1's [ɑ] vowel spoken in the *normal*, *yawny*, and *twangy* qualities. The arrangement and description of this figure is identical to Fig. 4.

tually produce the vowels with a more tightly constricted portion of the vocal tract, potentially as a means of creating a highly intelligible, but extreme, version of the vowel.

In a few cases, synergistic expansions and contractions between front and back regions were necessary. The normal [æ$_n$] and [u$_n$] (Figs. 5 and 7) fall most distinctly into this category where, in each case, the front cavity needed to be expanded and the back cavity constricted. For this type of modification, the source of the error is most likely due to performance differences between the scanning session and the audio recording session; error due to image artifact is more likely to be systematic (e.g., always underestimating or always overestimating cross-sectional area). Modifications to both vowels moved $F2$ and $F3$ down in frequency and $F1$ up in frequency.

In three of the yawn quality vowels, [i$_y$], [æ$_y$], and [ɑ$_y$] (see Figs. 4–6), a reduction of the lip aperture area was needed in order to move the formant peaks to the appropriate locations. While the speaker may have actually reduced the lip area during production of these vowels, it is also possible that the speaker lengthened the tract during the audio recording. Lip area contraction achieves formant changes similar to those of tract lengthening.

Also of note is that the male speaker produced the [i$_y$] (Fig. 4) vowel with the tongue tip in contact with the hard palate, creating two lateral air flow paths at this point instead of a single conduit as is typical of vowels. Thus the production was similar to an [l] except the contact point was several centimeters posterior with the lateral airways beginning at 11.2 cm from the glottis and rejoining as a single tube at 15.3 cm from the glottis. This portion of the area function is represented as the sum of the two lateral pathways. The pseudo-midsagittal profile, which has an unnatural, right-angle shape, was based on the average of the centerlines determined for each lateral pathway. Considering the increased complexity of this vocal tract shape, the computed formants

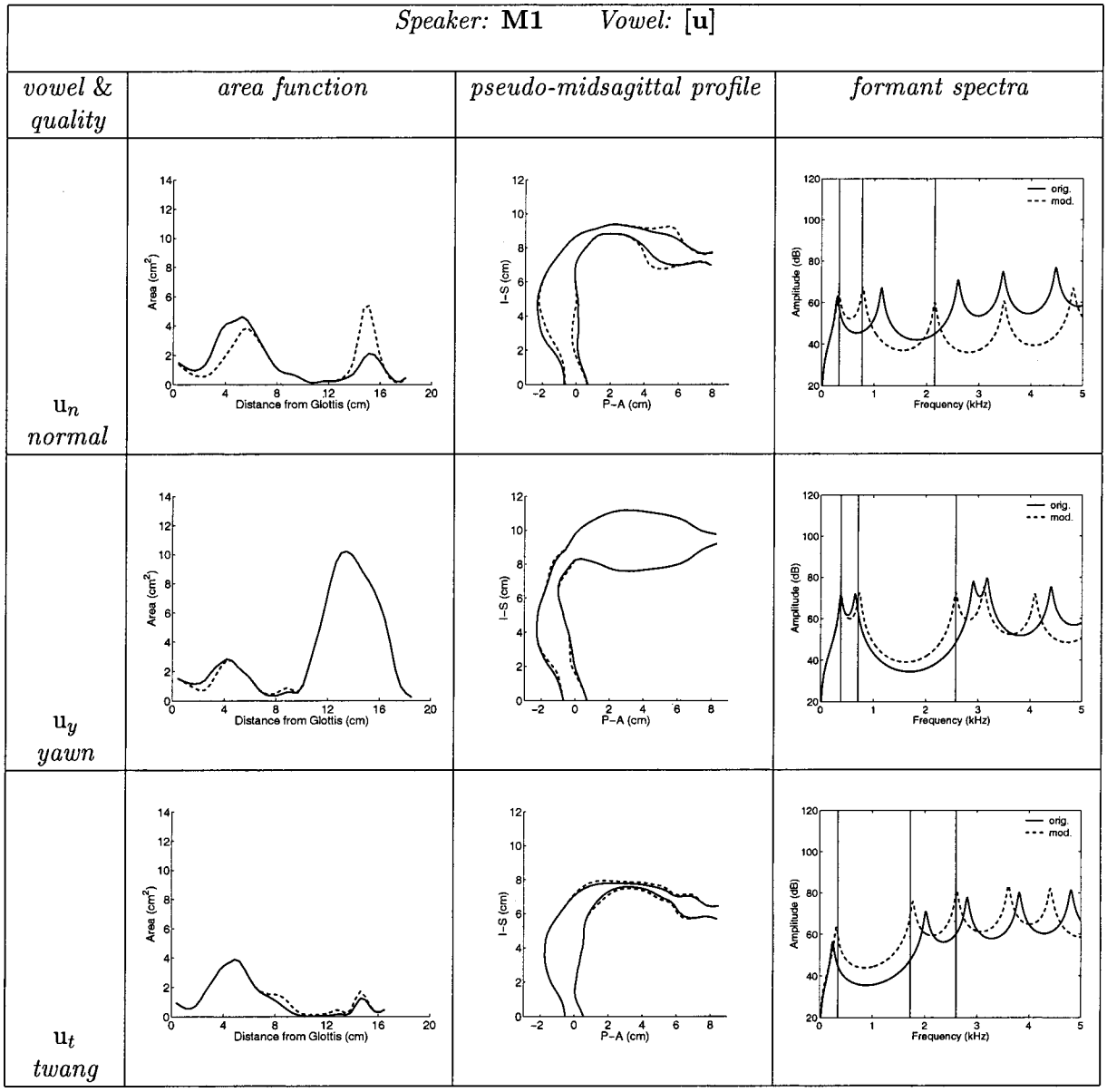


FIG. 7. Area functions, pseudo-midsagittal profiles, and formant spectra for M1's [u] vowel spoken in the *normal*, *yawny*, and *twangy* qualities. The arrangement and description of this figure is identical to Fig. 4.

were surprisingly close to those measured from recorded speech.

With regard to the voice qualities themselves, it is observed that the yawny vowels maintain significantly larger cross-sectional areas in the expanded portions of the vocal tract than do either the normal or twang qualities. For example, the pharyngeal cavity for [i_y] has nearly twice the peak area than do the [i] vowels of either of the other two qualities. Similarly, the front cavities for [$æ_y$], [$ɑ_y$], and [u_y] have areas significantly larger than those of their normal and twang counterparts. The epi-larynx tube, which is roughly defined as the first 2–2.5 cm of area function length, is also seen to possess slightly larger cross-sectional areas in the yawn quality. An obvious result of enlarged cavities is

TABLE I. Length (L_t), volume (V_o and V_m), and fundamental frequency (F_0) measurements for 12 male vowels (speaker M1). Length is given in cm, volume in cm^3, and fundamental frequency in Hz. V_o and V_m are the volumes of the original and modified area functions, respectively.

	i_n	i_y	i_t	$æ_n$	$æ_y$	$æ_t$	$ɑ_n$	$ɑ_y$	$ɑ_t$	u_n	u_y	u_t
L_t	16.32	18.39	15.14	16.06	15.84	14.70	16.10	17.56	15.22	18.00	18.52	16.50
V_o	23.7	60.0	20.0	18.5	44.3	20.2	48.0	76.3	47.9	27.9	63.7	19.4
V_m	24.5	59.3	20.4	19.9	41.0	22.0	47.9	75.6	47.0	26.9	62.9	21.7
F_o	116	116	116	116	116	116	116	116	116	116	116	116

TABLE II. Length (L_t), volume (V_o and V_m), and fundamental frequency (F_o) measurements for 12 female vowels (speaker W1). Length is given in cm, volume in cm^3, and fundamental frequency in Hz. V_o and V_m are the volumes of the original and modified area functions, respectively.

	[i$_n$]	[i$_y$]	[i$_t$]	[æ$_n$]	[æ$_y$]	[æ$_t$]	[ɑ$_n$]	[ɑ$_y$]	[ɑ$_t$]	[u$_n$]	[u$_y$]	[u$_t$]
L_t	14.34	14.52	12.94	13.82	15.14	13.64	13.95	17.07	13.38	15.31	13.46	13.64
V_o	29.9	54.9	27.1	25.6	61.7	21.7	36.0	37.4	21.5	24.4	24.2	19.0
V_m	27.7	29.4	24.2	25.3	41.7	24.2	38.8	42.9	24.5	23.9	29.4	14.8
F_o	185	175	233	185	175	233	185	175	233	185	175	233

that the yawn vowels maintain a larger vocal tract volume than either normal or twang vowels, as can be observed in Table I. In fact, for three of the yawn vowels ([i$_y$], [æ$_y$], [u$_y$]), the total volume is nearly double (or more) than that of the normal and twang vowels.

The [i$_y$], [ɑ$_y$], and [u$_y$] also show a larger total length (L_t in Table I) relative to normal and twang. This increased overall tract length could arise from either a lowered larynx or lip protrusion or possibly both. The exception is the [æ$_y$] in which the vocal tract length is slightly shorter than the [æ$_n$] but still more than a centimeter longer than [æ$_t$]. It is worth noting, however, that the modification to the [æ$_y$] area function that was necessary to bring the computed formant peaks in line with those extracted from recorded speech, largely consisted of constricting the lip section. As mentioned previously, an alternative modification that would have a similar effect on formant frequencies would be to lengthen the vocal tract.

The most prominent feature of the *twang* quality is a tendency toward a large lip opening and a slightly constricted epi-larynx, both of which will cause formants to be relatively high. For [ɑ$_t$] and [i$_t$], the lip opening is larger than both normal and yawn. Another strong feature of the twang tract shapes is that, relative to a normal voice quality, the vocal tract length is shortened. The amount of shortening is more than 1 cm for the [i$_t$], [æ$_t$], and [u$_t$] while the [ɑ$_t$] is shortened by about 0.9 cm.

2. Acoustic characteristics

The measured formants and computed formant spectra for the three qualities (fourth column in Figs. 4–7) show consistently that the locations of formants $F1$ and $F2$ in the yawn quality tend to be more closely spaced than in either the normal or twang qualities. Furthermore, for the twang quality $F1$ and $F2$ tend to be spread far apart.

Figures 8(a) and (b) show $F1$-$F2$ plots of the yawny and twangy vowels, respectively, relative to the normal vowels. For the yawny quality [Fig. 8(a)] all vowels are shifted downward (relative to the normal vowels) in the $F2$ dimension. In addition, the first formants ($F1$) of the [i$_y$] and [u$_y$] vowels are higher in frequency than in the normal quality while those of [æ$_y$] and [ɑ$_y$] are lower. Overall, the range of variation in both the $F1$ and $F2$ dimensions are reduced in the yawny quality.

For the twangy quality [Fig. 8(b)] all vowels are shifted upward in the $F2$ dimension. Interestingly, the [ɑ$_t$] which shows an increase in $F1$ of more that 100 Hz, is the only twangy vowel that shows a large change in the $F1$ dimension. The [i$_t$] and [u$_t$] show a slight increase in $F1$ while [æ$_t$] showed a small decrease.

B. Female speaker (W1)

1. Vocal tract shape

The female speaker's (W1) vocal tract characteristics for the three qualities and four vowels are depicted in Figs. 9–12. To maintain consistency the axes on these figures have been made identical to those for the male speaker. The ''accuracy'' of W1's area functions is poorer than for those of the male speaker as can be seen by the more extensive modifications required to align the computed formants with those measured from the audio recording. This is not surprising considering that this subject felt somewhat constrained by the MR scanner and often had the desire to move. The problem was not, however, one of claustrophobia, but simply that this speaker is one who tends to use more body movement

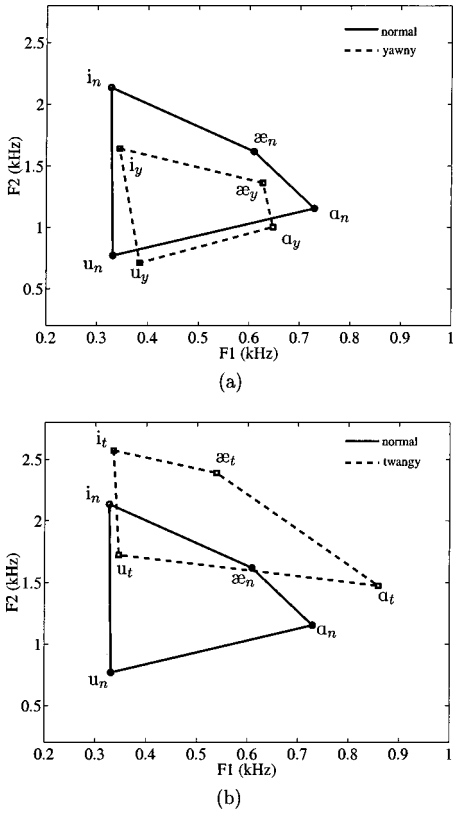


FIG. 8. Vowel space plot for the male speaker's (M1) vowels with $F2$ on the y-axis and $F1$ on the x-axis, (a) yawny (dashed) and normal (solid) voice qualities, (b) twangy (dashed) and normal (solid) voice qualities.

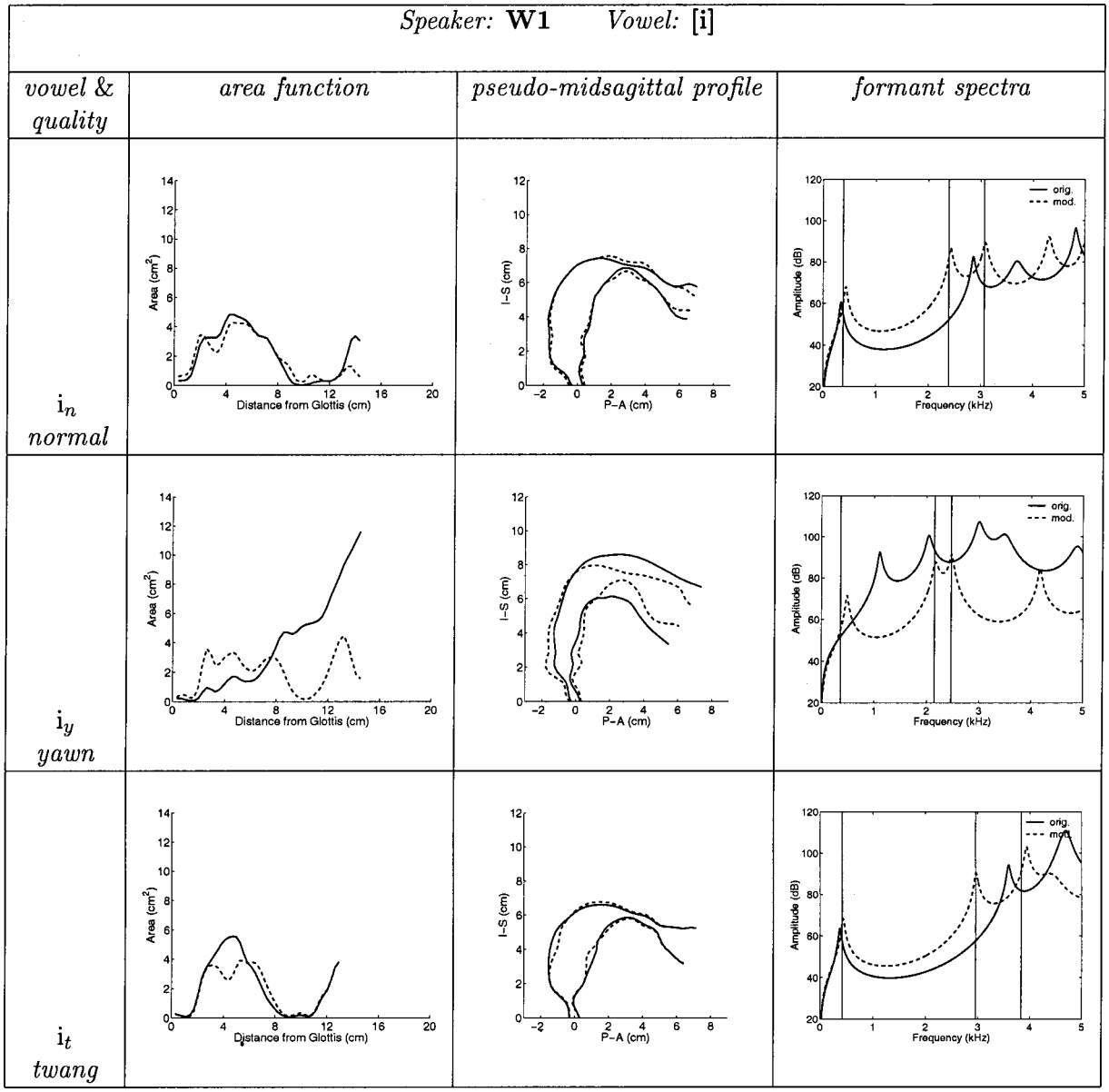


FIG. 9. Area functions, pseudo-midsagittal profiles, and formant spectra for W1's [i] vowel spoken in the *normal*, *yawny*, and *twangy* qualities. The arrangement and description of this figure is identical to Fig. 4.

during vocalization, making the required supine position along with the enclosed space of the scanner a constraining environment.

The most extensive modification had to be performed on the yawny [i_y] (Fig. 9) where the original shape seems to have little or no resemblance to an [i] vowel; in fact, it would appear to have been produced more like an [æ]. The applied modification constricted the front portion of the tract and expanded the back part, giving a more standard [i]-like shape. Even with this, $F1$ is still higher than the measured value of $F1$. Interestingly, M1's yawny [i_y] (Fig. 4) was also produced rather unconventionally (with the tongue tip in contact with the hard palate), suggesting that an [i] vowel and yawn quality are somewhat incompatible because of the simultaneous need for a front constriction (to produce [i]) and an oral expansion (to create the yawn quality). As a result the subjects may have been confused as to how to configure the vocal tract for this shape. This, coupled with possible fatigue from the experiment, perhaps forced the subjects to adopt a compromised tract shape.

The remaining three yawny vowels also required significant modifications but not as extensive as for the [i_y]. In the case of the [$æ_y$] (Fig. 10), the oral cavity volume needed to be reduced in the region 8–13 cm from the glottis but slightly expanded near the lips. A synergistic modification was required for the [$ɑ_y$] (Fig. 11) in which the cross-sectional areas in the pharynx were moderately reduced while those in the oral cavity were expanded slightly except at the lips where a large expansion was needed. A similar modification was also needed for the [u_y] (Fig. 12) except that both a contraction and expansion of areas was needed in the pharynx while a large expansion was required for the areas in the oral cavity between 10 and 12 cm from the glottis. The modifications applied to the normal and twangy vowels were also more extensive than those for the male but subtle in comparison to W1's yawny vowels.

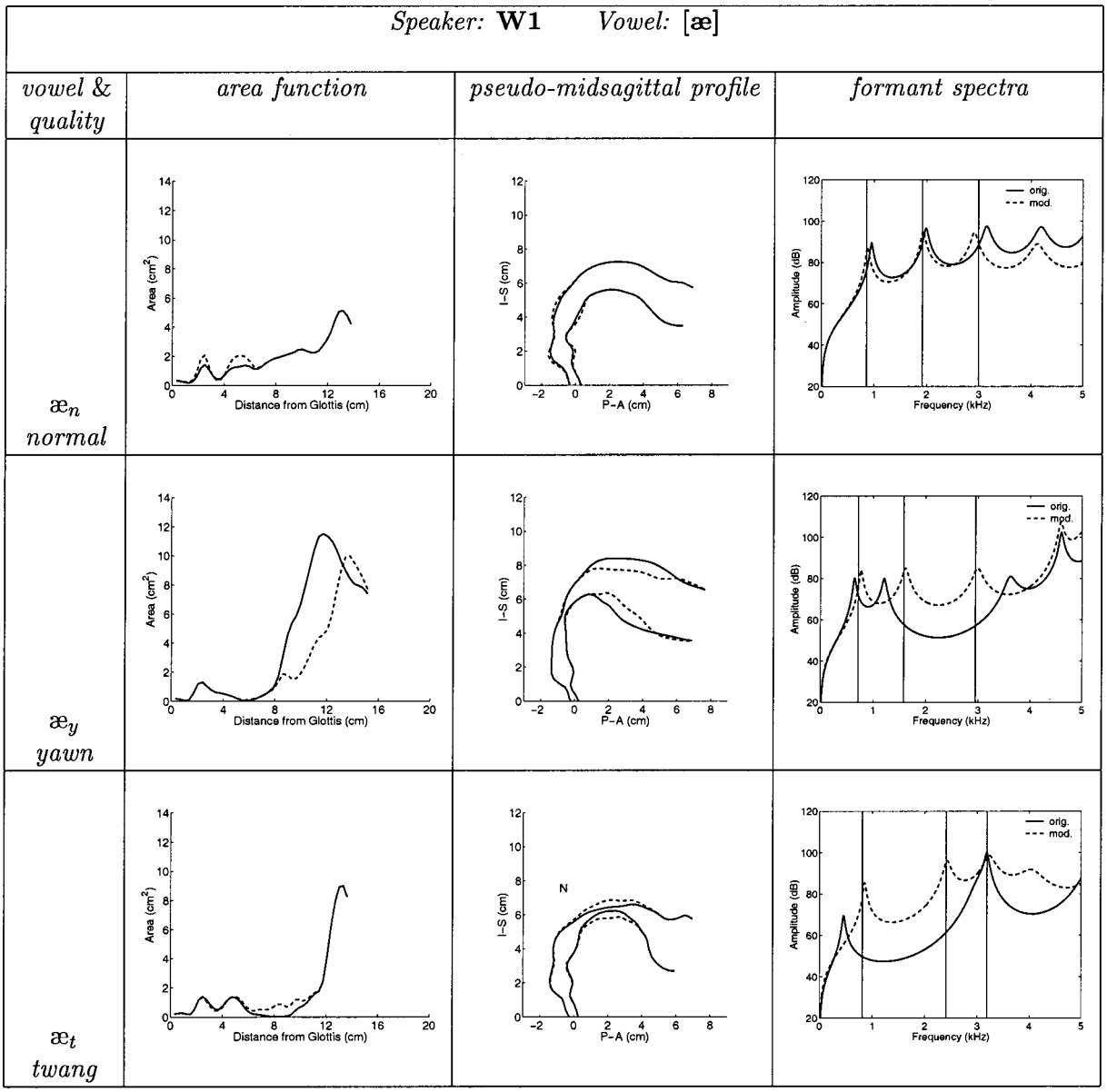


FIG. 10. Area functions, pseudo-midsagittal profiles, and formant spectra for W1's [æ] vowel spoken in the *normal*, *yawny*, and *twangy* qualities. The arrangement and description of this figure is identical to Fig. 4. The ''*N*'' shown in the pseudo-midsagittal profile for $[æ_t]$ signifies that this vowel was nasalized.

The fact that W1's yawny vowels required a great deal of modification suggests that she may have been the most uncomfortable or unfamiliar with this type of voice quality. In other words, the yawny quality may have been farther from her natural (normal) speaking voice than the twang. It is also notable that expanded vocal tract regions in the yawny vowels tend to be only slightly larger than those in the normal or twang vowels. In addition, the tract volumes for the modified area functions shown in Table II also indicate only slightly larger values for the yawn quality. This differs from M1's yawny vowels where the cross-sectional areas of expanded vocal tract regions and the tract volumes were much larger than that of the other two qualities.

With regard to vocal tract lengths, Table II shows that for the $[æ_y]$ and $[ɑ_y]$, the total tract lengths are approximately 1.5 cm and 3.0 cm longer, respectively, than their normal and twang counterparts. The total length of the $[i_y]$ is just slightly longer than the normal $[i_n]$ and about 1.5 cm longer than the twang $[i_t]$. Oddly, for the $[u_y]$ the tract length is the shortest of the three qualities. The modification of the $[u_y]$, however, still leaves the computed $F1$ and $F2$ at higher frequency locations than the measured formants, suggesting that the vowel may have actually been produced with a longer tract length during the audio recording.

Like the male speaker, W1's twang vowels have a tendency toward a large lip opening. However, unlike the male, the cross-sectional area of the epi-larynx remains similar to that of the normal quality (in which it is already quite constricted). What seems to be the most prominent feature for W1's twangy vowels is an overall constrictive effect of the entire tract, except at the lips where the effect is expansive. This is seen in the oral cavity of the $[i_t]$, the mid-tract region of the $[æ_t]$, the pharynx of the $[ɑ_t]$, and the mid-tract region

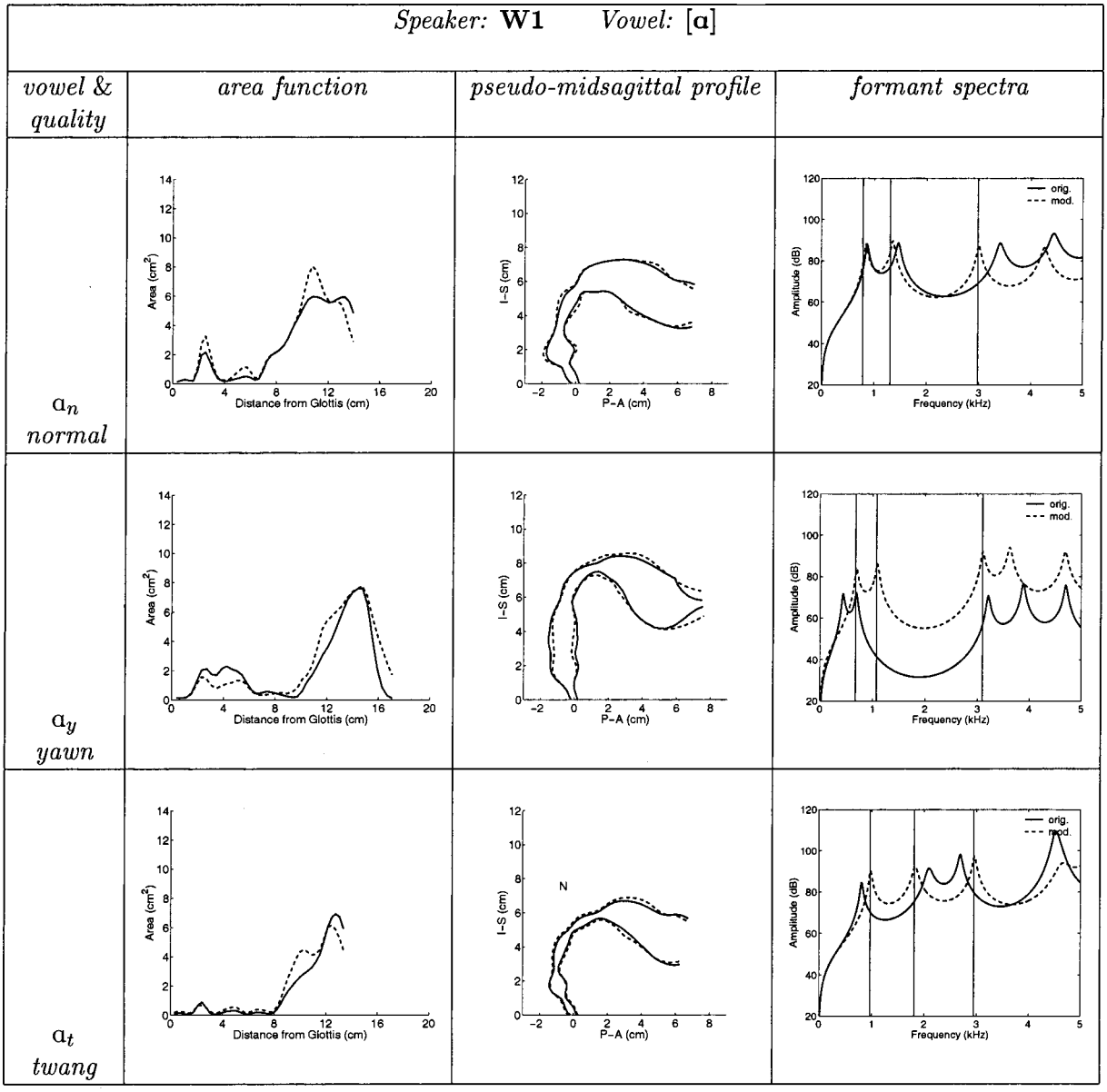


FIG. 11. Area functions, pseudo-midsagittal profiles, and formant spectra for M1's [ɑ] vowel spoken in the *normal*, *yawny*, and *twangy* qualities. The arrangement and description of this figure is identical to Fig. 4. The ''*N*'' shown in the pseudo-midsagittal profile for [$ɑ_t$] signifies that this vowel was nasalized.

of the [u_t]. This constrictive effect is also demonstrated by the tract volumes in Table II where the twangy vowels have the smallest values. It is also of interest that the constricted region in the [$æ_t$] at about 8 cm from the glottis is located at nearly the same place as the mid-tract constriction for the [u_t]. But, the front cavity shapes for each of these vowels is dramatically different with a rapid increase of area out to the lip termination in the [$æ_t$] and a contracted frontal space for the [u_t]. A feature unique to two of W1's twang vowels was that of nasalization. The [$æ_t$] and [$ɑ_t$] each showed a distinct airway branching off the main vocal tract coursing up into the nasal passages where nasal port cross-sectional areas were measured to be 0.2 cm^2 and 0.1 cm^2, respectively. Estill *et al.* (1996) notes that twang is sometimes observed to be produced with nasalization. In the case of these two vowels, the discrepancy between the formants obtained from the audio recordings and those observed in the computed frequency responses may be somewhat due to the fact the nasalization was not taken into account for the calculation. With the exception of the [u_t], all of W1's twangy vowels have tract lengths shorter than the other two qualities as can be seen in Table II.

2. Acoustic characteristics

Like the male speaker, the spectral structure of the frequency response functions and the measured formant frequencies (fourth column of Figs. 9–12) again show that the yawn quality tends toward reducing the distance between $F1$ and $F2$. The vowel space plot in Fig. 13(a) shows that, for all of the yawny vowels, $F2$ is reduced relative to the normal quality. Also, the [i_y] and [u_y] show little change in the frequency of $F1$ but $F1$ is reduced by more than 100 Hz for

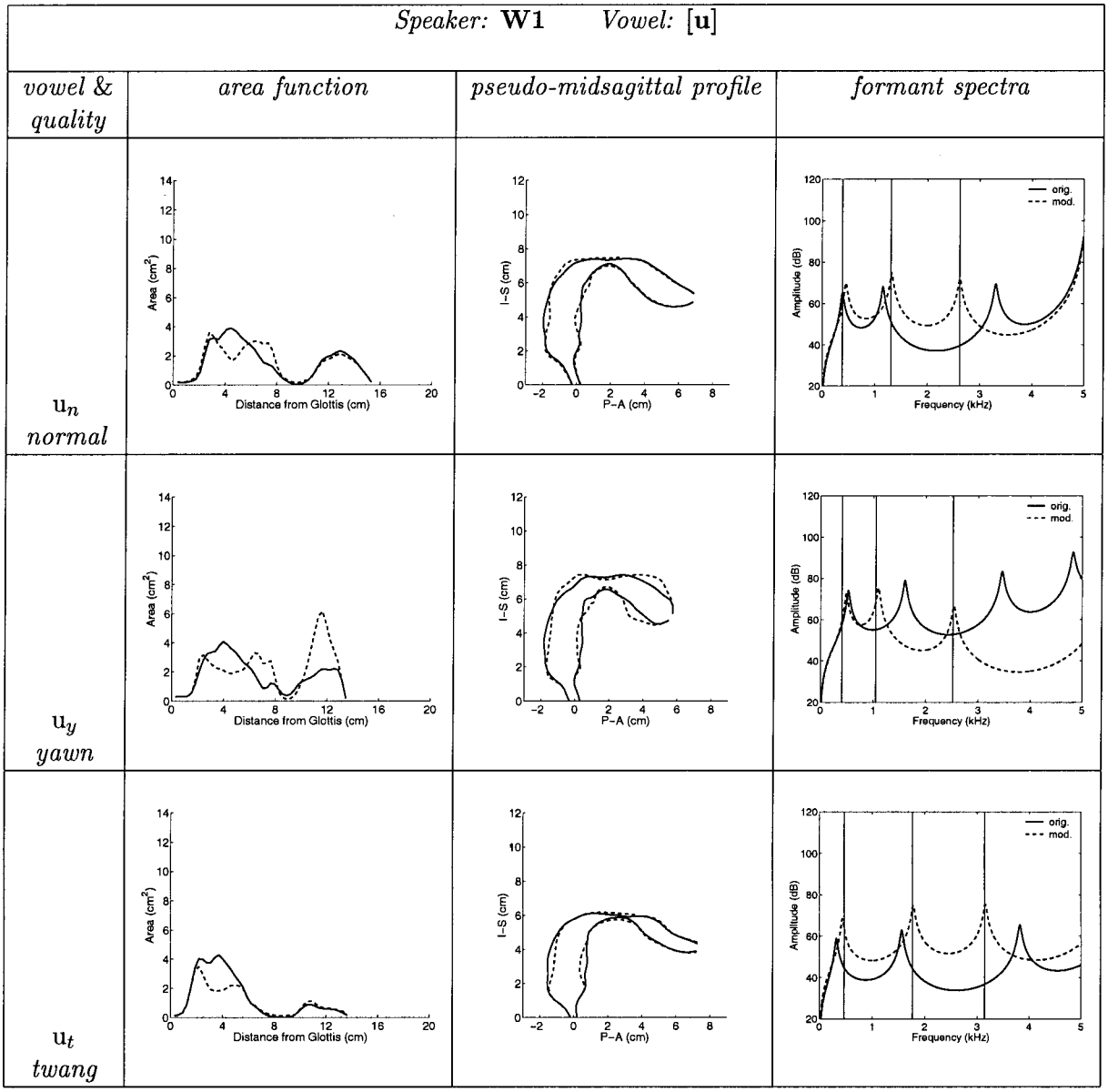


FIG. 12. Area functions, pseudo-midsagittal profiles, and formant spectra for W1's [u] vowel spoken in the *normal*, *yawny*, and *twangy* qualities. The arrangement and description of this figure is identical to Fig. 4.

both [æ$_y$] and [ɑ$_y$]. An overall lengthening of the vocal tract as well as enlarging some tract regions apparently serve to maintain these features.

The twang quality [Fig. 13(b)] tends toward increasing the distance between $F1$ and $F2$, in large part by raising $F2$. In fact, all of the twang vowels show an increase in the frequency of $F2$. Furthermore, all vowels except [ɑ$_t$] show an increase in the frequency of $F1$.

While the shifting of the formant frequencies for the twang and yawn qualities was similar for both M1 and W1, it should be noted that W1 chose to use a different voice fundamental frequency (F_o) to produce each voice quality (see Table I). In contrast, M1 maintained an F_o of 116 Hz throughout the experiment and never expressed any desire to use different F_o's. W1's use of different F_o's was motivated by achieving a level of comfort with the different qualities, however, it was not clear if this comfort level was associated with biomechanical or perceptual considerations. Perhaps W1 should have been required to maintain a consistent F_o across all productions thus eliminating the ambiguity. This, however, may have forced her into an uncomfortable and even less natural pattern of vowel production.

C. Mean area functions

In order to condense the area function data, mean area functions were computed for each speaker across the four vowels in each voice quality. They were calculated by finding the average area of each of the 44 vocal tract sections across the four vowels within a specific voice quality (this was performed on the *modified* area functions). Thus, three mean area functions were created for each speaker. The length intervals Δl were also averaged across vowels so that each mean area function has a corresponding mean length. Figure 14(a) shows the mean area functions for the male speaker. The observations made in the previous sections are

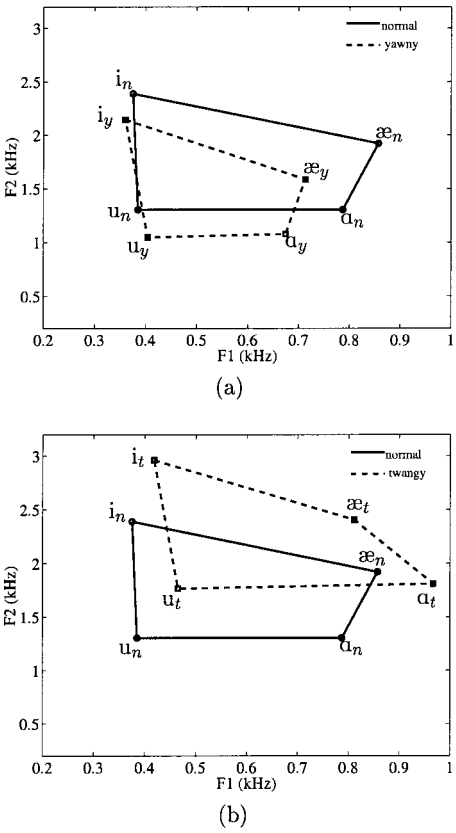


FIG. 13. Vowel space plot for the female speaker's (W1) vowels with $F2$ on the *y*-axis and $F1$ on the *x*-axis, (a) yawny (dashed) and normal (solid) voice qualities, (b) twangy (dashed) and normal (solid) voice qualities.

essentially summarized by this figure. The yawn quality has the largest cross-sectional areas and consequently the largest tract volume. It is also the longest of the three area functions and has the most expanded epi-larynx tube. Conversely, the mean area function for twang has the shortest tract length and the most constricted epilarynx. In addition, the lip opening is by far the largest of the three qualities. The normal mean area function is quite similar to the twang except that the epi-larynx is wider and the lip termination area is more than 2 cm^2 smaller.

Nearly the same characteristics can be described for the female speaker's mean area functions in Fig. 14(b). The yawn has the longest length and a large volume, at least in the oral cavity, while the twang again has a large lip termination area. Unlike the male, however, the cross-sectional areas within the epi-larynx tube are constricted (<0.5 cm^2) and change very little across the three qualities.

Figures 14(c) and (d) summarize the vocal tract characteristics of the yawn and twang qualities by showing the difference between their respective mean area functions and that of the normal quality. In a strict sense these differences are not valid because each mean area function has a different mean length and cannot be subtracted from each other centimeter by centimeter. However, the area functions can be normalized to a single length (in this case to the mean length across all 15 vowels for each speaker) and then subtracted, giving a difference function that is perhaps useful as an indicator of the global characteristics of each voice quality relative to the normal quality. For example, the difference function for M1's yawn quality [Fig. 14(c), dashed line] shows an increase in cross-sectional area along the entire vocal tract length with largest increase in the oral cavity. M1's twang difference function [Fig. 14(c), dotted line] oscillates between small positive and negative area values for about the first 8 cm from the glottis (essentially the pharynx), then shows a constrictive effect in the oral cavity out to approximately 15 cm from the glottis, and terminates the vocal tract with an increase of the lip aperture area. For W1, the yawn difference function differs from that of M1 in that a constrictive effect occurs in the lower pharynx (0–7 cm from the glottis) before showing the oral cavity expansion. The difference function for W1's twang quality shows a progressively increasing constrictive effect from 0 to 12.5 cm from the glottis, afterwhich the function becomes positive indicating an expansion at the lip termination.

Plotted in Figs. 14(e) and (f) are the frequency response functions calculated for each of the male and female mean area functions, respectively. Each spectrum is shown with a frequency range of 0–3 kHz so that the first three formants can be easily seen. There is only a small amount of variation observed in $F1$ with a 36 Hz range for the male and a 90 Hz range for the female. Relative to the normal quality, both speakers mean area function shapes reduce the distance between $F1$ and $F2$ for the yawn quality and increase the same distance for the twang; because of the small range of variation for $F1$, it is primarily $F2$ that is increased or decreased.

Both the difference functions [Figs. 14(c) and (d)] and the frequency response functions [Figs. 14(e) and (f)] confirm and summarize the earlier observations of the individual vowels with regard to spatial variations of the tract shape and formant locations. In both an acoustic sense and an articulatory sense, it might be proposed that the yawn quality is biased toward the vowel [ɑ] since the tendency is to expand the oral cavity and minimize the distance between $F1$ and $F2$. Conversely, the twang could be said to be [i]-biased because of the constriction of the oral cavity and widening of lip area as well as a tendency to maximize the distance between $F1$ and $F2$. In more traditional phonetic terms (e.g., Ladefoged, 1993) the yawny vowels have a higher degree of *backness* and lip rounding than the normal quality vowels. The twangy vowels tend toward the opposite with a more *fronted* and lip spread production. Both qualities show a less systematic variation in the *height* ($F1$) dimension.

The acoustic properties of the twang quality (e.g., increased distance between $F1$ and $F2$) share some similarity to languages that use the so-called Advanced Tongue Root or +ATR. Ladefoged and Maddieson (1996, p. 305) show that +ATR vowels in six languages (Akan, Ateso, DhoLuo, Ebira, Igbo, and Ijo) are advanced in the ($F2-F1$) or front/back dimension much as were the twang vowels in this study. However, the +ATR vowels also indicated a systematic decrease in $F1$ frequency (increase in vowel height) which was not always the case for the twang vowels. Furthermore, the physiologic realization of +ATR is apparently a considerable increase in cross-sectional area within the pharynx. On average, the twang vowels do show a slight increase in cross-sectional area within the distance of about 2 cm to 5 cm from the glottis [see Figs. 14(a) and (c)] for the

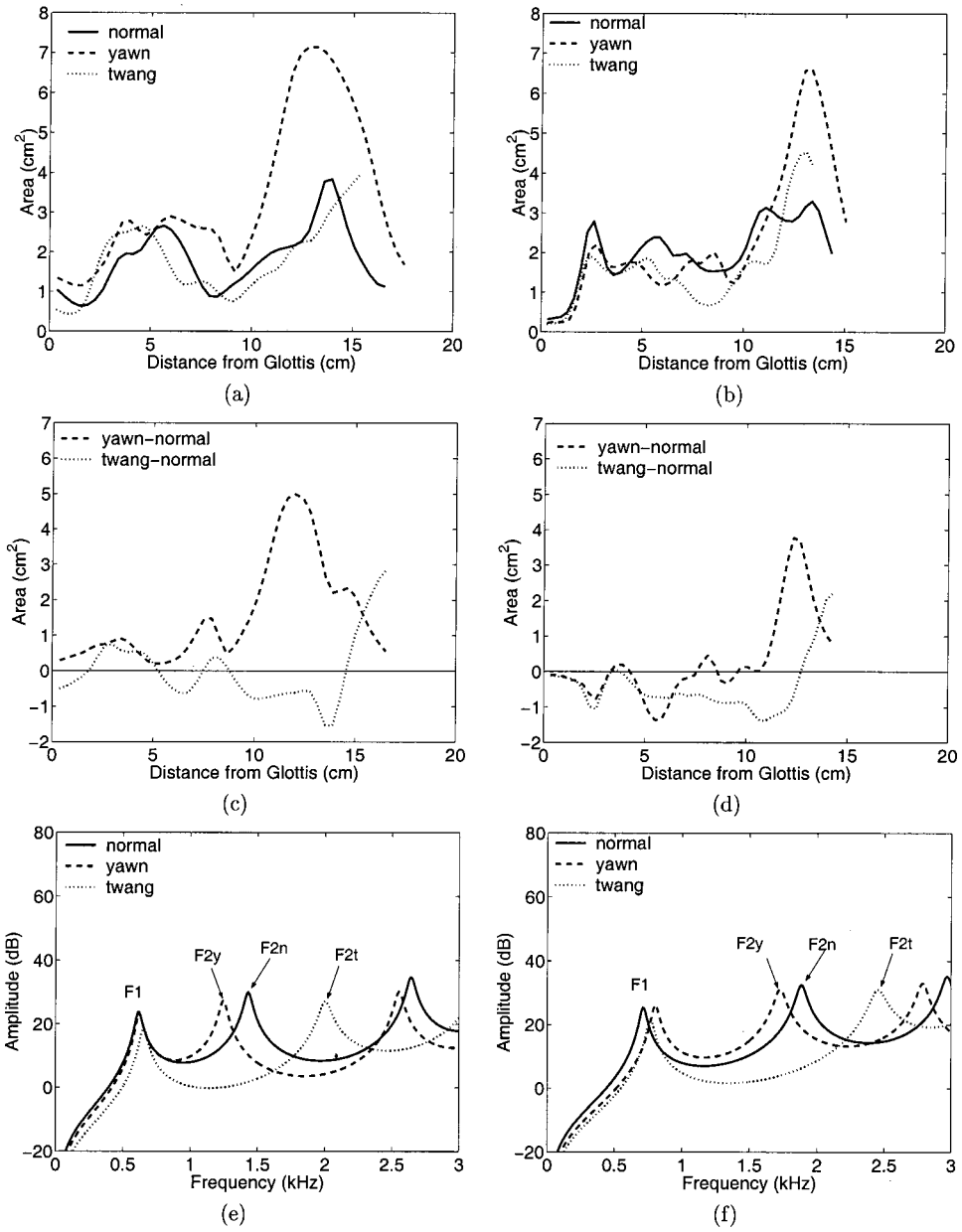


FIG. 14. Summary of the results in Figs. 4–7 and Figs. 9–12 based on mean area functions, (a) mean area functions of M1's normal, yawny, and twang qualities, (b) mean area functions of W1's normal, yawny, and twang qualities, (c) functions representing the difference between M1's yawny and normal qualities (dashed line) and between M1's twangy and the normal qualities (dotted line), (d) functions representing the difference between W1's yawny and normal qualities (dashed line) and between W1's twangy and the normal qualities (dotted line), (e) frequency response functions computed for each of M1's mean area functions (the $F2$'s for each quality are denoted with the subscripts ''*n*,'' ''*y*,'' and ''*t*''), (f) frequency response functions computed for each of W1's mean area functions (the $F2$'s for each quality are denoted with the subscripts ''*n*,'' ''*y*,'' and ''*t*'').

male but a constrictive effect in this same region for the female [see Figs. 14(b) and (d)]. Thus while acoustic properties of the twang quality are similar to +ATR vowels in various languages, the vocal tract shapes themselves do not show the same degree of similarity.

IV. CONCLUSION

This paper has described a simple study investigating some connections between voice quality and vocal tract shape for a male and female speaker. Relative to their normal speaking voice, both speakers interpreted the yawny quality to have a reduced distance between $F1$ and $F2$, acoustically, and both ''reset'' their vocal tract shape to realize this voice quality by widening the oral cavity and lengthening the entire vocal tract. The twang quality was interpreted by both speakers to have the acoustic characteristics of an increased distance between $F1$ and $F2$, the opposite of those for the yawn quality. Furthermore, the twangy vowels for both speakers were produced by resetting the vocal tract shape to have a shortened vocal tract, a widened lip aperture, and a moderate constriction of the oral cavity.

This study has shown how Laver's (1980) concept of vocal tract ''settings'' can be realized in terms of ''tendencies to maintain a particular constrictive (or expansive) effect'' within some region or regions located along the length of the vocal tract and/or to maintain a particular vocal tract length. The tract shape tendencies observed for the yawny and twangy qualities were pervasive throughout the four spoken (and imaged) vowels suggesting that they create a dominant pattern or ''background setting'' upon which speech articulation may be superimposed. Such a background setting may serve as the acoustic context in which a person produces their speech.

An obvious remaining question is how well each of the vowels and voice qualities can be identified by listeners. As stated in Sec. II, the production of each vowel within a particular voice quality was verified during data collection only

by the experimenters. A fair perceptual test of the voice qualities and the vowels would include an identification of both recorded and synthetic vowels (generated with obtained area functions) as well as an intelligibility test of connected speech produced with each quality. Such a test was beyond the scope of this paper but should constitute a future study.

ACKNOWLEDGMENTS

The authors would like to thank Steve Baker at the University of Iowa Hospitals and Clinics for his technical expertise of MRI and willingness to do late night scanning as well as the staff of the Division of Physiologic Imaging at the University of Iowa for allowing generous use of equipment and software. In addition, we would like to thank two reviewers for constructive comments on an earlier version of this paper. This study was supported by Grant No. R01 DC02532 from the National Institutes on Deafness and Other Communication Disorders.

Abercrombie, D. (**1967**). *Elements of General Phonetics* (Edinburgh University Press, Edinburgh).

Beautemps, D., Badin, P., and Laboissiere, R. (**1995**). ''Deriving vocal-tract area functions from midsagittal profiles and formant frequencies: A new model for vowels and fricative consonants based on experimental data,'' Speech Commun. **16**, 27–47.

Baer, T., Gore, J. C., Gracco, L. C., and Nye, P. W. (**1991**). ''Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels,'' J. Acoust. Soc. Am. **90**, 799–828.

Dang, J., and Honda, K. (**1997**). ''Acoustic characteristics of the piriform fossa in models and humans,'' J. Acoust. Soc. Am. **101**, 456–465.

Estill, J., Fujimura, O., Sawada, M., and Beechler, K. (**1996**). ''Temporal perturbation and voice qualities,'' in *Vocal Fold Physiology: Controlling Complexity and Chaos*, edited by P. J. Davis and N. H. Fletcher, pp. 237–252.

Fant, G. (**1960**). *The Acoustic Theory of Speech Production* (Mouton, The Hague).

Fant, G., and Pauli, S. (**1975**). ''Spatial characteristics of vocal tract resonance modes,'' in Proc. Speech Comm. Sem. Vol. 74, Stockholm, Sweden, Aug. 1–3, pp. 121–132.

Fitch, W. T., and Giedd, J. (**1999**). ''Morphology and development of the human vocal tract: A study using magnetic resonance imaging,'' J. Acoust. Soc. Am. **106**, 1511–1522.

Goldstein, U. G. (**1980**). ''An articulatory model for the vocal tracts of growing children,'' Doctoral dissertation, Department of Electrical Engineering and Computer Science, MIT.

Joos, M. (**1948**). *Acoustic Phonetics*, Supplement to Language (Journal of the Linguistic Society of America), Vol. 24, no. 2.

Ladefoged, P. (**1993**). *A Course in Phonetics*, 3rd ed. (Harcourt Brace, Fort Worth, TX).

Ladefoged, P., and Broadbent, D. E. (**1957**). J. Acoust. Soc. Am. **29**, 98–104.

Ladefoged, P., and Maddieson, I. (**1996**). *The Sounds of the World's Languages* (Blackwell, Oxford, Cambridge, MA).

Laver, J. (**1980**). *The Phonetic Description of Voice Quality* (Cambridge University Press, Cambridge, MA).

Mermelstein, P. (**1966**). ''Determination of the vocal-tract shape from measured formant frequencies,'' J. Acoust. Soc. Am. **41**, 1283–1294.

Narayanan, S. S., Alwan, A. A., and Haker, K. (**1995**). ''An articulatory study of fricative consonants using magnetic resonance imaging,'' J. Acoust. Soc. Am. **98**, 1325–1347.

Sondhi, M. M., and Schroeter, J. (**1987**). ''A hybrid time-frequency domain articulatory speech synthesizer,'' IEEE Trans. Acoust., Speech, Signal Process. **ASSP-35**, 955–967.

Stevens, K. N., and House, A. S. (**1955**). ''Development of a quantitative description of vowel articulation,'' J. Acoust. Soc. Am. **27**, 484–493.

Story, B. H., Titze, I. R., and Hoffman, E. A. (**1996**). ''Vocal tract area functions from magnetic resonance imaging,'' J. Acoust. Soc. Am. **100**, 537–554.

Story, B. H., Titze, I. R., and Hoffman, E. A. (**1998**). ''Vocal tract area functions for an adult female speaker based on volumetric imaging,'' J. Acoust. Soc. Am. **104**, 471–487.

Titze, I. R., Horii, Y., and Scherer, R. C. (**1987**). ''Some technical considerations in voice perturbation measurements,'' J. Speech Hear. Res. **30**, 252–260.

Traunmüller, H. (**1994**). ''Conventional, biological and environmental factors in speech communication: A modulation theory,'' Phonetica **51**, 170–183.

Yang, C-S., and Kasuya, H. (**1994**). ''Accurate measurement of vocal tract shapes from magnetic resonance images of child, female, and male subjects,'' Proc. ICSLP **94**, 623–626, Yokohama, Japan.

Influence of emotion and focus location on prosody in matched statements and questions

Marc D. Pell[a)]
School of Communication Sciences and Disorders, McGill University, 1266 Pine Avenue West, Montréal, Québec H3G 1A8, Canada

(Received 5 May 2000; accepted for publication 4 January 2001)

Preliminary data were collected on how emotional qualities of the voice (sad, happy, angry) influence the acoustic underpinnings of neutral sentences varying in location of intra-sentential focus (initial, final, no) and utterance ''modality'' (statement, question). Short (six syllable) and long (ten syllable) utterances exhibiting varying combinations of emotion, focus, and modality characteristics were analyzed for eight elderly speakers following administration of a controlled elicitation paradigm (story completion) and a speaker evaluation procedure. Duration and fundamental frequency ($f0$) parameters of recordings were scrutinized for ''keyword'' vowels within each token and for whole utterances. Results generally re-affirmed past accounts of how duration and $f0$ are encoded on key content words to mark linguistic focus in affectively neutral statements and questions for English. Acoustic data on three ''global'' parameters of the stimuli (speech rate, mean $f0$, $f0$ range) were also largely supportive of previous descriptions of how happy, sad, angry, and neutral utterances are differentiated in the speech signal. Important interactions between emotional and linguistic properties of the utterances emerged which were predominantly (although not exclusively) tied to the modulation of $f0$; speakers were notably constrained in conditions which required them to manipulate $f0$ parameters to express emotional and nonemotional intentions conjointly. Sentence length also had a meaningful impact on some of the measures gathered. [DOI: 10.1121/1.1352088]

PACS numbers: 43.70.Fq [AL]

I. INTRODUCTION

Recent decades have yielded numerous insights into the acoustic properties of human discourse, including its suprasegmental or ''prosodic'' attributes. For the most part, the purview of these studies has revolved around presumed linguistic or emotional features of the signals under scrutiny, limiting acoustic data on prosody to a description of one of these two operational contexts. Meager attention has been accorded to how linguistic and emotional attributes of spoken utterances *interact* to influence the acoustic form of speech. The present study addresses this issue, contributing to an initial understanding of the *conjoint* influences of emotive and linguistic processes on the acoustic form of spoken utterances in English. The current design also permitted important corroborative data on previously described prosodic phenomena in the acoustic literature (e.g., Eady and Cooper, 1986; Williams and Stevens, 1972) for a somewhat older group of speakers.

A. Acoustic investigations of emotional or linguistic prosody

Many researchers have sought to characterize external acoustic modifications linked to a speaker's emotional disposition and associated processes (e.g., Banse and Scherer, 1996; Frick, 1985; Sobin and Alpert, 1999). A sizable, independent literature has accumulated on how speakers exploit suprasegmental parameters to assign linguistic meaning in discourse (e.g., Bolinger, 1955; Cooper *et al.*, 1985; Ladd, 1996). In both areas, most studies have concentrated on the operation of fundamental frequency ($f0$) and duration in the transmission of prosodic meaning (Murray and Arnott, 1993), although important data on vocal intensity and ''voice quality'' have also been gathered (Bachorowski and Owren, 1995; Cummings and Clements, 1995; Denes, 1959; Lieberman and Michaels, 1962; Scherer, 1974; Sobin and Alpert, 1999; Turk and Sawusch, 1996).

Acoustic properties of *affective* vocalizations, like other discernable cues to emotional processes (e.g., facial, gestural, and body movements), are shaped by patterned neurophysiological responses coupled with a speaker's attempt to regulate their response due to social-contextual factors (Ekman *et al.*, 1983; Scherer, 1986; Siegman *et al.*, 1992). The speaker's physiological arousal during communication, and to a lesser extent, the relative pleasantness (*valence*) of the emotion being experienced, exert a particularly strong influence on the external configuration of emotion-related cues when ''internal'' factors are considered (Davitz, 1964; Huttar, 1968; Ladd *et al.*, 1985; Murray and Arnott, 1993; Pakosz, 1983; Scherer, 1986; Siegman and Boyle, 1993; Uldall, 1960). Given their origin in relatively widespread alterations in a speaker's emotional, physiological, and cognitive status, specifying the acoustic cues to particular types of emotional responses has proven a challenge (see Bachorowski, 1999, for a discussion). Of particular note, collecting naturalistic exemplars of vocal emotion in speech while controlling for linguistic/segmental variables and the recording quality of obtained samples has proven complicated, although some

a)Electronic mail: marc.pell@mcgill.ca

spontaneous recordings of emotion have been analyzed (Streeter *et al.*, 1983; Williams and Stevens, 1972).

Despite methodological obstacles—and the inherent compromises associated with analysis of both spontaneous and ''simulated'' emotions—research on vocal emotion expression has identified several acoustic parameters associated with purported ''basic'' emotion states (Ekman, 1992; Izard, 1977). To summarize this relationship for the emotions typically investigated, happy, angry, and fearful speech tend to display a higher mean $f0$ and amplitude, with greater $f0$ and amplitude variation, than emotionally neutral utterances; sad utterances exhibit the opposite pattern relative to neutral utterances (lower mean $f0$ and amplitude, minimal $f0$/amplitude variation) (Banse and Scherer, 1996; Breitenstein *et al.*, in press; Davitz, 1964; Fairbanks and Pronovost, 1939; Huttar, 1968; Sobin and Alpert, 1999; Williams and Stevens, 1972). Speech rate also differentiates among emotional modes, neutral and angry utterances demonstrating an accelerated rate of delivery relative to sad utterances, for example (Breitenstein *et al.*, in press; Fairbanks and Hoaglin, 1941; Johnson *et al.*, 1986; Williams and Stevens, 1972). Many of these tendencies in how ''discrete'' emotions are differentiated acoustically are captured by a hypothetical model of affect expression formulated by Scherer (1986).

Not all suprasegmental forms reflect the outward manifestation of emotional processes and associated responses. Speakers routinely employ prosody to encode *linguistic* messages based on conventionalized knowledge about their language system and its sociolinguistic applications (Bolinger, 1978). Unlike emotional attributes, acoustic correlates of linguistic prosody operate at circumscribed levels of linguistic representation such as the syllable, word, or utterance, directing the manner in which these phenomena have been studied. For example, research on linguistic prosody in English has illustrated (among others) that ''stressed'' syllables are associated with an elevated peak $f0$ and amplitude, increased $f0$ and amplitude variation, and an elongated vocalic segment when compared to unaccented syllables produced in the same linguistic environment (Brown and McGlone, 1974; Cooper and Sorensen, 1981; Klatt, 1976; Lea, 1977; McClean and Tiffany, 1973; Morton and Jassem, 1965). Words assigned linguistic focus within an utterance display a similar constellation of phonetic features as stressed syllables or words (Eefting, 1990; Ferreira, 1993; Folkins *et al.*, 1975; Weismer and Ingrisano, 1979), although more extensive modifications are witnessed in the case of focus (Bolinger, 1958; Fry, 1955, 1958).

Speech acts, or utterances varying in their linguistic-pragmatic intention with respect to the listener (e.g., statements/questions), have also been characterized acoustically; this research points to the critical importance of $f0$ parameters in the terminal portion of the utterance for many languages (Bolinger, 1978; Lieberman, 1967; Ohala, 1983). However, additional points in the $f0$ contour may also differentiate the declarative and interrogative ''modality'' (Hadding-Koch and Studdert-Kennedy, 1964; Majewski and Blasdell, 1968; O'Shaughnessy, 1979; Studdert-Kennedy and Hadding, 1973). A series of investigations undertaken by Cooper, Eady, and colleagues (Cooper *et al.*, 1985; Eady and Cooper, 1986; Eady *et al.*, 1986) furnished important data on how linguistic applications of prosody operating at the word- and utterance-levels *interact* in the speech signal for English. Utilizing a structured elicitation paradigm with a limited number of speakers, these authors analyzed prominent duration and $f0$ attributes of sentential ''key words'' according to the modality of the utterance (statement, question) and the location of contrastive focus within the stimulus (initial, medial, final key word). These reports corroborated prominent duration and $f0$ cues associated with linguistic focus and speech acts summarized above, and were instrumental in exemplifying the *interaction* of these linguistic processes on localized acoustic parameters of spoken utterances, as well as the overall shape of the intonation contour (see Eady and Cooper, 1986). The issue of utterance *length* and its potential impact on the acoustic properties of focus in statements and questions was also raised by the researchers (Eady and Cooper, 1986; Eady *et al.*, 1986), although length, focus, and modality effects on prosody implementation were not examined directly by the researchers.

B. Intersection of emotional and linguistic prosody in English

Thus increasingly sophisticated views of how speech prosody operates in emotional or nonemotional contexts are tempered by a lack of controlled experimentation looking at the *intersection* between emotional and linguistic constructs on the acoustic form of utterances (cf. Cosmides, 1983; McRoberts *et al.*, 1995; Ross *et al.*, 1986). Understanding how emotional and linguistic prosody interact in the speech signal will supply vital information on human vocal communication with its concurrent demands for affective and propositional signaling, despite access to a limited and overlapping set of acoustic features available to express these different intentions through speech. The present investigation adopts methods employed by Cooper, Eady, and colleagues in their exploration of linguistic prosody in English, adapting this procedure to evaluate acoustic dimensions of focused words in matched statements and questions in three simulated *emotional* contexts (happy, angry, sad). New data regarding the interaction of emotional and linguistic factors on suprasegmental parameters were further extended to ''short'' and ''long'' utterances elicited from a single group of encoders (Eady and Cooper, 1986; Weismer and Ingrisano, 1979). This design will allow novel insights into factors that affect the acoustic realization of linguistic meanings through prosody and how these representations are accommodated by ''prototypical'' emotional qualities of the voice during speech in a manner capable of advancing future research in this area.

II. METHODS

A. Subjects

Five female and five male normally aging adults (mean = 66.1 years, range = 59–72) volunteered for the study; these individuals also served as a control group in studies examining speech production characteristics following unilateral brain damage (Pell, 1999a, b, c). Subjects were

native speakers of Canadian English with at least eight years of formal education and no reported history of speech, language, or neurological disturbance. All participants displayed good hearing following a pure-tone air conduction screening (entry criteria: 30 dB HL at 0.5, 1, and 2 kHz, for the better ear).

B. Materials

Stimuli were two six-syllable (''*short*'') and two ten-syllable (''*long*'') English utterances matched for stress assignment at designated ''keyword'' positions (italicized):

Short								
1.	*Barry*	*took*	the	*sailboat*				
2.	*Mary*	*sold*	the	*teapot*				
Keyword:	FOC1	INT1		FOC2				
Long								
1.	*Barry*	*took*	the	*sailboat*	for	the	*weekend*	
2.	*Mary*	*sold*	the	*teapot*	for	a	*dollar*	
	FOC1	INT1		INT2			FOC2	

Long and short items were differentiated by a terminal prepositional phrase and by the number of keyword sites in the stimulus. Both short and long stimuli contained two keyword ''candidates'' for focus realization in particular contexts (FOC—underlined above). However, short stimuli contained only one rather than two ''intervening'' keywords (INT) where acoustic measures were later derived. Stimuli could be produced in three focus contexts: no focus, sentence-initial focus (focus on FOC1), or sentence-final focus (focus on FOC2). Each test item could also be intoned as a statement or yes–no question without subject-auxiliary inversion in each focus context. Finally, each item was conducive to four emotional interpretations (neutral, sad, happy, angry) which could not be inferred from the lexical-semantic content, but only from the prosody.[1] An exhaustive combination of focus, modality, and emotion features yielded 24 prosodically unique exemplars of each item or 96 productions per speaker (2 items×2 lengths×3 focus contexts×2 modalities×4 emotions).

A recorded vignette preceded each trial to elicit tokens conveying specific combinations of prosodic attributes (Cooper *et al.*, 1985; Eady and Cooper, 1986; Ryalls *et al.*, 1994). Vignettes consisted of a question or short passage that biased the target response; half of the recordings biased a response with a statement intonation and half with a question intonation. Differences in focus were rendered by modifying the vignette to place information to be used contrastively as ''new'' or unresolved within the situational context (e.g., Eady and Cooper, 1986). For example, the scenario preceding a neutral, interrogative reading of *Mary sold the teapot* with final focus (i.e., [final focus, question, neutral]) was the following:

> *You are holding a garage sale at your house with the help of some friends. After the sale, you notice that Mary sold many of the articles on her table. To find out whether the teapot was one of the items that was sold, you turn to another friend and ask*:
> [Mary sold the teapot]

To elicit emotional renditions of target utterances, vignettes provided explicit situational cues consistent with the target emotion. For example, to bias a sad interpretation of the passage above, it was explained to the listener that the teapot had strong sentimental value and had been included in the sale by accident; the vignette then terminated with a direct prompt to respond in the target emotion (e.g., *you say/ask sadly*...). Vignettes were recorded in a soundproof chamber by a male speaker who was coached to produce the passages in a neutral, ''reporting'' tone. The tape containing the 96 scenarios was edited to randomize trials for order of presentation and to insert a 5-s interstimulus pause between vignettes for subjects to execute a response.

C. Procedure

Subjects were tested individually during two 30-min sessions to limit fatigue and possible inattention to pre-recorded vignettes. Subjects were seated comfortably with a directional microphone (Sony ECM-909) 20 cm in front of their mouths and stimulus cards placed on a table in front of them. Participants were encouraged to pay close attention to the vignettes (presented free-field) and to complete the story by producing the sentence on the card in front of them (cards also contained visual information reinforcing target prosodic features, such as bold italics to signify focused words). Repetition of subjects' responses occurred rarely but was permitted in the event of reading errors, dysfluencies, or subject-initiated ''corrections''; for such items, the final production was considered for further analysis. Five practice trials acquainted subjects with the experimental procedure. All responses were recorded onto digital audio tape and re-digitized using the BLISS speech analysis system (Mertus, 1989) at a sampling rate of 20 kHz, with a 9 kHz low-pass filter setting and 12-bit quantization.

D. Speaker evaluation

To accurately portray the conjoint influences of specific emotion and focus characteristics on the acoustic form of utterances, it was imperative to limit acoustic analyses to speakers who adequately encoded these target meanings in their productions. This selection process was critical to the aims of the current investigation by mitigating the potential effects of encoders who did not accommodate well to the elicitation procedure on the acoustic findings (also Eady and Cooper, 1986). Speaker evaluation criteria took the form of listener judgments about the emotion expressed or about the word being focused for a subset of each speaker's utterances (the 12 declarative readings of ''*Mary sold the teapot*''). These stimuli elicited from each of the ten encoders were randomized within a single task and presented to ten listeners twice. On one occasion, listeners identified where the heard focus within the utterance (initial word, final word, none). On a separate occasion listeners identified the speakers' emotional tone for these tokens (neutral, sad, happy, angry). Results confirmed anticipated differences among encoders in how reliably they transmitted both types of intentions to listeners (e.g., Sobin and Alpert, 1999, for data on emotional

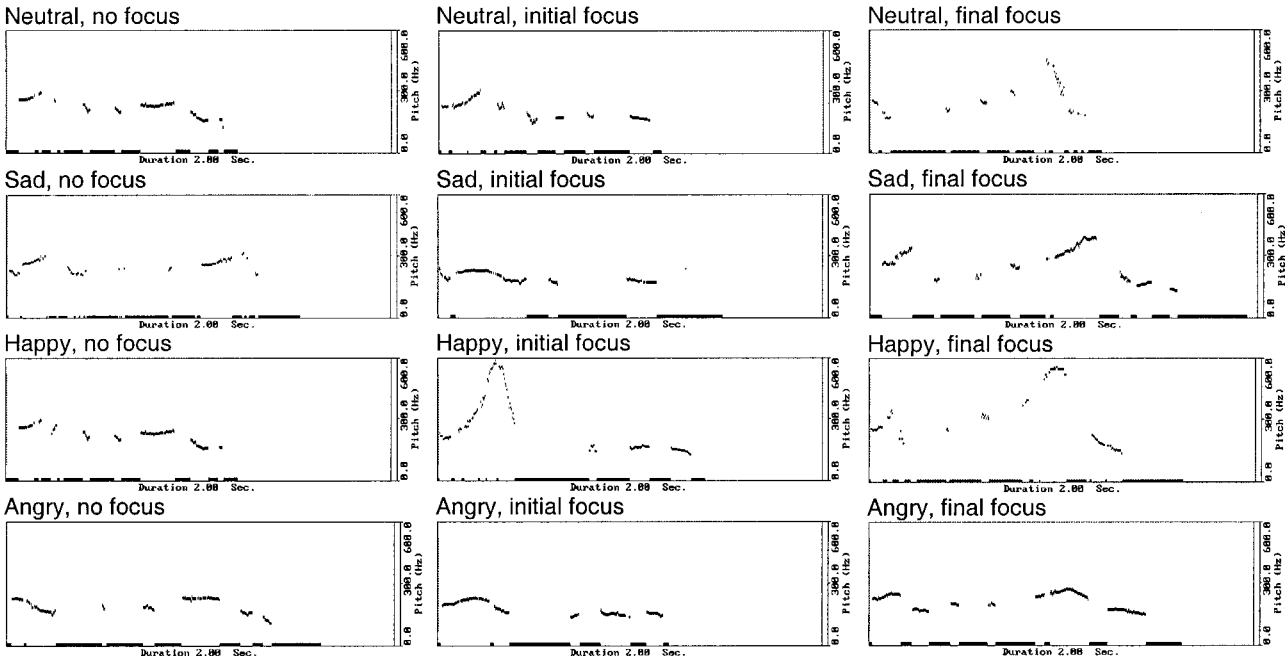


FIG. 1. $F0$ contours representing 12 distinct productions of *Barry took the sailboat* by one of the ''proficient'' female speakers.

judgments); correct listener identifications of focus and emotion targets in productions elicited from the ten encoders was 78.6% ($\pm$13.7) and 67.2% ($\pm$14.6), respectively.[2] Based on the overall group ratings, two male speakers who fell below *both* the focus and emotion condition means were excluded from further acoustic/statistical consideration. It was observed that three female encoders were especially adept at communicating linguistic and emotional meanings in tandem, yielding correct perceptual judgments of their speech exceeding 80% in both emotion and focus conditions. A sample of utterances produced by one of these ''proficient'' encoders is showcased in Fig. 1.

E. Acoustic/statistical analyses

Duration and $f0$ measures were extracted at critical ''keyword'' sites in each utterance (Eady and Cooper, 1986) and for each stimulus as a whole (e.g., Ladd *et al.*, 1985; Williams and Stevens, 1972). For *keyword measures*, duration (in ms) was computed for the full vowel (transition and steady state) on the stressed syllable of each content word (e.g., Ma-ry, sold, tea-pot, dol-lar). The corresponding $f0$ of these vowels—roughly the ''topline'' of the intonation contour (Cooper *et al.*, 1985; Eady and Cooper, 1986)—was also computed, as was the mean $f0$ of the terminal 150 ms (T) of the intonation contour (Lieberman, 1967). For both vowel and terminal measures, $f0$ was extracted via an autocorrelation algorithm and then five contiguous pulses were isolated visually at the center of the target constituent by placing cursors at zero crossings and the inverted period in Hz was averaged (Behrens, 1988; Ouellette and Baum, 1993). For whole utterance measures, three parameters were isolated for analysis: *speech rate*, or the number of syllables divided by the total sentence duration in ms; *mean* $f0$, the average of all keyword $f0$ values within a token including the terminal measure; and $f0$ *range*, the calculated difference between the maximum and minimum $f0$ value derived at keyword sites including the terminal point.

Acoustic measures were normalized for each speaker prior to statistical analysis. Keyword vowel durations were divided by the corresponding utterance duration to adjust for inter-speaker differences in speech rate. Keyword and utterance $f0$ measures were normalized for inter-speaker differences in $f0$ range by applying the following z score transform to each value: $f0i=(f0i-f0\text{mean})/s$, where $f0i$ is the observed $f0$ value, $f0$ mean is the mean $f0$ across all utterances produced by the speaker (both short and long), and s is the standard deviation (Gandour *et al.*, 1995; Rose, 1987). $F0$ range was normalized for inter-subject differences in mean $f0$ by dividing the range by the corresponding utterance mean for each speaker. Statistical analyses were performed separately on each acoustic parameter, and for keyword measures, at each keyword position [FOC1, INT1, INT2 (long only), FOC2, T]. Data gathered for the two ''short'' and the two ''long'' items were first collapsed in each condition. A total of eight $2\times3\times2\times4$ repeated measures ANOVAs were performed, with LENGTH (short, long), FOCUS (no, initial, final), MODALITY (statement, question), and EMOTION (neutral, sad, happy, angry) serving as within-subjects fixed variables. Geisser–Greenhouse $\ddot{\varepsilon}$ adjusted critical values determined the significance of tests involving more than 1 df on within-subjects factors (Max and Onghena, 1999) and a conservative p of 0.01 was adopted to focus discussion on the most robust patterns in the acoustic data. *Post hoc* comparisons were conducted using Tukey's HSD procedure ($p<0.01$), where appropriate.

III. RESULTS

Results first examine how localized ''keyword'' measures of duration and $f0$ are associated with focus realization

TABLE I. Mean duration[a] of sentence-initial (FOC1) and sentence-final (FOC2) keyword vowels produced in four emotional tones as a function of utterance length, modality and focus context. Shaded cells indicate keywords when focus was realized in each condition.

			NEUTRAL Keyword position		**SAD** Keyword position		**HAPPY** Keyword position		**ANGRY** Keyword position	
Length	Modality	Focus	FOC1	FOC2	FOC1	FOC2	FOC1	FOC2	FOC1	FOC2
SHORT	(.)	no	134	106	127	99	118	113	122	107
		initial	150	92	159	89	153	100	159	86
		final	113	118	118	113	113	124	113	121
	(?)	no	116	102	127	102	124	108	115	106
		initial	164	104	160	96	147	101	150	101
		final	120	109	116	103	115	115	117	114
LONG	(.)	no	88	78	87	83	85	89	83	86
		initial	112	69	109	74	106	79	106	78
		final	77	89	79	92	76	94	80	91
	(?)	no	83	82	86	82	82	88	85	88
		initial	109	80	112	79	105	86	113	79
		final	78	85	76	92	77	92	76	87

[a]Vowel durations (in ms) were adjusted for inter-speaker differences in speaking rate and then multiplied by 1000.

in declarative utterances, in matched interrogative utterances, in short versus long stimuli, and in stimuli expressed in different emotional modes. Data reflecting ''global'' dimensions of the same utterances are subsequently examined to elaborate on how emotion influences acoustic properties of spoken utterances that encode specific combinations of the prosodic features.

A. Keyword measures

Table I shows the duration of sentence-initial (FOC1) and sentence-final (FOC2) keyword vowels and Table II displays the corresponding $f0$ of these vowels, of ''intervening'' content words (INT1, INT2), and of the utterance terminal (T). All values are an average of two items elicited in each condition, collapsed (for expository purposes) across eight speakers (Table I) or across female or male speakers (Table II).

1. Focus realization in neutral statements: FOCUS main effect

Duration and $f0$ attributes of contrastive focus in English have been established by several investigators (e.g., Cooper *et al.*, 1985; Eady and Cooper, 1986; Weismer and Ingrisano, 1979) and are largely corroborated here. Focus realization corresponded to a lengthening of focused versus unfocused target vowels at both FOC1 and FOC2 [FOCUS: $F_{\mathrm{FOC1}}(2,14)=53.94$, $p<0.001$; $F_{\mathrm{FOC2}}(2,14)=28.22$, $p<0.001$]. Focused vowels simultaneously exhibited higher $f0$ peaks than corresponding unfocused vowels at each of these two positions [FOCUS: $F_{\mathrm{FOC1}}(2,14)=11.86$, $p<0.01$; $F_{\mathrm{FOC2}}(2,14)=27.12$, $p<0.001$]. These well-accepted parameters likely represent central features of how ''new'' information is highlighted for the listener in many spoken languages (Bolinger, 1986; Fowler and Housum, 1987).

Inspection of the data further looked at how the context for focus realization (initial, final, no focus) influenced the acoustic form of specific keywords in statements. For duration, signaling focus on the initial word of utterances (FOC1) had a significant effect on the acoustic properties of *final* keywords (FOC2); as illustrated in Fig. 2, FOC2 vowels spoken ''post-focus'' (initial-focus condition) were systematically shorter than when FOC2 was itself focused (final focus), and importantly, than in speaking conditions where *no* contrastive focus was realized in the utterance. Post-focal reduction of the $f0$ of FOC2 was also witnessed for the stimuli, yielding significant decrements in the peak $f0$ values of FOC2 in the context of ''post'' (initial) focus [see Fig. 3(a)]. These patterns indicate maximal acoustic differentiation of FOC2 as an index of the three focus contexts for both duration and $f0$. In contrast, a reciprocal process of acoustically ''de-accentuating'' FOC1 vowels in anticipation of focus at FOC2 (i.e., pre-focus effects) beyond those cues expected in utterances without focus altogether was not a statistically reliable attribute of the stimuli for either duration or $f0$.

2. Focus realization in statements versus questions: Effects of MODALITY

Stimuli formulated as a question rather than statement led to expected changes in the $f0$ of all keyword values following sentence-initial words (FOC1), consistent with a fall/nonfall pattern for statements and questions respectively (e.g., Eady and Cooper, 1986; Studdert-Kennedy and Hadding, 1973). This diverging trend was significantly distinctive within the final 150 ms of the $f0$ contour where statements displayed a lower terminal $f0$ than questions in all conditions [MODALITY: $F_T(1,7)=88.41$, $p<0.001$]. The *conjoint* influences of linguistic modality and focus context on the $f0$ contour are illustrated in Fig. 3 and were most notable when focus was realized at FOC1; this led to an abrupt lowering (statement) or elevation (question) of $f0$ by the speaker on immediate post-focus content words [MODALITY×FOCUS: $F_{\mathrm{INT1}}(2,14)=14.34$, $p<0.01$]. This marked $f0$ contrast was maintained by speakers at all remaining keyword locations in utterances spoken in the initial-focus context [MODALITY×FOCUS: $F_{\mathrm{INT2}}(2,14)=33.54$, $p<0.0001$; $F_{\mathrm{FOC2}}(2,14)=22.63$, $p=0.001$] (Eady and Cooper, 1986). Rendering matched utterances in con-

TABLE II. Mean *f*0 of vowels produced in sentence positions of potential focus (FOC), for intervening content words (INT), and for the ''terminal'' (*T*) portion of utterances spoken in four emotional tones (A–D). Data are presented separately for female ($n=5$) and male ($n=3$) speakers. Shaded cells indicate keywords where focus was realized in each condition as a function of utterance length and modality.

			SHORT Keyword Position				**LONG** Keyword Position				
Modality	Focus	Gender	FOC1	INT1	FOC2	*T*	FOC1	INT1	INT2	FOC2	*T*
A-**NEUTRAL**											
(.)	no	female	195	172	202	158	217	194	190	194	150
		male	113	115	162	74	116	118	130	154	97
	initial	female	242	143	136	111	263	157	153	150	132
		male	179	100	91	74	210	104	100	85	78
	final	female	213	186	306	125	211	193	189	208	129
		male	117	136	233	89	126	117	135	184	94
(?)	no	female	191	236	211	349	187	233	210	169	341
		male	134	119	147	259	120	126	138	121	234
	initial	female	183	280	264	369	184	260	264	241	348
		male	147	164	171	252	149	157	173	159	238
	final	female	210	214	214	331	201	215	204	171	316
		male	132	121	181	275	126	127	128	149	262
B-**SAD**											
(.)	no	female	208	187	192	134	203	179	192	189	159
		male	137	127	140	75	129	111	118	131	89
	initial	female	203	154	157	140	220	166	162	151	150
		male	139	104	90	72	155	107	120	109	84
	final	female	223	189	223	140	217	181	183	200	130
		male	124	118	125	83	133	110	125	159	111
(?)	no	female	190	225	204	301	214	203	189	160	294
		male	130	133	152	241	123	120	131	129	226
	initial	female	185	225	230	314	190	218	207	178	264
		male	128	158	166	242	142	141	169	151	225
	final	female	187	201	191	307	205	191	192	176	273
		male	130	110	160	234	123	115	131	146	241
C-**HAPPY**											
(.)	no	female	275	208	311	143	307	211	260	298	141
		male	174	128	253	89	197	139	155	214	102
	initial	female	367	180	168	129	341	181	178	207	111
		male	229	140	129	92	240	131	144	140	96
	final	female	287	224	409	144	257	193	212	335	154
		male	154	132	245	85	151	149	160	234	98
(?)	no	female	261	257	258	430	236	235	224	218	360
		male	134	162	172	293	137	125	135	145	220
	initial	female	221	258	272	414	244	255	256	238	370
		male	153	160	200	251	181	155	188	182	241
	final	female	219	213	239	413	239	208	207	222	373
		male	124	135	178	323	132	134	132	134	286
D-**ANGRY**											
(.)	no	female	222	190	261	120	216	187	214	222	104
		male	154	158	134	86	149	142	161	166	87
	initial	female	235	152	158	113	244	184	169	176	153
		male	160	102	100	86	166	124	121	107	84
	final	female	232	185	277	115	217	187	209	242	153
		male	141	133	193	90	141	138	146	181	93
(?)	no	female	238	225	247	311	239	212	232	213	318
		male	143	144	190	280	152	120	157	150	261
	initial	female	219	238	239	330	232	235	250	221	321
		male	151	163	187	264	166	180	189	166	247
	final	female	201	193	231	376	225	197	212	212	347
		male	115	118	178	277	147	139	144	166	265

texts of no-focus or final-focus was associated with less pronounced modifications to the global shape of statement/question contours (review Fig. 3).

There was also evidence that sentences formulated as a question affected how isolated target vowels at FOC1 and FOC2 were encoded to express the context for focus, elaborating patterns reported in the preceding section for statements only. As may be seen in Figs. 2 and 3 for duration and

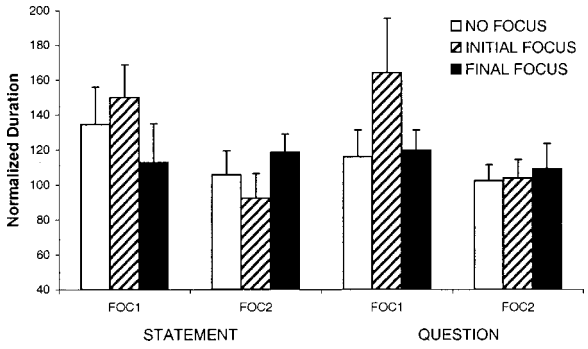


FIG. 2. Duration of sentence-initial (FOC1) and sentence-final (FOC2) vowels produced in short statements and questions in three focus contexts (averaged across eight speakers).

$f0$, respectively, stimuli posed as questions contained a more limited set of acoustic features that biased the focus context of the utterance than statements when measures were restricted to FOC1 and FOC2. In the case of duration (Fig. 2), questions were associated with focus-related vowel lengthening at both FOC1 and FOC2 as described earlier for statements; however, in questions, there was no significant evidence of post-focal vowel reduction at FOC2 in the initial-focus context beyond that associated with no focus [MODALITY×FOCUS: $F_{\mathrm{FOC2}}(2,14)=13.84$, $p=0.001$]. In the case of $f0$ (Fig. 3), formulating stimuli as a question had an even more pronounced impact on localized cues to focus at FOC1 and FOC2; unlike statements where reliable $f0$ differences characterized focused and unfocused vowels at each of these keyword positions, questions contained *no* statistically significant $f0$ differences indicative of focus context when measures were limited to FOC1 and FOC2

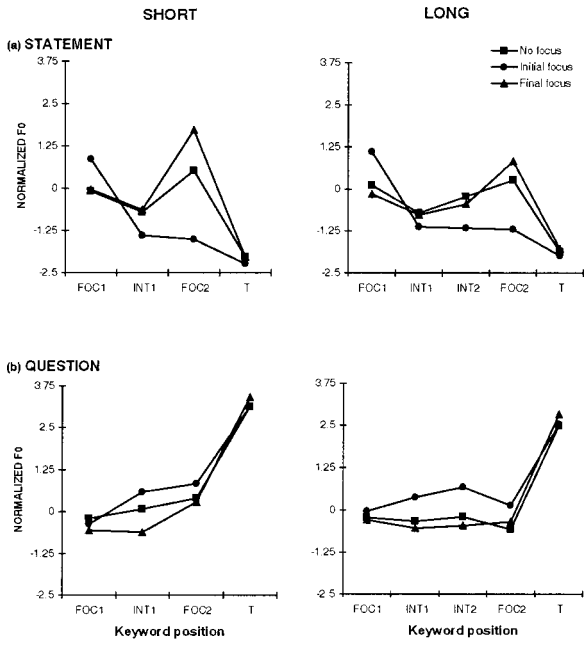


FIG. 3. Fundamental frequency of initial (FOC1), intervening (INT1, INT2), final (FOC2), and terminal (*T*) keyword vowels produced in short and long sentences as a (a) statement and (b) question (averaged across eight speakers and four emotions).

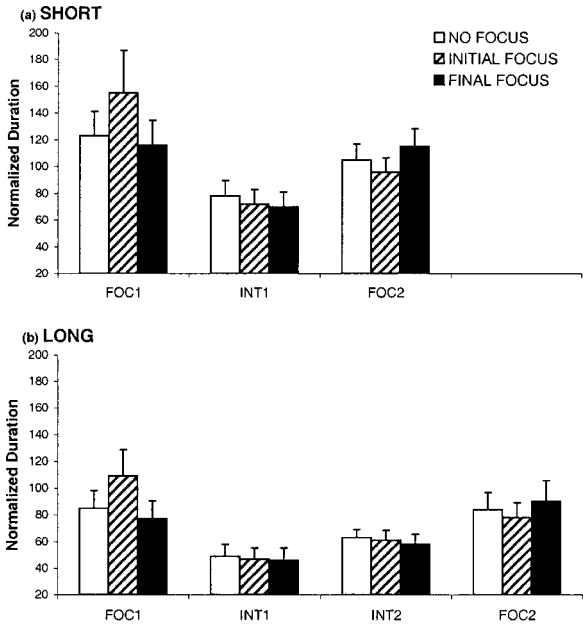


FIG. 4. Duration of initial (FOC1), intervening (INT1, INT2), and final (FOC2) keyword vowels spoken in three focus contexts for (a) short and (b) long stimuli (averaged across eight speakers and two modalities).

[MODALITY×FOCUS: $F_{\mathrm{FOC1}}(2,14)=12.57$, $p<0.01$; $F_{\mathrm{FOC2}}(2,14)=22.63$, $p=0.001$].

3. Focus realization in short versus long utterances: Effects of LENGTH

Producing ten—as opposed to six—syllable utterances did not alter the relative pattern of $f0$ cues implemented at FOC1 or FOC2 to signal different focus contexts. Only the endpoint of the $f0$ contour was influenced significantly by length, revealing differences in terminal measures for questions (''short'' terminated with a higher $f0$ than ''long'') but not for statements [LENGTH×MODALITY: $F_T(1,7)=24.59$, $p=0.002$] (review Fig. 3).

Figure 4 presents a graphic comparison of keyword duration values in short and long stimuli averaged across modality. At FOC2, it can be seen that duration parameters corresponding to the three focus contexts were not maximally distinctive in long stimuli [Fig. 4(b)] as they were in short stimuli [Fig. 4(a)], although obtained cues were in the expected direction (for long: final-focus=no-focus >initial-focus) [LENGTH×FOCUS: $F_{\mathrm{FOC2}}(2,14)=6.54$, $p=0.01$]. Of perhaps greater importance, temporal attributes of ''intervening'' words in the stimuli played a key role in marking focus interpretation for long utterances. Recall that long stimuli permitted two vowel measurements between experimentally designated focus positions (e.g., *sold* and *teapot* in *Mary sold the teapot for a dollar*). When focus was realized at FOC2 (sentence-final position), speakers systematically reduced the duration of the intervening content word directly preceding FOC2 (i.e., INT2) in this one context [FOCUS: $F_{\mathrm{INT2}}(2,14)=9.82$, $p<0.01$]. Interestingly, pre-focus anticipatory markers in long utterances were more pronounced at INT2 in questions (no-focus=initial-focus >finalfocus) than in statements (no-focus>final-focus =initial-focus) [MODALITY×FOCUS: $F_{\mathrm{INT2}}(2,14)=7.29$,

$p<0.01$]. This difference may reveal subtle mechanisms by which speakers attempt to compensate for apparent $f0$ constraints in signaling focus in interrogative utterances noted above. Focus-related alterations in the duration of INT1 (e.g., *sold*) were never observed for either short or long stimuli ($p>0.01$) (Weismer and Ingrisano, 1979).

4. Focus realization in emotional utterances: Effects of EMOTION

Documenting the impact of emotion on acoustic features reported above constituted the primary aim of the current investigation. First, as expected, varying emotional qualities of the stimuli was associated with acoustic changes that were detectable at isolated keyword locations but unrelated to other specifications of the stimuli (such as focus). These main effects were predominantly tied to the operation of $f0$, were significant at all keyword locations [EMOTION: $F_{\mathrm{FOC1}}(3,21)=34.12$, $p<0.001$; $F_{\mathrm{INT1}}(3,21)=6.85$, $p=0.01$; $F_{\mathrm{INT2}}(3,21)=9.03$, $p<0.01$; $F_{\mathrm{FOC2}}(3,21)=30.26$, $p<0.001$], and will be discussed in a later section on mean $f0$ and $f0$ variation at the *sentence* level. Of singular note here, emotion marginally influenced the terminal point of utterances overall, with happy contours demonstrating a higher endpoint than sad contours [EMOTION: $F_T(3,21)=4.92$, $p=0.017$]. Terminal $f0$ values of angry or neutral stimuli did not significantly differ in either case from those of happy or sad stimuli.

Manipulating emotional attributes of experimental stimuli had a strong influence on patterns of $f0$ implemented at FOC1 and FOC2 to mark focus context, which were both dependent on the linguistic modality of the utterance. At FOC1, significant interactions emerged for EMOTION×FOCUS [$F_{\mathrm{FOC1}}(6,42)=7.98$, $p=0.001$], EMOTION×MODALITY [$F_{\mathrm{FOC1}}(3,21)=10.30$, $p<0.01$] and MODALITY×FOCUS (reported above). Analysis of FOC2 vowels yielded a three-way interaction of EMOTION×MODALITY×FOCUS [$F_{\mathrm{FOC2}}(6,42)=6.29$, $p<0.01$] which was marginally significant at FOC1 ($p=0.08$). An overview of the interplay among emotion, focus, and modality on $f0$ characteristics of FOC1 and FOC2 vowels is illustrated in Fig. 5 for short stimuli only.

In summarizing the three-way relationship, it is first worth stating that differences in emotion did not lead to systematic alterations in how $f0$ was implemented at FOC1 or FOC2 to communicate the context for focus; rather, the interplay of focus and emotion was mostly attributable to differences among the four levels of emotion that were not entirely uniform when each of the focus contexts was examined separately. In general, the $f0$ of FOC1 and FOC2 produced in a happy tone consistently surpassed that spoken in a sad or neutral tone. The same keywords expressed in angry stimuli exhibited a higher $f0$ than those in sad stimuli (where $f0$ tended to be lowest) but this comparison was significant in fewer focus contexts. Importantly, these qualitative patterns were strongly dependent on the linguistic modality of the utterance and, when inspected carefully, were only true of statements. For questions, rather, there were no statistically reliable differences among the four emotions in any of the three focus contexts when measures were restricted to the purview of FOC1 and FOC2 [compare Figs. 5(a)–(c)].

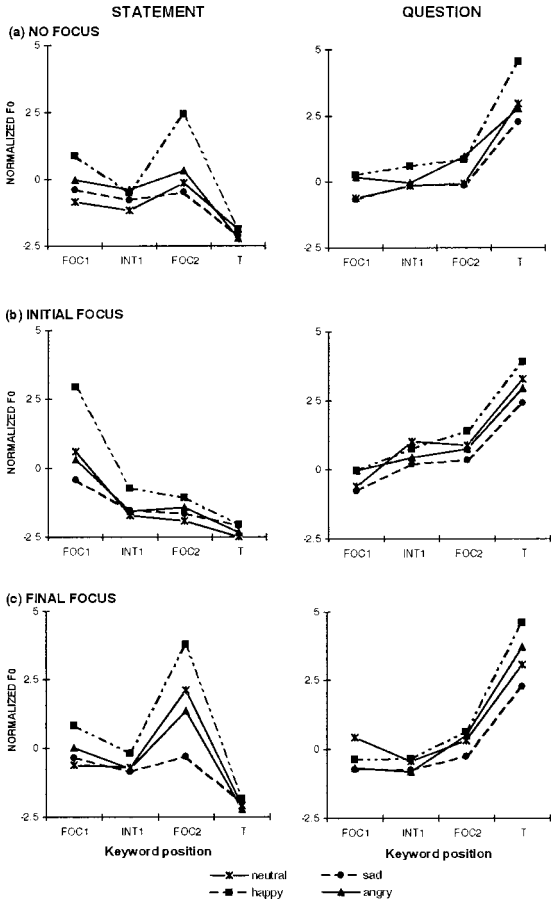


FIG. 5. Influence of emotion on the $f0$ of initial (FOC1), intervening (INT1), final (FOC2) and terminal (*T*) keyword vowels in short statements and questions in the (a) no focus, (b) initial focus, and (c) final focus context (averaged across eight speakers).

Varying emotional qualities of speech had comparatively little effect on *duration* attributes of keywords, affecting only FOC2 and in a manner unrelated to focus or any other prosodic manipulation [EMOTION: $F_{\mathrm{FOC2}}(3,21)=6.71$, $p<0.01$]. Speakers systematically prolonged final words in utterances expressed in a happy tone relative to those expressed in an angry, sad, or neutral tone in this position (which did not differ in any case). This effect did not interfere with temporal parameters of FOC2 vowels that signalled focus realization in the utterance, including post-focus vowel reduction in the production of neutral statements.

B. Utterance measures

Interplay of emotion and linguistic variables on global acoustic properties

Acoustic properties of whole utterances were analyzed to further illuminate potential emotion effects on target productions for three measures: speech rate, mean $f0$, and $f0$ range. For the speech rate data, significant effects were confined to LENGTH [$F_{\mathrm{RATE}}(1,7)=95.69$, $p<0.001$] and EMOTION [$F_{\mathrm{RATE}}(3,21)=28.14$, $p<0.001$]. Long utterances were produced at a faster rate than short utterances. The EMOTION main effect [depicted in Fig. 6(a)] was ex-

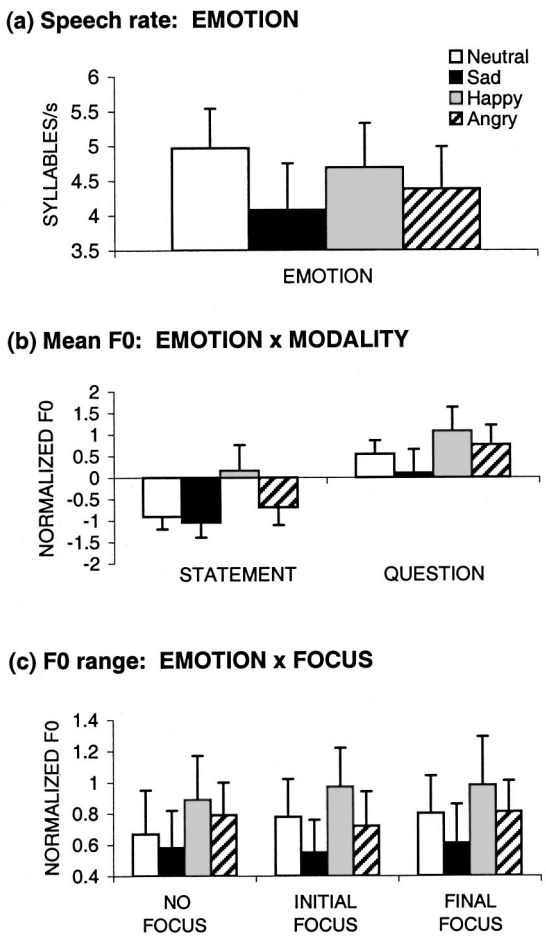


FIG. 6. Influence of emotion on three ''global'' measures of prosody as a function of relevant linguistic properties of the stimuli (averaged across eight speakers).

plained by an accelerated rate of delivery for neutral and happy utterances (which did not differ) relative to angry and sad utterances (which also did not differ). No interactions emerged from this analysis.

Analysis of mean utterance $f0$ yielded three related interactions centred around the MODALITY factor: LENGTH×MODALITY [$F_{\mathrm{MEAN}F0}(1,7)=64.67$, $p<0.001$]; FOCUS×MODALITY [$F_{\mathrm{MEAN}F0}(2,14)=28.03$, $p<0.001$]; and EMOTION×MODALITY [$F_{\mathrm{MEAN}F0}(3,21)=6.73$, $p<0.01$]. *Post hoc* comparisons indicated that in all conditions of each of these three interactions, questions were associated with a higher overall $f0$ than statements. However, for LENGTH×MODALITY, short utterances displayed a higher $f0$ mean than long utterances when formulated as a question but not when produced as a statement. The FOCUS×MODALITY interaction revolved around stimuli produced with *sentence-initial* focus: these tokens were characterized by a lower (statements) or higher (questions) mean $f0$ than matched stimuli in most other focus conditions. These trends were undoubtedly linked to the large post-focus fall or rise following FOC1 in this focus context, described earlier. The relationship between emotion and modality is presented in Fig. 6(b). It is shown that happy sentences exhibited a consistently higher mean $f0$ than sad or neutral sentences in both statements and questions. For statements, happy sentences were also higher than angry sentences, which did not differ from neutral or sad. For questions, angry and neutral sentences were associated with a comparable mean $f0$ and were both significantly higher than sad sentences.

Finally, analysis of $f0$ range yielded effects for EMOTION [$F_{F0\mathrm{RANGE}}(3,21)=20.27$, $p<0.001$] and EMOTION×FOCUS [$F_{F0\mathrm{RANGE}}(6,42)=4.88$, $p=0.01$], the interaction being supplied by Fig. 6(c). Happy utterances contained greater $f0$ variation than all other emotional modes examined (angry, neutral, sad). Angry utterances also demonstrated a larger range than sad utterances, and neutral utterances exceeded sad utterances in certain focus contexts (initial- and final-focus). Considered another way, it appeared that utterances with overt emotional features (happy, sad, angry) were associated with a relatively uniform $f0$ range across the three levels of focus, whereas the $f0$ range of neutral utterances was more dependent on focus context, displaying a larger range in contexts where focus was realized in the utterance (initial- and final-focus) than where it was not (no focus).

IV. DISCUSSION

Prosodic aspects of discourse provide a vehicle for speakers to concurrently express linguistic and emotional attributes of their mental states to listeners. Despite longstanding recognition of the multiple intentions represented in suprasegmental structure (e.g., Bolinger, 1986; Ladd, 1996), the acoustic realization of these conjoint intentions in spoken utterances has been accorded little empirical attention (cf. Cosmides, 1983; McRoberts *et al.*, 1995).

Discussion initially revolves around linguistic properties of the utterances sampled. There was a robust tendency for focused words in matched statements and questions to display higher $f0$ peaks and extended vowel segments than corresponding words produced without focus, supporting well-established data (Cooper *et al.*, 1985; Eady and Cooper, 1986; Ferreira, 1993; Weismer and Ingrisano, 1979). As was also expected, the local distribution of cues marking focus at designated keyword sites (FOC1 and FOC2 in the present study) was influenced by the *position* of focus realization in the utterance (Eady and Cooper, 1986; Weismer and Ingrisano, 1979). For both duration and $f0$, acoustic differentiation of the initial word of utterances (FOC1) was contingent on whether FOC1 was focused or not and did not vary for either parameter in the ''final'' and ''no'' focus speaking contexts. This description contrasts with how duration and $f0$ were regulated in sentence final position (FOC2); here, encoders differentiated ''focused'' elements from those in both ''unfocused'' contexts in the predicted manner. However, they also reduced the temporal and $f0$ prominence of FOC2 when occurring ''post-focus'' (i.e., in the ''initial'' focus context) relative to the cues provided in utterances that did not contain focus (Eady and Cooper, 1986; Weismer and Ingrisano, 1979 for duration only). Patterns observed for sentence-final words imply the operation of a ''retroactive'' strategy governing a speaker's on-line decision about focus production; the semantic value of initial focused words is reinforced by diminishing the duration and $f0$ salience of

later, potentially contrastive forms (Cooper *et al.*, 1985; Eady *et al.*, 1986). However, the same encoders did not supply ''anticipatory'' cues to focus produced at the end of utterances by attenuating the acoustic prominence of concepts at the onset of the stimulus (Eady and Cooper, 1986; Weismer and Ingrisano, 1979).

Further consideration of the keyword measures argues that for duration, focus effects are not always strictly localized to the target word as described for previous stimuli (Cooper *et al.*, 1985; Eady and Cooper, 1986). Rather, such effects manifest more broadly throughout carrier utterances as a function of stimulus characteristics. As recognized by Eady *et al.* (1986), acoustic studies that have concentrated their analyses on relatively long (10–12 syllables) utterances have reported localized durational effects for focus (Cooper *et al.*, 1985; Eady and Cooper, 1986) whereas those examining shorter (5–7 syllables) stimuli have not (Folkins *et al.*, 1975; Weismer and Ingrisano, 1979; also trends in Eady *et al.*, 1986). The impact of sentence length was directly testable in the current study which elicited six-syllable (short) and ten-syllable (long) utterances from a uniform sample of encoders.

Results confirmed that utterance length contributes to the temporal properties of focused versus unfocused words, although not entirely in the projected manner. ''Long'' utterances resembling those employed by Eady and Cooper (1986) and others exhibited fewer instances where the duration of neighboring words was modified due to focus placement (especially for final keywords). This pattern is suggestive of localized cue use. However, post-focal reduction in the duration of FOC2, a nonlocalized effect, was observed for both long and short utterances. Moreover, content words directly preceding FOC2 in long stimuli (*teapot* in *Mary sold the teapot for a dollar*) were sensitive to focus position, displaying reduced durations in *pre*-focal environments (Cooper *et al.*, 1985). Thus neither short nor long utterances contained purely localized alterations in duration due to the operation of linguistic focus. Possibly, syntactic-semantic attributes of experimental stimuli in which such pragmatic decisions are marked—which vary considerably across studies in this literature (Cooper *et al.*, 1985; Eady and Cooper, 1986; Folkins *et al.*, 1975; Weismer and Ingrisano, 1979)—further dictate the extent to which speakers manipulate prosodic aspects of the message to communicate focus, influencing the localizability of associated cues (particularly for long stimuli).

The $f0$ correlates of focus were not expected to be localized in the utterance but have a pervasive impact on the shape of the intonation contour. This assumption was borne out for both statements and questions, particularly in the initial-focus condition where marked falling (statement) or rising (question) $f0$ excursions characterized sentence-initial words and their immediate post-focus environment (Cooper *et al.*, 1985; Eady and Cooper, 1986; O'Shaughnessy, 1979). These abrupt transitions influenced the $f0$ of all subsequent points of the contour topline for this one focus condition when matched statements and questions in the other focus conditions were compared. In neutral- and final-focus utterances, the divergence between falling (statement) and rising (question) contours progressed more gradually and did not always differ significantly until the ''terminal'' point (i.e., final 150 ms) where all speakers reliably distinguished these two speech acts (Eady and Cooper, 1986; Hadding-Koch and Studdert-Kennedy, 1964; Lieberman, 1967; O'Shaugnessy, 1979; Studdert-Kennedy and Hadding, 1973).

Thus discontinuities in the $f0$ topline may furnish powerful clues to focus position within statements and questions. This being said, it is interesting to consider that for interrogative utterances, speakers did not furnish *any* statistically detectable $f0$ cues that biased the position of focus in the utterance when measures were limited to sentence-initial or sentence-final keywords (Eady and Cooper, 1986). A potentially related trend was noted for keyword duration, where acoustic differences among the three focus conditions were smaller and fewer in number for questions than for statements (although recall that focus lengthening was distinctive in both modality types). These findings imply that English utterances produced with the terminal yes/no interrogation rise place observable constraints on a speaker's capacity to modulate $f0$ and duration to realize focus using the range of cues typical of matched words produced in declarative utterances (McRoberts *et al.*, 1995). Presumably, such prosodic differences have a negligible effect on *listeners'* ability to infer what linguistic elements are highlighted by the speaker in declarative versus interrogative speech acts, although the present study was not designed to address this issue (perceptual evaluation of speakers was restricted to declarative renditions of the stimuli).

Specifying the acoustic form of 'basic' emotional displays was estimated via three frequently cited parameters of emotional vocalizations: mean $f0$, $f0$ range, and speech rate (Frick, 1985; Scherer, 1986). In broad terms, results obtained here for elderly speakers accord with past descriptions of younger subjects engaged in simulated or naturalistic displays of vocal emotion. Specifically, sad utterances were spoken with a relatively slow rate, a low mean $f0$ and highly restricted $f0$ range (Breitenstein *et al.*, in press; Huttar, 1968; Johnson *et al.*, 1986; Williams and Stevens, 1972). Neutral utterances were reliably faster than sad utterances (Breitenstein *et al.*, in press; Fairbanks and Hoaglin, 1941; Williams and Stevens, 1972) with a somewhat higher mean $f0$ and extended $f0$ range than sad in many conditions. Happy utterances exhibited a higher mean $f0$ and broader $f0$ range than sad or neutral stimuli (Fonagy, 1978; Huttar, 1968; Scherer, 1974). In the current data set, angry utterances also displayed a higher overall $f0$ and broader $f0$ range than sad and neutral stimuli, and happy exceeded angry stimuli on these dimensions in many speaking conditions (cf. Williams and Stevens, 1972, for a comparison of happy and angry).[3] Finally, happy vocalizations tended to be faster than angry or sad utterances for the present stimuli (happy=neutral), although attempts to characterize the rate of ''happy'' voices have led to mixed conclusions overall (Murray and Arnott, 1993). As an overview, the current set of measures—while clearly inadequate descriptors of the array of vocal features contributing to emotional communication for any speaker group (e.g., ''angry'' voices may rely heavily on aspects of voice quality; e.g., Cummings and

Clements, 1995; Fonagy and Magdics, 1963; Johnson *et al.*, 1986)—were nonetheless relatively successful in separating the four emotional targets at the acoustic level, and in a manner which was largely predicted by past research (Murray and Arnott, 1993; Scherer, 1986).

The impact of simulated emotion on the acoustic correlates of linguistic focus and utterance modality was of central interest here. First, note that emotion was not expected to substantially alter the pattern of acoustic associations between focused and unfocused words beyond that related to utterance modality and this prediction was confirmed. However, the especially vital contribution of pitch/$f0$ in communicating both linguistic and emotional aspects of target messages yielded meaningful interactions among levels of emotion, focus, and modality when $f0$ parameters of sentence-initial and sentence-final keywords were inspected. Notably, speaking conditions in which encoders were required to modulate $f0$ to convey both focus and the ''marked'' terminal rise of questions in English (Lieberman, 1967) were associated with an *elimination* of $f0$ distinctions conducive to different emotional interpretations at the location of key content words. Such tendencies were not witnessed for declarative speech acts which seemed to invest speakers with greater flexibility in manipulating emotion-related $f0$ parameters in most focus contexts, permitting the relatively distinctive array of acoustic gestures typical of these four emotions (Murray and Arnott, 1993). Presumably, past data on the vocal characterization of emotion have been derived almost exclusively for declarative utterances, explaining previous failures to detect such modality-related differences.

Thus there are tentative indications in the data that modulating $f0$ parameters that preserve the communicative features of both focus position and yes/no questions mitigate the scope of *emotion*-related $f0$ differences permitted on key content words when compared to those expressed in declarative utterances. Certainly, this is not to claim that emotion markers were absent in interrogative contours; in addition to speech rate and other unexplored parameters of emotion that were undoubtedly manifest in both types of speech acts, analysis of mean utterance $f0$ revealed reliable cross-emotion trends that were qualitatively similar for statements and questions (e.g., happy exhibited a consistently higher overall $f0$ than sad and neutral). As such, it may be concluded that emotion-related $f0$ characteristics of declarative and interrogative speech acts are encoded consistently as a global function of the intonation contour (e.g., Ladd *et al.*, 1985; Scherer, 1986), but that for interrogative utterances, emotional differences in $f0$ are comparatively opaque at points where contrastive focus is operational. At such points in the contour, $f0$ features of the four emotions tend to converge.

Given these interactive effects on $f0$, one may speculate that the need to assign contrastive focus within rising intonation contours—joint linguistic intentions achieved through regulation of vocal fold tension due to laryngeal (primarily cricothyroid) activity (Atkinson, 1973; Lieberman *et al.*, 1970)—places unique constraints on English speakers in the modulation of $f0$ for *additional* purposes such as emotional inflection, at least in simulated speaking contexts. This explanation, which appeals to the idea of the ''prosodic load'' of an utterance and its impact on speakers, resembles an account proposed by McRoberts and colleagues (1995) in a related study of simple English utterances (e.g., *November*). Those authors attributed a trade-off in the ability to program $f0$ for simultaneous *linguistic* purposes (stress and interrogation rise) to articulatory constraints imposed on their four male speakers. However, requiring the same speakers to manipulate the affective (positive/negative) valence of interrogative stimuli did not yield a similar $f0$ trade-off for their stimuli, suggesting to the investigators that affective and linguistic programming of $f0$ may be functionally separate (McRoberts *et al.*, 1995).

Many important differences characterize the methods and stimuli employed here and in this earlier investigation, especially the form in which emotionally inflected speech was characterized and evaluated for experimental stimuli. Furthermore, McRoberts and colleagues did not require their subjects to manipulate focus, speech acts, *and* emotional attributes in tandem (the effects of focus and affect on interrogative contours were evaluated in separate conditions). As such, the prosodic load and its presumed effect on $f0$ implementation was greater for the current set of speakers. Indeed, it is precisely this need to combine ''marked'' linguistic-prosodic forms and specific emotional targets that imposed the most severe articulatory limitations on the present speakers, implying that linguistic and emotional uses of $f0$ are not always functionally separate in speech production. Further work on the interaction between emotional and linguistic aspects of prosody will help shed light on these promising, yet preliminary hypotheses.

Brief commentary is reserved for data on the form of ''happy'' utterances. In this condition, the duration of sentence-final keywords was robustly longer than in the other three emotional modes independent of a broad range of variables (focus position, modality, or sentence length). Moreover, the $f0$ glide of questions terminated at a significantly higher point when speakers were happy than in some other emotional modes (in contrast, the endpoint of statements was a stable feature across emotional contexts; Liberman and Pierrehumbert, 1984). Differences in the terminal point of the interrogative rise, although consistent with described trends in the $f0$ data for happy (and formulations proposed by Ohala, 1983), are again curious in light of McRoberts *et al.*'s (1995) data; those authors reported that the positive or negative valence of short utterances had no effect on the extent of $f0$ rise marking yes/no questions. This discrepancy will again benefit from future testing employing a relatively broad range of stimuli.

Evidence of the impact of emotion on aspects of intonation and word duration underscore the diverse ways in which humans modulate a small set of acoustic parameters to communicate complex linguistic and emotional intentions. By the same token, findings are limited by the restricted purview of the acoustic measures employed here which do not capture the range of cues available to speakers to signal such intentions (particularly for emotional expression; e.g., Scherer, 1986). Despite relative success in simulating three

prototypical emotional contexts for analysis, future endeavors which address the influence of emotion on intonation will benefit from recordings of naturalistic emotional displays, and a broader range of emotions to weigh against the current data (e.g., ''fear'' and ''disgust'' may constitute basic human emotions; Ekman, 1992; Izard, 1977). The important role of gender on the ability to encode emotion is also showcased here, as the four individuals who were best able to communicate emotional qualities to listeners were all female (Sobin and Alpert, 1999; Zuckerman *et al.*, 1975).

ACKNOWLEDGMENTS

Special thanks to Marta Fundamenski, Nazma Mohammed, and Anita Shuper for help in subject testing and manuscript preparation, and to Dr. A. Löfqvist and two anonymous reviewers for valuable comments received on an earlier version of this paper. This research was supported by the Québec Fonds pour la formation des chercheurs et l'aide à la recherche (FCAR) and a Fraser, Monat, and McPherson scholarship awarded by the McGill Faculty of Medicine.

[1]It is assumed that speech samples elicited by Cooper, Eady, and others resemble emotionally ''neutral'' speech generated in the current paradigm.

[2]Overall values are not inconsistent with related studies obtaining perceptual ratings of prosody from untrained listeners using a forced-choice paradigm (e.g., Williams and Stevens, 1972).

[3]Frick (1985) and others have distinguished between anger which represents ''frustration'' versus ''threat,'' with only the former correlating with raised pitch (Fairbanks and Pronovost, 1939; Williams and Stevens, 1972). Contextual cues provided in this study biased responses indicative of ''frustrated'' anger, yielding the predicted rise in mean $f0$.

Atkinson, J. (**1973**). ''Aspects of intonation in speech: implications from an experimental study of fundamental frequency,'' unpublished doctoral dissertation, University of Connecticut.

Bachorowski, J. (**1999**). ''Vocal expression and perception of emotion,'' Curr. Dir. Psychol. Sci. **8**, 53–57.

Bachorowski, J., and Owren, M. J. (**1995**). ''Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context,'' Psycholog. Sci. **6**, 219–224.

Banse, R., and Scherer, K. (**1996**). ''Acoustic profiles in vocal emotion expression,'' J. Personality Soc. Psychol. **70**, 614–636.

Behrens, S. J. (**1988**). ''The role of the right hemisphere in the production of linguistic stress,'' Brain Lang. **33**, 104–127.

Bolinger, D. (**1978**). ''Intonation across languages,'' in *Universals of Human Language*, edited by J. H. Greenberg (Stanford University Press, Stanford, CA), pp. 471–524.

Bolinger, D. (**1986**). *Intonation and its Parts* (Stanford University Press, Stanford, CA).

Bolinger, D. L. (**1955**). ''Intersections of stress and intonation,'' Word **11**, 195–203.

Bolinger, D. L. (**1958**). ''A theory of pitch accent in English,'' Word **14**, 109–149.

Breitenstein, C., Van Lancker, D., and Daum, I. (in press). ''The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample,'' Cognition and Emotion.

Brown, W. S., and McGlone, R. E. (**1974**). ''Aerodynamic and acoustic study of stress in sentence productions,'' J. Acoust. Soc. Am. **56**, 971–974.

Cooper, W., and Sorensen, J. (**1981**). *Fundamental Frequency in Sentence Production* (Springer, New York).

Cooper, W. E., Eady, S. J., and Mueller, P. R. (**1985**). ''Acoustical aspects of contrastive stress in question-answer contexts,'' J. Acoust. Soc. Am. **77**, 2142–2156.

Cosmides, L. (**1983**). ''Invariances in the acoustic expression of emotion during speech,'' J. Exp. Psychol. **9**, 864–881.

Cummings, K. E., and Clements, M. A. (**1995**). ''Analysis of the glottal excitation of emotionally styled and stressed speech,'' J. Acoust. Soc. Am. **98**, 88–98.

Davitz, J. R. (**1964**). ''Auditory correlates of vocal expressions of emotional meanings,'' in *The Communication of Emotional Meaning*, edited by J. R. Davitz (McGraw-Hill, New York), pp. 101–112.

Denes, P. (**1959**). ''A preliminary investigation of certain aspects of intonation,'' Lang. Speech **2**, 107–122.

Eady, S. J., and Cooper, W. E. (**1986**). ''Speech intonation and focus location in matched statements and questions,'' J. Acoust. Soc. Am. **80**, 402–415.

Eady, S. J., Cooper, W. E., Klouda, G. V., Mueller, P. R., and Lotts, D. W. (**1986**). ''Acoustical characteristics of sentential focus: Narrow vs broad and single vs. dual focus environments,'' Lang Speech **29**, 233–251.

Eefting, W. (**1990**). ''The effect of ''information value'' and ''accentuation'' on the duration of Dutch words, syllables, and segments,'' J. Acoust. Soc. Am. **89**, 412–424.

Ekman, P. (**1992**). ''An argument for basic emotions,'' Cognition and Emotion **6**, 169–200.

Ekman, P., Levenson, R. W., and Friesen, W. V. (**1983**). ''Autonomic nervous system activity distinguishes among emotions,'' Science **221**, 1208–1210.

Fairbanks, G., and Hoaglin, L. W. (**1941**). ''An experimental study of the durational characteristics of the voice during the expression of emotion,'' Speech Monographs **8**, 85–90.

Fairbanks, G., and Pronovost, W. (**1939**). ''An experimental study of the pitch characteristics of the voice during the expression of emotion,'' Speech Monographs **6**, 87–104.

Ferreira, F. (**1993**). ''Creation of prosody during sentence production,'' Psychol. Rev. **100**, 233–253.

Folkins, J. W., Miller, C. J., and Minifie, F. D. (**1975**). ''Rhythm and syllable timing in phrase level stress patterning,'' J. Speech Hear. Res. **18**, 739–753.

Fonagy, I. (**1978**). ''A new method of investigating the perception of prosodic features,'' Lang Speech **21**, 34–49.

Fonagy, I., and Magdics, K. (**1963**). ''Emotional patterns in intonation and music,'' Z. Phonetik **16**, 293–326.

Fowler, C., and Housum, J. (**1987**). ''Talkers' signaling of ''new'' and ''old'' words in speech and listeners' perception and use of the distinction,'' J. Memory Lang. **26**, 489–504.

Frick, R. W. (**1985**). ''Communicating emotion: The role of prosodic features,'' Psychol. Bull. **97**, 412–429.

Fry, D. B. (**1955**). ''Duration and intensity as physical correlates of linguistic stress,'' J. Acoust. Soc. Am. **27**, 765–768.

Fry, D. B. (**1958**). ''Experiments in the perception of stress,'' Lang Speech **1**, 126–152.

Gandour, J., Larsen, J., Dechongkit, S., Ponglorpisit, S., and Khunadorn, F. (**1995**). ''Speech prosody in affective contexts in Thai patients with right hemisphere lesions,'' Brain Lang. **51**, 422–443.

Hadding-Koch, K., and Studdert-Kennedy, M. (**1964**). ''An experimental study of some intonation contours,'' Phonetica **11**, 175–185.

Huttar, G. (**1968**). ''Relations between prosodic variables and emotions in normal American English utterances,'' J. Speech Hear. Res. **11**, 467–480.

Izard, C. E. (**1977**). *Human Emotions* (Plenum, New York).

Johnson, W. F., Emde, R. N., Scherer, K. R., and Klinnert, M. D. (**1986**). ''Recognition of emotion from vocal cues,'' Arch. Gen. Psychiatry **43**, 280–283.

Klatt, D. H. (**1976**). ''Linguistic uses of segmental duration in English: Acoustic and perceptual evidence,'' J. Acoust. Soc. Am. **59**, 1208–1221.

Ladd, D. R. (**1996**). *Intonational Phonology* (Cambridge University Press, Cambridge).

Ladd, D. R., Silverman, K. E. A., Tolkmitt, F., Bergmann, G., and Scherer, K. R. (**1985**). ''Evidence for the independent function of intonation contour type, voice quality, and $F0$ range in signaling speaker effect,'' J. Acoust. Soc. Am. **78**, 435–444.

Lea, W. (**1977**). ''Acoustic correlates of stress and juncture,'' in *Studies in Stress and Accent*, edited by L. M. Hyman (Dept. of Linguistics, U.C.L.A., Los Angeles), pp. 83–120.

Liberman, M., and Pierrehumbert, J. (**1984**). ''Intonational invariance under changes in pitch range and length,'' in *Language Sound and Structure*, edited by M. Aronoff and R. Oehrle (MIT Press, Cambridge, MA), pp. 157–233.

Lieberman, P. (**1967**). *Intonation, Perception, and Language* (MIT Press, Cambridge, MA).

Lieberman, P., and Michaels, S. B. (**1962**). ''Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech,'' J. Acoust. Soc. Am. **34**, 922–927.
Lieberman, P., Sawashima, M., Harris, K., and Gay, T. (**1970**). ''The articulatory implementation of the breath-group and prominence: Cricothyroid muscular activity in intonation,'' Language **46**, 312–327.
Majewski, W., and Blasdell, R. (**1968**). ''Influence of fundamental frequency cues on the perception of some synthetic intonation contours,'' J. Acoust. Soc. Am. **45**, 450–457.
Max, L., and Onghena, P. (**1999**). ''Some issues in the statistical analysis of completely randomized and repeated measures designs for speech, language, and hearing research,'' J. Speech Lang Hear Res. **42**, 261–270.
McClean, M. D., and Tiffany, W. R. (**1973**). ''The acoustic parameters of stress in relation to syllable position, speech loudness and rate,'' Lang. Speech **16**, 283–290.
McRoberts, G. W., Studdert-Kennedy, M., and Shankweiler, D. P. (**1995**). ''The role of fundamental frequency in signaling linguistic stress and affect: Evidence for a dissociation,'' Percept. Psychophys. **57**, 159–174.
Mertus, J. (**1989**). *BLISS User's Manual* (Brown University, Providence, RI).
Morton, J., and Jassem, W. (**1965**). ''Acoustic correlates of stress,'' Lang. Speech **8**, 159–181.
Murray, I. R., and Arnott, J. L. (**1993**). ''Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion,'' J. Acoust. Soc. Am. **93**, 1097–1108.
Ohala, J. (**1983**). ''Cross-language use of pitch: An ethological view,'' Phonetica **40**, 1–18.
O'Shaughnessy, D. (**1979**). ''Linguistic features in fundamental frequency patterns,'' J. Phonetics **7**, 119–145.
Ouellette, G. P., and Baum, S. R. (**1993**). ''Acoustic analysis of prosodic cues in left- and right-hemisphere-damaged patients,'' Aphasiology **8**, 257–283.
Pakosz, M. (**1983**). ''Attitudinal judgments in intonation: Some evidence for a theory,'' J. Psycholinguist. Res. **12**, 311–326.
Pell, M. D. (**1999a**). ''Fundamental frequency encoding of linguistic and emotional prosody by right hemisphere-damaged speakers,'' Brain Lang. **69**, 161–192.
Pell, M. D. (**1999b**). ''Some acoustic correlates of perceptually ''flat affect'' in right-hemisphere damaged speakers,'' Brain Cogn. **40**, 219–223.
Pell, M. D. (**1999c**). ''The temporal organization of affective and nonaffective speech in patients with right-hemisphere infarcts,'' Cortex **35**, 455–477.
Rose, P. (**1987**). ''Considerations in the normalization of the fundamental frequency of linguistic tone,'' Speech Commun. **6**, 343–351.
Ross, E. D., Edmondson, J. A., and Seibert, G. B. (**1986**). ''The effect of affect on various acoustic measures of prosody in tone and nontone languages: A comparison based on computer analysis of voice,'' J. Phonetics **14**, 283–302.
Ryalls, J., Le Dorze, G., Lever, N., Ouellet, L., and Larfeuil, C. (**1994**). ''The effects of age and sex on speech intonation and duration for matched statements and questions in French,'' J. Acoust. Soc. Am. **95**, 2274–2276.
Scherer, K. R. (**1974**). ''Acoustic concomitants of emotional dimensions: Judging affect from synthesized tone sequences,'' in *Non-Verbal Communication*, edited by S. Weitz (Oxford University Press, New York), pp. 105–111.
Scherer, K. R. (**1986**). ''Vocal affect expression: A review and a model for future research,'' Psychol. Bull. **99**, 143–165.
Siegman, A. W., and Boyle, S. (**1993**). ''Voices of fear and anxiety and sadness and depression: The effects of speech rate and loudness on fear and anxiety and sadness and depression,'' J. Abnorm. Psychol. **102**, 430–437.
Siegman, A. W., Dembroski, T. M., and Crump, D. (**1992**). ''Speech rate, loudness, and cardiovascular reactivity,'' J. Behav. Med. **15**, 519–532.
Sobin, C., and Alpert, M. (**1999**). ''Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy,'' J. Psycholinguist. Res. **23**, 347–365.
Streeter, L., MacDonald, N., Apple, W., Krause, R., and Galotti, K. (**1983**). ''Acoustic and emotional indicators of emotional stress,'' J. Acoust. Soc. Am. **73**, 1354–1360.
Studdert-Kennedy, M., and Hadding, K. (**1973**). ''Auditory and linguistic processes in the perception of intonation contours,'' Lang. Speech **16**, 293–313.
Turk, A. E., and Sawusch, J. R. (**1996**). ''The processing of duration and intensity cues to prominence,'' J. Acoust. Soc. Am. **99**, 3782–3790.
Uldall, E. (**1960**). ''Attitudinal meanings conveyed by intonation contours,'' Lang. Speech **3**, 223–234.
Weismer, G., and Ingrisano, D. (**1979**). ''Phrase-level timing patterns in English: Effects of emphatic stress location and speaking rate,'' J. Speech Hear. Res. **22**, 516–533.
Williams, C. E., and Stevens, K. N. (**1972**). ''Emotions and speech: Some acoustical correlates,'' J. Acoust. Soc. Am. **52**, 1238–1250.
Zuckerman, M., Lipets, M., Hall, J., and Rosenthal, R. (**1975**). ''Encoding and decoding nonverbal cues of emotion,'' J. Pers. Soc. Psychol. **32**, 1068–1076.

A study of sentence stress production in Mandarin speakers of American English

Yang Chen[a)]
Division of Speech-Language Pathology and Audiology, University of Wyoming, Laramie, Wyoming 82071

Michael P. Robb, Harvey R. Gilbert, and Jay W. Lerman
Department of Communication Sciences, University of Connecticut, Storrs, Connecticut 06269

(Received 7 August 2000; accepted for publication 12 January 2001)

Acoustic characteristics of American English sentence stress produced by native Mandarin speakers are reported. Fundamental frequency (F_0), vowel duration, and vowel intensity in the sentence-level stress produced by 40 Mandarin speakers were compared to those of 40 American English speakers. Results obtained from two methods of stress calculation indicated that Mandarin speakers of American English are able to differentiate stressed and unstressed words according to features of F_0, duration, and intensity. Although the group of Mandarin speakers were able to signal stress in their sentence productions, the acoustic characteristics of stress were not identical to the American speakers. Mandarin speakers were found to produce stressed words with a significantly higher F_0 and shorter duration compared to the American speakers. The groups also differed in production of unstressed words with Mandarin speakers using a higher F_0 and greater intensity compared to American speakers. Although the acoustic differences observed may reflect an interference of L1 Mandarin in the production of L2 American English, the outcome of this study suggests no critical divergence between these speakers in the way they implement American English sentence stress. [DOI: 10.1121/1.1356023]

PACS numbers: 43.70.Fq [AL]

I. INTRODUCTION

Individuals who speak English as a second language (L2) vary in their ability to produce phonetic features of English precisely. Phonetic characteristics of the native or first language (L1) are thought to interfere with the production of L2 (Cheng, 1987). The L1 interference with L2 production can occur at both the segmental and suprasegmental level (Ingram and Pittam, 1997; Shen, 1990). The more interference which exists between L1 and L2, the less likely phonetic features of the L2 will be produced accurately. There have been many acoustic and perceptual studies examining the effects of various Asian languages on the phonetic features of English spoken as L2 (Flege, 1989; Flege and Davidian, 1985; Hutchinson, 1973; Ingram and Park, 1997; Ingram and Pittam, 1987; Kim, 1972; Magnuson and Akahane-Yamada, 1996; Pittam and Ingram, 1992; Robson, 1982; Tarone, 1980). The magnitude of phonetic inaccuracies appears to be correlated to the amount of experience (or length of time) speaking English, or the age period during which L2 was acquired (Flege, 1995; Guion *et al.*, 2000; Johnson and Newport, 1991). It is believed that after a ''critical period,'' L2 learners have difficulty not only accurately articulating L2 segments (Flege, 1987; Lennenberg, 1967; Magnuson and Akahane-Yamada, 1996; Pittam and Ingram, 1992), but also in acquiring the suprasegmental features of L2 (Chun, 1982; Guion *et al.*, 2000; Guzma, 1973; Scuffil, 1982).

One aspect of American English noted to be of difficulty for native speakers of tonal languages, specifically East Asian languages, is the production of stress placed on syllables or words (Cheng, 1968, 1987; Chun, 1982). Wijk (1966) noted over 30 years ago that correct stressing of words presents a major difficulty in the pronunciation of English for individuals who learn English as L2. There are two forms of stress which occur in the production of American English: lexical stress and sentence stress. Lexical stress is concerned with the emphasis of individual syllables comprising a polysyllabic word. Sentence stress is concerned with the stress placed on words in order to indicate (or contrast) their importance/prominence in relation to other words in a sentence. Varying degrees of syllable/word stress are indicated by changes in vocal fundamental frequency (F_0), vowel duration, and vowel intensity (Berinstein, 1979; Bolinger, 1958; Crystal, 1969; Fry, 1958; Potisuk *et al.*, 1996).

The focus of the present study was on the production of American English sentence stress by native speakers of Mandarin.[1] Clear differences exist in the segmental and suprasegmental characteristics of Mandarin and English (Chao, 1948, 1968; Cooper and Sorensen, 1981; Ho, 1976; Kratochvil, 1962; Lehiste, 1970; Pike, 1945; Tseng, 1981). Such being the case, we wished to determine whether L1 Mandarin interferes with the production of sentence stress in L2 English. Prior to undertaking this research, literature related to sentence stress production in American English and Mandarin was reviewed. Results of this review were used to formulate specific research predictions. The literature review and predictions follow.

A. Sentence stress in English

English is a polysyllabic language with diverse syllable structure. English is often described as a stress-timed language, whereby the speech rhythm of English involves an

a)Author to whom correspondence should be addressed. Electronic mail: yc@uwyo.edu

interplay of prominent (or long duration) syllables and less prominent (or short duration) syllables (Chun, 1982; van Santen and Shih, 2000).[2] English is also a nontone language whereby the meaning of a word is derived from the phonetic composition of the word. In the case of English, tones are part of what is usually called intonation, and can be spread across any number of syllables (Chun, 1982; Clark and Yallop, 1995). The primary acoustic correlate of a tone is the F_0 contour and in English the fundamental tonal choice is between rise and fall. The tone is normally placed on the last word of an utterance to convey either a statement (''She gave him the *KEYS*.'') or a question (''She gave him the *KEYS*?''). However, tones can also be placed on virtually any syllable within a sentence (i.e., sentence stress), in which case the tone has contrastive value (''He doesn't live *IN* Auckland'' [but nearby]) (Clark and Yallop, 1995).

There have been several examinations of the sentence stress of English (Brown and McGone, 1974; Crystal, 1969; Fromkin and Ohala, 1968; Ohala and Hirano, 1967; Lieberman, 1960, 1967; Pike, 1945). Notable are the studies reported by Cooper and his colleagues (Cooper *et al.*, 1985; Cooper and Sorensen, 1981; Eady and Cooper, 1986; Eady *et al.*, 1986). Results from a majority of studies examining sentence stress indicate that the most consistent acoustic correlates underlying stress are an increase in magnitude of F_0, duration, and intensity of stressed words in relation to unstressed words. However, an increase in magnitude of each acoustic correlate is not a necessary requirement for stress to occur. Liberman (1967) noted that typically two of these three correlates show an increase in magnitude during production of sentence stress.

B. Sentence stress in Mandarin

A syllable in Mandarin consists of segmental and suprasegmental features (Shih, 1986). Segmental features consist of a vocalic nucleus which may or may not be accompanied by pre-vocalic and post-vocalic consonants (Chao, 1968). The syllable structure of Mandarin, unlike English, is monosyllabic with primarily a basic CV word shape (Guo, 1992). Suprasegmental features include four ''basic'' tones[3] (Cheng, 1968; Gandour, 1978; Howie, 1976; Leben, 1978). In a tone language such as Mandarin, a change in tone of a syllable (i.e., a word) leads to a change of lexical meaning (Chao, 1948; Cheng, 1968; Chun, 1982; Dreher and Lee, 1966; Tseng, 1981; Pike, 1948). Mandarin is also a syllable-timed language and is generally thought of as showing no strong pattern of stress, with syllable duration remaining relatively constant across a sentence (Clark and Yallop, 1995; van Santen and Shih, 2000).

There are perceptual and acoustic studies which demonstrate that sentence stress occurs in the production of Mandarin. For example, Chao (1932) and Chun (1982) examined sentence stress as perceived in the production of Mandarin phrases (i.e., a sequential CV production). Results from both studies indicated that sentence stress in Mandarin can be achieved through differential use of F_0 or by lengthening of syllable duration, depending on the position of the stressed word in the sentence. The acoustic studies performed by Lin and Yan (1980) and Shen (1993) found Mandarin speakers to produce stressed words with higher intensity compared to unstressed words. Cao (1986) and Shen (1993) reported lengthening of stressed words compared to unstressed words, and Xu (1999) found Mandarin speakers to vary F_0 during the production of Mandarin phrases to differentiate stressed from unstressed words.

C. Sentence stress in Mandarin speakers of American English

When Mandarin speakers produce English, they may encounter difficulty determining the appropriate location for stress within a sentence (Cheng, 1987; Chun, 1982). For example, Chun (1982) reported that L1 Mandarin speakers learning English were perceived to either misplace sentence stress or produce stressed words with abnormally short durations. Chun (1982) attributed the findings to the lack of a fixed pitch pattern in Mandarin sentences and to the syllable-timed nature of Mandarin, which appeared to interfere with the production of English. To date, information related to English sentence stress production by Mandarin speakers is limited to perceptual judgments. An acoustic examination of the production of American English sentence stress by native speakers of Mandarin remains to be performed.

Based on what is known about the characteristics of sentence stress in the English and Mandarin languages, three predictions were made regarding the production of English sentence stress by native speakers of Mandarin. The predictions were developed with reference to the specific influence of F_0, duration, and intensity on sentence stress. The first prediction was that Mandarin speakers will not significantly differ from American English speakers in their use of F_0 to signal stress in the production of English sentences. Although tonal languages such as Mandarin are assumed to alter F_0 exclusively to signal a change in word meaning rather than intent (Chao, 1948; Howie, 1976), the recent report by Xu (1999) would suggest otherwise. The second prediction was that Mandarin speakers will show no significant differences from American English speakers in their use of duration to signal English sentence stress. Being a syllable-timed language, Mandarin might be expected to show minimal variation in syllable duration across a phrase; however, Cao (1986) and Shen (1993) have found contrary evidence. The third prediction was that Mandarin speakers will show no significant differences from American English speakers in their use of intensity to signal sentence stress. Mandarin speakers have been shown to alter intensity to signify stress in phrases (Lin and Yan, 1980; Shen, 1993). It remains to be determined whether native Mandarin speakers are capable of using intensity to signify stress in English sentences.

II. METHOD

A. Participants

Two groups of subjects were recruited from within the University of Connecticut community (Storrs, CT). The first group consisted of 40 adults (20 males, 20 females) who spoke Mandarin as L1 and American English as L2. All Mandarin subjects were born in mainland China. The average age of the Mandarin males was 33 years

(range=30–46 years). The average age of the Mandarin females was 28 years (range=21–42 years). Selection criteria for inclusion in the Mandarin group consisted of: (1) a university education; (2) formal instruction in English; (3) the ability to speak standard Mandarin as judged by the first author (who is a native speaker of Mandarin); (4) speaking English for a minimum of 30% of their daily conversation; (5) the ability to orally read English fluently; and (6) residence in the U.S. for a minimum of 2 years. The average length of U.S. residence was 3 years, 9 months for Mandarin female speakers (range=2–12 years), and 4 years, 7 months for Mandarin male speakers (range=2–17 years). The second group consisted of 40 adults (20 males, 20 females) who spoke American English as L1. The average ages of the American male and female speakers were 33 years (range =22–46 years) and 27 years (range=23–41 years), respectively. All subjects (native Mandarin and English) were judged to be free of speech, language, or hearing disorders.[4]

B. Speech materials

Sentence stress was evaluated during production of the sentence, ''I bought a cat there.'' The sentence was produced four ways with primary stress placed on one of four different words. The four sentences with varying stress (noted in uppercase italics) were: ''*I* bought a cat there,'' ''I *BOUGHT* a cat there,'' ''I bought a *CAT* there,'' and ''I bought a cat *THERE*.'' A similar procedure was used by Copper and Sorenson (1981). Each of the 4 sentences was produced 3 times by each participant for a total of 12 sentences per participant. Individuals were allowed to practice the speech materials as often as they wished prior to the actual audio recordings. A sentence was deemed acceptable if it was perceptually fluent, contained no misarticulations, conformed to the prescribed placement of primary stress, and was judged by one of the researchers to be produced at a conversational pitch and loudness. For the present study, sentence stress was defined as the stressed placed on the vocalic nucleus of each target word.

C. Audio recordings

All recordings took place in a sound-attenuated booth. Order of sentence presentation was randomized across participants. Audio recordings were made using a cassette recorder (Marantz PMD-360) and a unidirectional dynamic microphone (Shure, 515SD). The microphone was placed approximately 20 cm from the speaker's mouth. Three pure tones of varying intensities were recorded on audiotape for later intensity calibration. A sound level meter was placed alongside the microphone for recording the actual sound intensity level. Gain control settings remained in fixed positions during the recording of the calibration and speech signals.

D. Acoustic analysis

Each of the four sentences, along with the corresponding intensity calibration signals, were introduced via a cassette recorder (Nakamichi, MR-2) into a computer (CPU586). All acoustic signals were digitized at 10 000 Hz using a speech analysis system (CSL 4300B). Each sentence produced by each subject was displayed as an amplitude-by-time waveform. The four target words in each sentence recitation were measured for F_0, intensity, and vowel duration. Measurements of F_0, intensity, and duration were made only for the vowel nucleus of the four target words. The measurements performed were as follows.

1. Fundamental frequency (F_0)

Three portions of the vocalic nucleus were measured for F_0. A 50-ms windowed cursor was placed at beginning, midpoint, and end locations of the vowel segment, respectively. To minimize possible coarticulatory influences of pre- and post-vocalic consonants on the vowel of interest, the beginning of the vowel was defined as the first 50 ms of periodic activity following the third glottal pulse. The end of the vowel was defined as the last 50 ms of periodic activity preceding the last three glottal pulses. Vowel midpoint was the mathematical middle point between the beginning and end points. Once the 50-ms window was fixed on the waveform, the signal was transformed into an amplitude-by-frequency ''power'' spectrum using a Hamming window weighting. The center frequency of the first harmonic peak was extracted to represent the F_0. Instances when the resulting F_0 value was thought be questionable were handled by re-examining the vowel nucleus using a narrow band (45-Hz) spectrogram. The F_0 determined for each vocalic nucleus was based on the median of the F_0 values identified at the three measurement locations. Group data were calculated as an average of the median values.

2. Vowel intensity

Five data points (i.e., dB values) were extracted from each of the three calibration tones, and a calibration equation was then generated by using a linear regression method (Ng, 1996). The calibration equation was used to calculate the intensity level in dB SPL during the speech sample recordings. The onset and offset of each vowel sample was demarcated using the same beginning and end locations used to measure the F_0 for each target word. The demarcated vowel was then transformed into an energy-by-time plot. The same three vowel portions used for F_0 measurement were extracted from the energy-by-time plot and the intensity levels were obtained. These intensity values were then ''corrected'' on the basis of the calibration equation to determine the true intensity values. The intensity value for each vocalic nucleus was based on the median of the intensity values identified at the three measurement locations. Group data were calculated as an average of the median values.

3. Vowel duration

Measurement of vowel duration for each target word was based on the amplitude-by-time waveform display. Vertical cursors were manually placed at the first and last glottal pulses of the vowel to demarcate the onset and offset points. The time interval between the two cursors was taken to be the vowel duration.

TABLE I. Mean (M), standard deviation (s.d.), and range values for F_0 (Hz), intensity (dB), and duration (ms) for each stressed word produced by the group of American (A) female and Mandarin (M) female speakers.

Group	Stressed word	F_0 (Hz)			Intensity (dB)			Duration (ms)		
		M	s.d.	Range	M	s.d.	Range	M	s.d.	Range
A-Female	*I*	219	28	166–303	61	7	43–75	277	89	103–508
	BOUGHT	225	34	166–332	63	7	44–75	209	44	125–305
	CAT	232	52	168–491	58	8	38–76	160	29	113–247
	THERE	201	27	146–285	56	9	39–73	335	48	238–454
M-Female	*I*	257	28	215–329	63	13	37–83	265	75	70–437
	BOUGHT	258	29	204–336	65	11	42–83	207	48	111–370
	CAT	270	31	222–336	60	13	41–80	160	43	71–290
	THERE	237	27	182–290	58	14	29–82	344	68	211–503

E. Sentence stress calculation

The acoustic measurements of each speaker's production of sentence stress were analyzed using two different calculations. The first calculation involved computing mean values of F_0, intensity, and duration for stressed words only. This was referred to as *average* sentence stress. Analysis of average sentence stress provided absolute values of the acoustic features for stressed words. This analysis alone might not indicate whether Mandarin speakers produced sentence stress in a manner similar to American speakers. Laver (1994) indicated that a stressed syllable/word is one that is produced more prominently than another syllable/word. Therefore, comparison of the acoustic characteristics of stressed and unstressed words would be necessary to comprehensively evaluate the contrastive (or differentiated) production of sentence stress.

The second calculation of sentence stress was designed to evaluate differentiated stress. This calculation involved comparing the acoustic characteristics of a stressed word (e.g., stressed ''*BOUGHT*'') to the same but unstressed words (e.g., the remaining unstressed ''bought'' productions) produced across the four different sentences. This was referred to as *across-sentence* stress. A ''difference'' value of an acoustic parameter (F_0, intensity, or duration) was obtained which indicated the change associated with the stressed word compared to the average of the remaining unstressed words. For example, to calculate the across-sentence $F_0(\Delta F_0)$ for the word ''bought'' across the four sentences:''*I* bought a cat there,'' ''I *BOUGHT* a cat there,'' ''I bought a *CAT* there,'' and ''I bought a cat *THERE*,'' the formula is:

$$\Delta F_0 = F_{0\ BOUGHT} - (F_{0\ \text{bought}} + F_{0\ \text{bought}} + F_{0\ \text{bought}})/3.$$

F. Measurement reliability

Ten percent of the sentence tokens (96 sentences) were randomly selected across Mandarin and American speakers (i.e., four Mandarin and four American speakers) for assessment of intra-judge and inter-judge measurement reliability. The first author remeasured the 96 sentences for intra-judge reliability assessment, while inter-judge reliability assessment was performed by another individual who was experienced in acoustic measurements. Average absolute errors for F_0, intensity, and vowel duration for intra-judge measurement were 4.10 Hz, 0.55 dB, and 1.82 ms, respectively. Pearson correlation coefficients for F_0, intensity, and vowel duration between the first and second measurements were 0.95, 0.79, and 0.97 ($p<0.01$), respectively. Average absolute errors for F_0, intensity, and vowel duration for inter-judge measurements were 1.48 Hz, 2.01 dB, and 26.88 ms, respectively. Pearson correlation coefficients for measurement of F_0, intensity, and vowel duration between the two judges were 0.79, 0.79, and 0.91 ($p<0.01$), respectively.

III. RESULTS

A. Average sentence stress

The F_0, duration, and intensity values for average sentence stress are shown in Table I for American and Mandarin

TABLE II. Mean (M), standard deviation (s.d.), and range values for F_0 (Hz), intensity (dB), and duration (ms) for each stressed word produced by the group of American (A) male and Mandarin (M) male speakers.

Group	Stressed word	F_0 (Hz)			Intensity (dB)			Duration (ms)		
		M	s.d.	Range	M	s.d.	Range	M	s.d.	Range
A-Male	*I*	140	29	87–222	61	7	46–77	299	82	163–471
	BOUGHT	144	25	106–203	67	6	54–79	208	44	118–327
	CAT	154	32	106–242	62	6	43–74	170	39	103–297
	THERE	127	26	121–203	61	6	41–74	350	66	218–488
M-Male	*I*	160	23	104–215	63	12	35–82	258	92	78–530
	BOUGHT	163	22	126–215	68	11	39–84	191	60	69–392
	CAT	167	33	107–257	65	11	40–82	140	31	71–209
	THERE	145	19	121–182	62	12	39–82	314	57	202–467

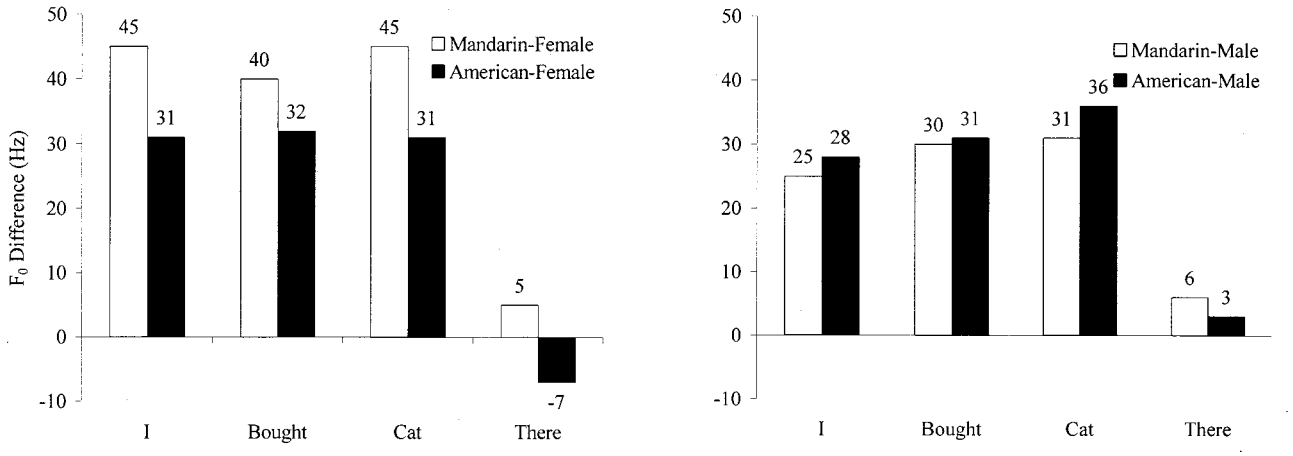


FIG. 1. Comparison of across-sentence F_0 between American and Mandarin speakers. The values represent the F_0 difference between stressed and unstressed words across sentences. Negative values indicate higher F_0 values in the nonstressed words compared to stressed words.

female speakers, and Table II for American and Mandarin males. Two-way repeated analysis of variance (ANOVA) tests were performed separately for each acoustic feature according to gender group. Native language (American English vs Mandarin) was the between groups factor and stressed words (*I, BOUGHT, CAT, THERE*) was the within groups factor.[5]

1. F_0

Results of the two-way ANOVA for the F_0 of female speakers revealed a significant main effect for language group [$F(1,479)=145.1$, $p<0.01$; $\eta^2=0.235$], with Mandarin females producing stressed words with a higher F_0 compared to American females. There was also a significant main effect for stressed word [$F(3,479)=19.8$, $p<0.01$; $\eta^2=0.112$]. Post hoc *t*-tests found *THERE* to be produced with a significantly lower F_0 compared to the remaining three stressed words ($p<0.01$). There was no significant interaction between female language groups and the F_0 of stressed words [$F(3,479)=0.14$, $p>0.01$; $\eta^2=0.001$]. Results for the male speakers revealed a significant main effect for language group [$F(1,479)=51.8$, $p<0.01$; $\eta^2=0.099$], with Mandarin males showing a higher F_0 for stressed words compared to American males. There was a significant main effect for stressed word [$F(3,479)=18.1$, $p<0.01$; $\eta^2=0.103$]. Post hoc *t*-tests found the F_0 of *THERE* to be significantly lower compared to the remaining three stressed words ($p<0.01$). No significant interaction was found between language group and stressed word [$F(3,479)=0.35$, $p>0.01$; $\eta^2=0.002$].

2. Duration

A two-way ANOVA for vowel durations produced by female speakers was nonsignificant for language group [$F(1,479)=0.06$, $p>0.01$; $\eta^2=0.000$]. A significant main effect occurred for stressed words [$F(3,479)=212.6$, $p<0.01$; $\eta^2=0.575$]. Post hoc *t*-tests found each of the four stressed words to be significantly different in duration compared to the other ($p<0.01$). There was no significant interaction between female language groups and the duration of stressed words [$F(3,479)=0.59$, $p>0.01$; $\eta^2=0.004$]. The male speakers demonstrated a significant main effect for language group [$F(1,479)=29.7$, $p<0.01$; $\eta^2=0.059$], with Mandarin males producing stressed words with shorter vowel durations compared to American males. There was also a significant effect for stressed words [$F(3,479)=196.2$, $p<0.01$; $\eta^2=0.555$] with each of the stressed words significantly different in duration compared to the other ($p<0.01$). There was no significant interaction between language group and stressed word [$F(3,479)=0.77$, $p>0.01$; $\eta^2=0.005$].

3. Intensity

Results of a two-way ANOVA for vowel intensity revealed a nonsignificant main effect for the female language groups [$F(1,479)=5.0$, $p>0.01$; $\eta^2=0.010$]. There was a significant main effect for the intensity of stressed words [$F(3,479)=11.5$, $p<0.01$; $\eta^2=0.068$], with *BOUGHT* being produced with greater intensity than *THERE* ($p<0.01$). There was no significant interaction between female language group and intensity of stressed words [$F(3,479)=0.04$, $p>0.01$; $\eta^2=0.000$]. The ANOVA results for male speakers were nonsignificant for language group [$F(1,479)=4.2$, $p>0.01$; $\eta^2=0.009$]. There was a significant main effect for stressed word [$F(3,479)=8.5$, $p<0.01$; $\eta^2=0.051$], with *BOUGHT* being produced with greater intensity than the three remaining stressed words ($p<0.01$). No significant interaction was found between language group and stressed word [$F(3,479)=0.25$, $p>0.01$; $\eta^2=0.002$].

B. Across-sentence stress

1. F_0

Results of the analysis for F_0 are shown in Fig. 1. Both American and Mandarin speakers produced stressed words with a higher F_0 compared to unstressed words. Results of a two-way ANOVA for the F_0 of female speakers revealed a significant main effect for language group [$F(1,479)=22.7$, $p<0.01$; $\eta^2=0.046$]. Mandarin females differentiated stressed from unstressed words with a higher F_0 compared to American females. There was a significant effect for

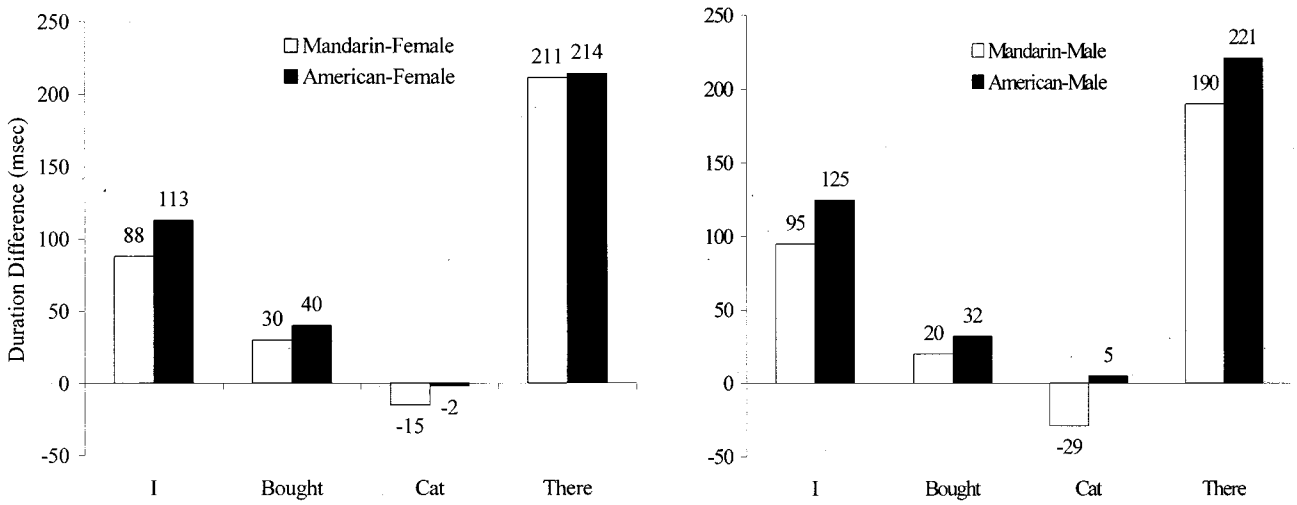


FIG. 2. Comparison of across-sentence duration between American and Mandarin speakers. The values represent the duration difference between stressed and unstressed words across sentences. Negative values indicate longer duration values in the nonstressed words compared to stressed words.

the F_0 of stressed words [$F(3,479)=15.4$, $p<0.01$; $\eta^2=0.089$]. Post hoc testing found the word *CAT* to have a higher differentiated F_0 compared to the remaining three words ($p<0.01$). There was no significant interaction between female language groups and stressed words [$F(3,479)=1.14$, $p>0.01$; $\eta^2=0.007$]. The ANOVA results for male speakers showed a nonsignificant effect for language group [$F(1,479)=0.89$, $p>0.01$; $\eta^2=0.002$], and a significant effect for the F_0 of stressed words [$F(3,479)=17.3$, $p<0.01$; $\eta^2=0.099$]. The word *CAT* had a higher differentiated F_0 compared to the remaining three words ($p<0.01$). There was no significant interaction between male language groups and stressed words [$F(3,479)=2.08$, $p>0.01$; $\eta^2=0.013$].

2. Duration

Results of the duration analysis are displayed in Fig. 2. Both American and Mandarin speakers produced stressed words with a longer duration compared to unstressed words. A two-way ANOVA for vowel durations produced by the female speakers showed a significant effect for language group [$F(1,479)=7.2$, $p<0.01$; $\eta^2=0.015$], with Mandarin females producing differentiated stress with shorter vowel durations. There was a significant effect for the duration of stressed words [$F(3,479)=123.4$, $p<0.01$; $\eta^2=0.44$]. The word *I* was produced with longer differentiated duration compared to the remaining three words ($p<0.01$). There was no significant interaction between female language groups and production of stressed words [$F(3,479)=2.53$, $p>0.01$; $\eta^2=0.016$]. A two-way ANOVA for male speakers was significant for language group [$F(1,479)=28.2$, $p<0.01$; $\eta^2=0.056$], with Mandarin males producing differentiated stress with shorter vowel durations. There was a significant effect for stressed word [$F(3,479)=126.3$, $p<0.01$; $\eta^2=0.445$], with *I* being produced with longer differentiated duration compared to the remaining three words ($p<0.01$). There was a significant interaction between male language groups and stressed word [$F(3,479)=9.35$, $p<0.01$; $\eta^2=0.056$]. Post hoc testing found Mandarin males to produce *I* and *THERE* with significantly shorter differentiated duration compared to American males ($p<0.01$).

3. Intensity

Results of the intensity analysis are shown in Fig. 3. American and Mandarin speakers produced stressed words with greater intensity compared to unstressed words. Results of the two-way ANOVA for female speakers revealed a nonsignificant effect for language group [$F(1,479)=1.9$, $p>0.01$; $\eta^2=0.004$]. There was a significant effect for stressed words [$F(3,479)=22.2$, $p<0.01$; $\eta^2=0.124$], with the differentiated intensity of *THERE* being higher than the remaining three words ($p<0.01$). There was no significant interaction between female language groups and stressed word [$F(3,479)=0.67$, $p>0.01$; $\eta^2=0.004$]. Results of the two-way ANOVA for male speakers was nonsignificant for language group [$F(1,479)=8.6$, $p>0.01$; $\eta^2=0.018$]. There was a significant effect for the intensity of stressed words [$F(3,479)=23.5$, $p<0.01$; $\eta^2=0.130$], with the differentiated intensity of *THERE* being higher than the remaining three words ($p<0.01$). There was no significant interaction between language group and stressed word [$F(3,479)=0.60$, $p>0.01$; $\eta^2=0.004$].

IV. DISCUSSION

A. Sentence stress according to F_0

It was predicted that Mandarin speakers would be able to alter F_0 in L2 English to differentiate stress from unstressed words. The present results indicate that Mandarin speakers were able to use F_0 to differentiate stressed from unstressed words in the production of American English sentences. Similar to American speakers, stressed words were consistently produced with the highest F_0. Although results obtained from the across-sentence analyses indicate Mandarin speakers' appropriate use of F_0 to signal stress, it is interesting to consider the results of the significance testing regarding the characteristics of F_0 for stressed and unstressed words. The ANOVA results obtained from the aver-

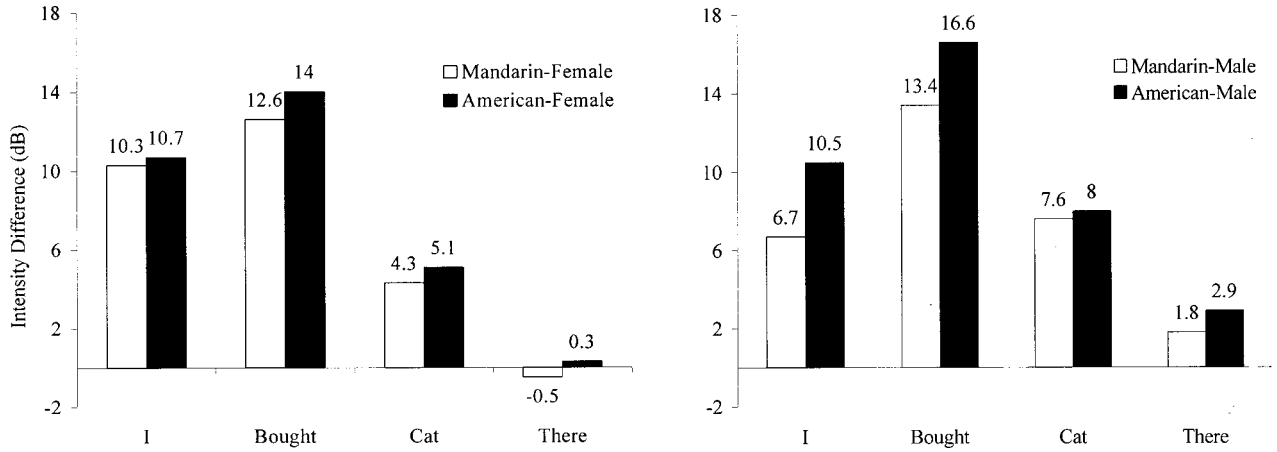


FIG. 3. Comparison of across-sentence intensity between American and Mandarin speakers. The values represent the intensity difference between stressed and unstressed words across sentences. Negative values indicate higher intensity values in the nonstressed words compared to stressed words.

age stress analysis found male and female Mandarin speakers to produce stressed words with a significantly higher F_0 than their American counterparts. A significantly higher F_0 was also found for the female Mandarin speakers in the across-sentence analysis.

Two possible explanations may account for the use of a higher F_0 for sentence stress among male and female Mandarin speakers. The first explanation is that the higher F_0 among Mandarin speakers may simply reflect basic anthropometric differences between Asian and Caucasian speakers. Yang (1996) examined the acoustic characteristics of Korean speakers compared to Caucasian speakers and found that the vocal F_0 and formant frequencies were higher in the Korean speakers. He attributed the higher acoustic values in Korean speakers to an overall smaller laryngeal framework compared to Caucasian speakers. However, Ng (1996) found no differences in vocal F_0 between Cantonese-speaking males compared to American-English speaking males. In the absence of anatomic measures of the larynx, it is unknown whether the observed differences in F_0 resulted from discrepancies in laryngeal anatomy.

Another reason for the higher F_0 among Mandarin speakers may relate to the influence of Mandarin tones on the production of American English. Eady (1982) found that the continuous speech of Mandarin contains significantly greater F_0 fluctuations at the syllable-level compared to the continuous speech of English. The greater F_0 fluctuations in Mandarin sentences were attributed to the fact that every syllable spoken in Mandarin has its own tone contour. This finding was examined in greater detail by Xu (1999) who found that the lexical tone determines the F_0 contour for a syllable, as well as the shape of F_0 contours for surrounding syllables. In addition, the location of stress within a sentence serves to vary the overall (global) F_0 of the entire sentence. Shen (1990) found that when native speakers of Mandarin produce French as L2 (a nontonal language), greater pitch fluctuations were found in the L2. Shen believed that the interplay between sentence intonation and lexical tones served to increase overall pitch level. The suggestion of an interplay between intonation and lexical tones was also espoused by Chao (1968) who suggested that the interplay was similar to small ripples riding on large waves which results in an algebraic sum of two kinds of waves. Considering the results from these various studies, it is possible that the higher average F_0 shown by Mandarin speakers in the present study was indicative of interference from L1, whereby Mandarin speakers produced words with greater pitch fluctuations compared to American speakers.

To evaluate whether the Mandarin groups produced a higher F_0 regardless of sentence stress, a series of post hoc analyses were performed. The F_0 values comprising the entire sample of unstressed vowels were summed and compared between Mandarin and American English groups. Results of the analysis found Mandarin females to produce unstressed words with a significantly higher F_0 compared to American females $[F(1,1439)=402.71,\ p<0.01;\ \eta^2=0.22]$. Results for the Mandarin males also indicated a significantly higher F_0 for unstressed words compared to American males $[F(1,1439)=253.9,\ p<0.01;\ \eta^2=0.15]$. Thus while Mandarin speakers were able to use F_0 to differentiate stressed from unstressed words, there was a tendency for these speakers to use a high F_0 across stressed and unstressed words in their production of L2 English. This finding lends further support to Shen's (1990) contention of an increased overall pitch level during L2 production because of an interplay between sentence intonation and lexical tones in L1 Mandarin. Future acoustic comparisons of Mandarin and American English speakers should consider the possible influence of anthropometric differences, as well as the interplay between sentence intonation and lexical tones to ascertain why Mandarin speakers of English use a significantly higher F_0.

B. Sentence stress according to duration

It was predicted that Mandarin speakers would show no significant differences from American speakers in their use of duration to signal American English sentence stress. Results of the analysis showed that Mandarin speakers produced stressed words with a longer vowel duration compared to unstressed words. The pattern of stress duration shown across the words was similar to that found for American

speakers (Fig. 2). However, female and male Mandarin speakers produced stressed words with significantly shorter vowel durations compared to the same productions by American speakers. This behavior was particularly apparent in the across-sentence analysis which examined differentiated vowel duration.

To evaluate whether Mandarin speakers differed from American speakers in vowel duration regardless of sentence stress, a series of post hoc analyses were performed. The vowel duration values comprising the entire sample of unstressed words were summed and compared between Mandarin and American English groups. Results of the analysis found Mandarin females to produce unstressed words with a significantly longer duration compared to American females [$F(1,1439)=27.3$, $p<0.01$; $\eta^2=0.02$], while there was no significant difference between Mandarin and American males in the vowel duration of unstressed words [$F(1,1439)=2.29$, $p>0.13$; $\eta^2=0.002$]. The behavior noted for production of unstressed words would suggest that Mandarin speakers' application of duration in L2 English is somewhat variable. These results are in agreement with a recent report by Guion *et al.* (2000) who found that temporal variability is a common feature of L2 English. Because Mandarin is considered a syllable-timed language, whereby syllable duration is relatively constant across a sentence, it is not surprising that Mandarin speakers would have difficulty regulating vowel duration in multisyllabic American English sentences (Chun, 1982).

C. Sentence stress according to intensity

The third prediction was that Mandarin speakers would show no significant differences from American English speakers in their use of intensity to signal sentence stress. Results from the across-sentence and average stress analysis found Mandarin and American speakers to produce stressed words with comparable intensity. The finding of no differences in intensity between the two language groups agrees with previous studies which indicated that intensity serves as a cue for stress in both Mandarin and American English languages (Fry, 1955; Lieberman, 1967; Lin and Yan, 1980; Shen, 1993).

There are reports of intensity being the least salient and less consistent acoustic parameter to differentiate stress from unstressed words (Clark and Yallop, 1995). Yet, this feature of sentence stress was perhaps the most comparable across Mandarin and American speakers. Additional post hoc analyses examined whether the intensity of unstressed words was also comparable across Mandarin and American speakers. The analysis revealed that Mandarin females produced unstressed words with a significantly higher intensity compared to American females [$F(1,1439)=23.2$, $p<0.01$; $\eta^2=0.02$]. Mandarin males also produced unstressed words with a significantly higher intensity than the American males [$F(1,1439)=40.2$, $p<0.01$; $\eta^2=0.16$]. Therefore, while it is likely that Mandarin and American groups used intensity to signify sentence stress, Mandarin subjects appeared to use intensity to a greater extent across both stressed and unstressed words.

V. SUMMARY AND CONCLUSIONS

This study confirmed the predictions that Mandarin speakers of English would be capable of signifying sentence stress using F_0, duration, and intensity. Although the present group of Mandarin speakers were able to signal sentence stress, the acoustic characteristics of stress were not identical to American speakers. Mandarin speakers produced stressed words with a significantly higher F_0 and shorter vowel duration compared to American speakers. Acoustic examination of unstressed words revealed additional differences between language groups. Mandarin speakers produced unstressed words with significantly higher F_0 and greater intensity than American speakers, and the vowel duration of unstressed words was either similar or longer than that of American speakers. The most reasonable explanation for these differences is an interference of L1 Mandarin in the production of L2 American English. The outcome of this study suggests that, although there are acoustic differences between Mandarin and American speakers in the production of sentence stress, there is no critical divergence between these speakers in the way they implement sentence stress in English. The work here adds to the limited body of data examining the influence of Mandarin on the accurate production of L2 American English.

[1]Putonghua (i.e., the general/standard Chinese language) has been promoted as the ''official language'' by the government throughout China since the 1950s (Chao, 1968; Cheng, 1987). It is based on the phonological and grammatical system known as Beifanghua (i.e., the northern dialect). In English, Beifanghua is also often referred to as Mandarin (Hua and Dodd, 2000).

[2]The interplay of stress and unstressed syllables/words within a sentence is thought by some researchers to reflect an isochronal (i.e., equal timing) pattern of speech rhythm (Chun, 1982; Crystal, 1969; Lehiste, 1980); however, this remains a matter of debate (cf. Cruttenden, 1986). The intent of the present study was to simply note the general timing pattern of English as contrasted to Mandarin.

[3]Besides the four ''basic'' tones in Mandarin Chinese, there is also a fifth tone, which is usually referred to as a neutral tone. According to Chao (1968), almost any morpheme normally having one of the regular four tones can be produced in a neutral tone under certain conditions. There are also a very small number of morphemes, such as suffixes, which tend to be produced in a neutral tone and do not possesses one of the four basic tones.

[4]Equal numbers of male and female subjects were selected for two reasons. First, there are limited data concerning the production of American English sentence stress by native speakers of Mandarin and several acoustic studies examining Mandarin speakers are based on subject samples of eight or less (Eady, 1982; Shen, 1990; Xu, 1997, 1999). The small sample of male and female speakers has led to the collapsing of data across genders, thereby requiring that absolute values for features such as F_0 be converted to logarithms (e.g., Xu, 1999). The relatively large number of male and female subjects sampled in the present study allowed for grouping subjects according to gender and to provide representative estimates of central tendency. Second, because a portion of the acoustic analyses involved examining absolute measures of F_0, it was deemed inappropriate to collapse the results across males and females because of the known influence of laryngeal and vocal tract anatomy on the subsequent speech product. Although it was not the intent of the present study to explore whether gender differences were apparent, we believed that reporting data according to gender served to provide additional descriptive data regarding English spoken as L2.

[5]The phonetic composition of the four words was not identical which might naturally contribute to observed acoustic differences between words. However, the ANOVA test applied to the present database was thought to provide the most comprehensive analysis within and between native Mandarin and English groups. Any significant differences which may have occurred

between groups in the production of specific word types would be identified by using word as a within-groups factor.

Berinstein, A. (**1979**). ''A cross-linguistic study on the perception and production of stress,'' UCLA Working Papers of Phonetics, **47** (UCLA Phonetics Laboratory, Los Angeles).

Bolinger, D. (**1958**). ''A theory of pitch accent in English,'' Word **14**, 109–149.

Brown, W., and McGlone, R. (**1974**). ''Aerodynamic and acoustic study of stress in sentence productions,'' J. Acoust. Soc. Am. **56**, 971–974.

Cao, J. (**1986**). ''Pu Tong Hua Qing Sheng Yin Jie Te Xin Fen Xi'' [Analysis of characteristics of unstressed syllables in Mandarin], Ying Yong Sheng Xue **5**, 23–37.

Chao, Y. (**1932**). ''A preliminary study of English intonation (with American variants and its Chinese equivalents),'' The Tsai Yuan Pei Anniversary Volume (Suppl. of the Bulletin of the Institute of History and Philosophy), Vol. 1, pp. 105–156.

Chao, Y. (**1948**). *Mandarin Primer* (Harvard University Press, Cambridge, MA).

Chao, Y. (**1968**). *A Grammar of Spoken Chinese* (University of California, Berkeley).

Cheng, C. (**1968**). ''English stress and Chinese tones in Chinese sentences'' Physica C **18**, 77–88.

Cheng, L. (**1987**). *Assessing Asian Language Performance: Guidelines for Evaluating Limited-English-Proficient Students* (Aspen, Rockville, MD).

Chun, D. (**1982**). ''A contrastive study of the suprasegmental pitch in modern German, American English and Mandarin Chinese,'' unpublished doctoral dissertation, University of California, Berkeley.

Clark, J., and Yallop, C. (**1995**). *An Introduction to Phonetics and Phonology*, 2nd ed. (Blackwell, Cambridge, MA).

Cooper, W., Eady, S., and Mueller, P. (**1985**). ''Acoustical aspects of contrastive stress in question-answer contexts,'' J. Acoust. Soc. Am. **77**, 2142–2156.

Cooper, W., and Sorensen, J. (**1981**). *Fundamental Frequency in Sentence Production* (Springer-Verlag, New York).

Cruttenden, A. (**1986**). *Intonation* (Cambridge University Press, Cambridge).

Crystal, D. (**1969**). *Prosodic Systems and Intonation in English* (Cambridge University Press, Cambridge).

Dreher, F., and Lee, P. (**1966**). ''Instrumental investigation of single and paired Mandarin tonemes'' (Douglas Advanced Research Laboratory).

Eady, S., (**1982**). ''Differences in the *F*0 patterns of speech: Tone language versus stress language,'' Lang. Speech **25**, 29–42.

Eady, S., and Cooper, W. (**1986**). ''Speech intonation and focus location in matched statements and questions,'' J. Acoust. Soc. Am. **80**, 402–415.

Eady, S., Cooper, W., Klouda, P., Mueller, D., and Lotts, P. (**1986**). ''Acoustic characteristics of sentential focus: Narrow vs broad and single vs dual focus environments,'' Lang. Speech **29**, 233–251.

Flege, J. (**1987**). ''The instrumental study of L2 speech production: Some methodological considerations,'' Lang. Learn. **37**, 285–296.

Flege, J. (**1989**). ''Chinese subjects' perception of the word-final English /t/–/d/ contrast: Performance before and after training,'' J. Acoust. Soc. Am. **86**, 1684–1697.

Flege, J. (**1995**). ''Second language speech learning: Theory, findings, and problems,'' in *Speech Perception and Linguistic Experience*, edited by W. Strange (York Press, Baltimore).

Flege, J., and Davidian, R. (**1985**). ''Transfer and developmental processes in adult foreign language speech production,'' J. Appl. Psycholing. Res. **5**, 323–347.

Fromkin, V., and Ohala, J. (**1968**). ''Laryngeal control and a model of speech production,'' UCLA Working Papers Phonet. **10**, 98–110.

Fry, D. (**1955**). ''Duration and intensity as physical correlates of linguistic stress,'' J. Acoust. Soc. Am. **27**, 765–768.

Fry, D. (**1958**). ''Experiments in the perception of stress,'' Lang. Speech. **1**, 126–152.

Gandour, J. (**1978**). ''The perception of tone,'' in *Tone: A Linguistic Survey*, edited by V. Fromkin (Academic, New York).

Guion, S., Flege, J., Liu, S., and Yeni-Komshian, G. (**2000**). ''Age of learning effects on the duration of sentences produced in a second language,'' Appl. Psycholing. **21**, 205–228.

Guo, J. (**1992**). *Zhong He Yu Yin Xue* [Comprehensive Phonetics] (Fujian Ren Min Chu Ban She, Fujian, China).

Guzma, S. (**1973**). ''*Les intérferences de l'intonation interrogative dans le français parlé par les Chiliens*,'' Thèse de 3ème cycle, Paris III.

Ho, A. (**1976**). ''Mandarin tones in relation to sentence intonation and grammatical structure,'' J. Chinese Ling. **4**, 1–13.

Howie, J. (**1976**). *Acoustical Studies of Mandarin Vowels and Tones* (Cambridge University Press, Cambridge).

Hua, Z., and Dodd, B. (**2000**). ''The phonological acquisition of Putonghua (Modern Standard Chinese),'' J. Child Lang. **27**, 3–42.

Hutchinson, S. (**1973**). ''An objective index of the English-Spanish pronunciation dimension,'' unpublished masters thesis, University of Texas, Austin, TX.

Ingram, J., and Pittam, J. (**1987**). ''Auditory and acoustic correlates of perceived accent change: Vietnamese school children acquiring Australian English,'' J. Phonetics **15**, 127–143.

Ingram, J., and Park, S. (**1997**). ''Cross-language vowel perception and production by Japanese and Korean learners of English,'' J. Phonetics **25**, 343–370.

Johnson, J., and Newport, E. (**1991**). ''Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language,'' Cognition **39**, 215–258.

Kim, D. (**1972**). ''A contrastive study of English and Korean phonology (on the basis of aural perception),'' Lang. Teach. **5**, 1–36.

Kratochvil, P. (**1962**). ''The role of stress in the syntactical analysis of modern Peking dialect,'' Arch. Orient. **30**, 147–49.

Laver, J. (**1994**). *Principles of Phonetics* (Cambridge University Press, Cambridge).

Leben, W. (**1978**). ''The representation of tone,'' in *Tone: A Linguistic Survey*, edited by V. Fromkin (Academic, New York).

Lehiste, I. (**1970**). *Suprasegmentals* (Cambridge University Press, Cambridge).

Lehiste, I. (**1980**). ''Phonetic Manifestation of Syntactic Structure in English,'' Annual Bulletin Research Institute of Logopedics and Phoniatrics, No. 14, pp. 1–27, Tokyo.

Lenneberg, E. (**1967**). *Biological Foundations of Language* (Wiley, New York).

Lieberman, P. (**1960**). ''Some acoustic correlates of word stress in American English,'' J. Acoust. Soc. Am. **32**, 451–454.

Lieberman, P. (**1967**). *Intonation, Perception and Language* (Research Monograph, 38) (M.I.T. Press, Cambridge, MA).

Lin, M., and Yan, J. (**1980**). ''Beijing Hua Qin Sheng De Shen Xue Xin Zhi [Acoustic characteristics of unstressed syllables in Beijing Dialect],'' Fang Yan **3**, 16–24.

Magnuson, J., and Akahane-Yamada, R. (**1996**). ''Acoustic correlates to the effects of talker variability on the perception of English /r/ and /l/ by Japanese listeners,'' in *Proceedings of the Fourth International Congress on Spoken Language Processing* (A. duPont Institute, France).

Ng, L. (**1996**). ''A perceptual and acoustic study of alaryngeal speech in adult Cantonese-speaking males,'' unpublished doctoral dissertation, University of Connecticut, Storrs.

Ohala, J., and Hirano, M. (**1967**). ''Studies of pitch change in speech,'' UCLA Working Papers Phonet. **7**, 80–84.

Pike, K. (**1945**). *The Intonation of American English* (University of Michigan Press, Ann Arbor, MI).

Pike, K. (**1948**). *Tone Languages* (University of Michigan Press, Ann Arbor, MI).

Pittam, J., and Ingram, J. (**1992**). ''Accuracy of perception and production of compound and phrasal stress by Vietnamese Australians,'' Appl. Psycholing. **13**, 1–12.

Potisuk, S., Gandour, J., and Harper, M. (**1996**). ''Acoustic Correlates of Stress in Thai,'' Phonetica **53**, 200–220.

Robson, B. (**1982**). ''Teaching English pronunciation to speakers of Korean,'' Focus **11**, 1–8.

Scuffil, M. (**1982**). *Experiments in Comparative Intonation—A Case Study of English and German* (Max Niemeyer, Tübingen).

Shen, X. (**1990**). ''Ability of learning the prosody of an intonational language by speakers of a tonal language: Chinese speakers learning French prosody,'' Int. Rev. Appl. Ling. Lang. Teach. **28**, 119–134.

Shen, X. (**1993**). ''Relative duration as a perceptual cue to stress in Mandarin,'' Lang. Speech **36**, 415–433.

Shih, C. (**1986**). *The Phonetics of the Chinese Tonal System: Technical Memorandum* (AT&T, Bell Laboratories).

Tarone, E. (**1980**). ''Some influences on the syllable structure of interlanguage phonology,'' Int. Rev. Appl. Ling. Lang. Teach. **18**, 139–152.

Tseng, C. (**1981**). ''An acoustic phonetic study on tones in Mandarin Chinese,'' unpublished doctoral dissertation, Brown University, Providence, RI.

van Santen, J., and Shih, C. (**2000**). ''Suprasegmental and segmental timing models in Mandarin Chinese and American English,'' J. Acoust. Soc. Am. **107**, 1012–1026.

Wijk, A. (**1966**). *Rules of Pronunciation for the English Language: An Account of the Relationship Between English Spelling and Pronunciation* (Oxford University Press, London).

Xu, Y. (**1997**). ''Contextual tonal variations in Mandarin,'' J. Phonetics **25**, 61–83.

Xu, Y. (**1999**). ''Effects of tone and focus on the formation and alignment of F_0 contours,'' J. Phonetics **27**, 55–105.

Yang, B. (**1996**). ''A comparative study of American English and Korean vowels produced by male and female speakers,'' J. Phonetics **24**, 245–261.

Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners

Winifred Strange[a)]
Speech & Hearing Sciences, The City University of New York Graduate School and University Center, New York, New York 10016

Reiko Akahane-Yamada and Rieko Kubo
Information Sciences Division, Advanced Telecommunications Research Institute International, Kyoto 619-0288, Japan

Sonja A. Trent and Kanae Nishi
Department of Psychology, University of South Florida, Tampa, Florida 33620

(Received 13 April 1999; revised 25 September 2000; accepted 1 December 2000)

This study investigated the extent to which adult Japanese listeners' *perceived phonetic similarity* of American English (AE) and Japanese (J) vowels varied with consonantal context. Four AE speakers produced multiple instances of the 11 AE vowels in six syllabic contexts /b-b, b-p, d-d, d-t, g-g, g-k/ embedded in a short carrier sentence. Twenty-four native speakers of Japanese were asked to categorize each vowel utterance as most similar to one of 18 Japanese categories [five one-mora vowels, five two-mora vowels, plus/ei, ou/ and one-mora and two-mora vowels in palatalized consonant CV syllables, $C^{j}a(a)$, $C^{j}u(u)$, $C^{j}o(o)$]. They then rated the ''category goodness'' of the AE vowel to the selected Japanese category on a seven-point scale. None of the 11 AE vowels was assimilated unanimously to a single J response category in all context/speaker conditions; consistency in selecting a single response category ranged from 77% for /eɪ/ to only 32% for /æ/. Median ratings of category goodness for modal response categories were somewhat restricted overall, ranging from 5 to 3. Results indicated that temporal assimilation patterns (judged similarity to one-mora versus two-mora Japanese categories) differed as a function of the voicing of the final consonant, especially for the AE vowels, /i, u, ɪ, ɛ, ʌ, ʊ/. Patterns of spectral assimilation (judged similarity to the five J vowel qualities) of /ɪ, ɛ, æ, ʌ/ also varied systematically with consonantal context and speakers. On the basis of these results, it was predicted that relative difficulty in the identification and discrimination of AE vowels by Japanese speakers would vary significantly as a function of the contexts in which they were produced and presented. [DOI: 10.1121/1.1353594]

PACS numbers: 43.71.An, 43.71.Es, 43.71.Hw [JMH]

I. INTRODUCTION

Early research in cross-language speech perception demonstrated that adult listeners had difficulty perceptually differentiating phonetic contrasts that were not distinctive in their native language (see Strange, 1995, for review). These learned, language-specific patterns of perception have been cited as a major contribution to adult language learners' difficulty in acquiring a second-language (L2) phonology. However, research over the last several years has shown that there is considerable variability in second-language learners' perceptual difficulties with non-native contrasts, and relative perceptual difficulty is often not predictable from contrastive analyses based on comparisons of native and non-native phoneme inventories or phonological features (Kohler, 1981). Some distinctions between non-native segments are rather easily perceived with no prior experience [e.g., Farsi velar versus uvular stops by English listeners (Polka, 1992); voicing and place contrasts in Zulu clicks by English listeners (Best *et al.*, 1988)]. In other cases, contrasts which are ''the same'' as native contrasts in terms of (abstract) phonological features are nevertheless difficult for second-language learners to differentiate perceptually [e.g., French /i-e/ by English L2 learners of French (Gottfried, 1984)].

Two current theories of second-language (L2) phonological acquisition hypothesize that difficulties in learning both to produce and to perceive non-native phonetic segments are predictable in large part on the basis of the perceived *phonetic* similarities and dissimilarities between native language (L1) segments and the to-be-learned segments (Best, 1994, 1995; Flege, 1992, 1995). In both of these theories, phonetic similarity is defined in terms of the acoustic and/or articulatory characteristics of the linguistically relevant speech sounds. That is, both theorists argue that a context-free phonemic level of analysis is too abstract to capture the details needed to make precise predictions about cross-language perceptual patterns. However, to date, there are limited data describing the phonetic details of segments in cross-language comparisons of phoneme inventories and little is known about how cross-language differences in the phonetic realization of segments affect their perception by non-native listeners. Most studies which examine cross-language perception utilize stimulus materials in which the

a)Electronic mail: strangepin@aol.com

target segments are presented in very limited (often single) phonetic and phonotactic contexts, and most often in citation-form monosyllables or even in isolation. Since the phonetic/phonotactic context and the larger prosodic context significantly affect the phonetic realization of phonological segments, evidence from these studies of cross-language perceptual patterns cannot be generalized beyond the contexts in which the target segments were examined.

Support for the necessity of considering phonetic/phonotactic variables in cross-language perception comes from studies showing that Japanese L2 learners of American English (AE) have significantly more difficulty learning to perceive the /ɹ/-/l/ contrast in prevocalic contexts (especially initial consonant clusters) than in postvocalic contexts (e.g., Mochizuki, 1981; Lively *et al.*, 1993; Yamada and Tohkura, 1992a, b) and that vowel context also seems to affect perceptual difficulty (Sheldon and Strange, 1982). Other studies have documented context-specific patterns of cross-language perception of voicing contrasts between stop consonants in initial position (cued by voice onset time) and in final position (cued by vowel duration, closure ''murmur'' and release bursts) (cf. Flege and Hillenbrand, 1987; Flege and Eefting, 1987), although seldom in the same study. Morosan and Jamieson (1989) also reported that French speakers trained to perceive English /ð-θ/ in syllable-initial position failed to generalize to new syllable positions. More recently, Schmidt (1996) reported effects of vowel context on cross-language perceptual similarity judgments of Korean and American English consonants by Korean L2 learners of English.

Previous cross-language studies of vowels have shown that L2 learners may have persistent difficulties learning to perceptually differentiate some non-native vowel contrasts (cf. Gottfried, 1984; Flege, 1995). Other studies have reported relatively accurate perception of non-native vowels even by listeners with no previous experience with the language (Polka, 1995; Best *et al.*, 1996). Further studies have demonstrated that the acoustic cues upon which L2 learners base their perceptual judgments may vary from those of native speakers (cf. Bohn, 1995; Gottfried and Beddor, 1988). These and other studies have usually employed stimulus materials in which vowels are produced (or synthesized) in isolation (e.g., Kuhl *et al.*, 1992) or in citation-form syllables with a single consonantal context (e.g., Flege, 1995); thus, little is known about contextual influences on perception of non-native vowels. The Gottfried (1984) study of L2 learners of French is of particular interest here because he tested perception of the same vowel contrasts when they were produced (and presented) as isolated utterances (#V#) and in citation-form /tVt/ syllables, and found that some vowels were significantly more difficult to discriminate in consonantal context than in isolation (e.g., 86% correct for /y-ø/, 65% correct for /tyt-tøt/; chance=50%).

According to the perceptual assimilation model (PAM) (Best, 1994, 1995), relative difficulty in perception of non-native contrasts is predictable from listeners' patterns of perceptual assimilation of non-native vowels to native vowel categories. Non-native vowels may be assimilated to native vowel categories on a continuum from ''good'' to ''poor'' instances of a native category, or perhaps as ''uncategorizable'' segments. If two non-native segments are assimilated to a single native vowel category as equally ''good'' instances, then they will be very difficult to discriminate (single category pattern). If one non-native vowel is judged to be a ''good'' instance of a native category and a contrasting non-native vowel is considered a ''poor'' instance of that same category or ''uncategorizable,'' then the two segments will be easier to perceptually differentiate (category goodness pattern). If two non-native vowels are perceptually assimilated to different native categories, they will be discriminated most accurately (two category pattern).

In order to predict the relative perceptual difficulty of non-native vowels, then, it is important to understand the patterns of perceptual assimilation of those vowels to native language categories and how those assimilation patterns differ as a function of the context in which the non-native phones occur. To this end, we have conducted experiments in which listeners are asked directly to categorize multiple instances of each L2 vowel as being ''most similar'' to one of their L1 vowel classes and to judge the ''category goodness'' of the L2 vowel with respect to that L1 category. The present study reports results of an experiment on the perceptual assimilation of AE vowels by native speakers of Japanese when the vowels were produced and presented in varying CVC syllables imbedded in a fixed sentence context.

The Japanese vowel inventory consists of five vowel ''qualities'' (which we refer to as spectral categories) /i, e, a, o, u/ which have phonologically distinctive short (one-mora) and long (two-mora) cognates. Both one- and two-mora vowels are produced as monophthongs, and long vowels are often described as ''double'' vowels, implying that the duration ratio of two-mora to one-mora vowels is approximately 2:1. In Standard (Kanto) dialect, the high back vowel /u/ is phonetically realized as unrounded [ɯ], while in western dialects (Kansai) it is more rounded. The American English (AE) vowel inventory consists of 11 spectrally distinct vowels /i, ɪ, e, ɛ, æ, ɑ, ɔ, ʌ, o, ʊ, u/, plus the rhotic vowel /ɝ/ and the ''true'' diphthongs. In stressed syllables, /e, o/ are diphthongized [e͡ɪ, o͡ʊ] and many of the other so-called monophthongs display considerable ''vowel-intrinsic spectral change'' (Andruski and Nearey, 1992; Nearey and Assmann, 1986; Hillenbrand *et al.*, 1995). In many dialects of AE, the /ɑ-ɔ/ distinction is neutralized to a single vowel [ɒ]. While vowel length is considered phonologically redundant in English, phonetically AE vowels can be divided into three categories: four intrinsically short vowels [ɪ, ɛ, ʌ, ʊ], two intermediate vowels [iː, uː], and five long vowels [e͡ɪ, æː, ɑː, ɔː, o͡ʊ] (Peterson and Lehiste, 1960).

Both temporal and spectral properties of AE vowels vary systematically with contextual variables. Vocalic durations differ as a function of phonetic context (e.g., voicing of the following consonant), phonotactic context (e.g., structure and number of syllables), prosodic structure (lexical stress, sentence focus, phrase boundaries), and the rate of speech (Crystal and House, 1988a, 1988b; Klatt, 1976; Fourakis, 1991). The spectral characteristics of vowels ($F1/F2/F3$ formant values or ratios) have been shown to differ as a function of consonantal context (Stevens and House, 1963), dialect (Wolfram, 1991), speaking rate and style (Moon and

Lindblom, 1994; Huang, 1991), and individual speaker characteristics such as gender, age, and physiognomy (Peterson and Barney, 1952; Hillenbrand *et al.*, 1995). While some of these causes of acoustic variability are due to physiological constraints of the human vocal system, and are thus universal across languages, other sources of variability are part of the linguistic codes of the language.

In the case of language-specific patterns of variability, perceivers must learn to compensate for the variability by forming equivalence classes appropriate to the language being acquired. Developmental research has demonstrated that language-specific patterns of equivalence classification of vowels are evident very early in L1 acquisition (Kuhl *et al.*, 1992; but see Polka and Bohn, 1996). For the L2 learner, then, part of learning to perceive non-native vowels accurately involves establishing new rules for equivalence classification (Flege, 1992, 1995) by which discriminable differences between instances of L2 vowels are correctly partitioned into within-L2-category variations (i.e., allophonic, stylistic, or free variations) versus across-L2-category variations (distinctive or phonemic variations). Before such new equivalence classes are acquired, contextual variations in the realization of an L2 vowel may very well lead to its being perceptually assimilated to different L1 categories (and with different category goodness) when heard in different contexts.

In a previous study (Strange *et al.*, 1998) we reported that Japanese listeners' perceptual assimilation of AE vowels to native categories (and their judged category goodness) differed when vowels were produced and presented in citation-form /hVba/ bisyllables versus /hVb/ syllables imbedded in the sentence, ''I say the /hVb/ on the tape.'' In the sentence condition, the intermediate and long AE vowels were heard as similar to two-mora Japanese vowels much more consistently than for the citation-form bisyllables. In both citation and sentence contexts, the short AE vowels were assimilated to one-mora categories; thus, assimilation patterns revealed better temporal differentiation of long versus short AE vowels in the sentence condition. This was true, even though the relative durations of intermediate/long versus short AE vowels (mean ratio=1.3) in citation-form and sentence utterances did not differ significantly, and were smaller than typically found for Japanese two-mora versus one-mora cognates. Assimilation of the 11 AE vowels to the five Japanese spectral categories (disregarding length) also varied with context. Four of the eleven vowels [ɪ, æː, ɔː, o͡ʊ], were actually assimilated most often to *different* Japanese spectral categories in citation and sentence conditions. Acoustical analysis of the AE vowel stimuli suggested that some, but not all, variations in spectral assimilation patterns could be accounted for by differences in speakers' productions of the vowels in the different contexts.

In summary, both temporal and spectral assimilation patterns of AE vowels by Japanese listeners were found to vary with prosodic context. In that study, the immediate phonetic context was chosen in order to minimize the coarticulatory influence of preceding and following consonants on the formant structure of the vowels (Stevens and House, 1963), and to preclude *variation* in spectral structure and vocalic duration within each condition that was not intrinsic to the vowel. In the present study, AE vowels were produced in CVC syllables in the same carrier sentence as used before, while the preceding and following stop consonant context was varied. The voicing of the final consonant was varied to investigate how *contextual* variation in vocalic duration affected temporal assimilation patterns. Place-of-articulation of preceding and following consonants was varied to examine the effects of coarticulatory variations in formant structure on patterns of spectral assimilation to Japanese vowel categories. Thus, systematic variation in perceptual assimilation of AE vowels by Japanese listeners as a function of immediate phonetic context was examined, while holding the phonotactic and prosodic context constant.

Although there are few systematic studies of the phonetic variability of Japanese vowels as a function of contextual variables, phonetic descriptions state that (1) the high-back vowel [ɯ] is fronted in coronal consonantal contexts (Homma, 1992; Shibatani, 1990), and (2) the high vowel /i/ and /u/ are ''devoiced'' when preceded and followed by voiceless consonants or following a voiceless consonant in word final position (Sugito, 1996). Although the nature of the devoicing process is controversial, acoustically this leads to a change in both source characteristics and temporal characteristics of these vowels (Beckman, 1982). Keating and Huffman (1984) reported acoustical data from citation-form utterances and continuous (read) speech of seven speakers, showing considerable variation in the formant structure of all five one-mora vowels, with the overall spread in spectral values greatest for /a/ and /ɯ/. In continuous speech, $F1/F2$ values for the five one-mora vowels overlapped considerably even within an individual speaker's utterances, suggesting significant coarticulatory effects on formant structure. Fitzgerald (1996) reported vowel centralization and systematic coarticulatory effects of preceding and following stop consonants for one-mora vowels produced in CVC syllables in a short carrier sentence. Temporal patterning of Japanese vowels also varies with speaking style and context. In continuous speech contexts, the duration ratio of two-mora to one-mora vowels is reduced, but two-mora vowels are still over 50% longer than one-mora vowels (Campbell and Sagisaka, 1991). Vowels before voiced consonants are not systematically longer in duration than vowels before voiceless consonants (Fitzgerald, 1996) and, in general, allophonic variation in vowel duration is said to be more conditioned by preceding (tautomoraic) consonants than by the following consonants (Homma, 1981).

In the present study, four speakers produced the 11 AE vowels in six contexts: /b-b, b-p, d-d, d-t, g-g, g-k/ in the sentence, ''I say the /CVC/ on the tape,'' with instructions to produce the sentences ''as if speaking to a native speaker,'' with no pauses or special emphasis on the target syllable. Thus, their rate and style of speech was thought to be somewhere between careful and casual speech styles and the temporal and spectral characteristics of the vowels were considered to be closer to those found in continuous speech utterances than are citation-form utterances. The specific questions asked were as follows:

A. Temporal assimilation patterns

(a) Are intermediate and long AE vowels consistently assimilated to two-mora J categories and the short vowels to one-mora J categories (as in the previous study) or does contextual variation in vocalic duration reduce consistency in temporal assimilation of AE vowels?

(b) Do temporal assimilation patterns vary systematically with the voicing of the following consonant? Because Japanese vowels do not vary systematically with the final consonant voicing, it was predicted that AE vowels would be assimilated more to two-mora Japanese categories when followed by voiced consonants than by voiceless consonants.

(c) To what extent does vocalic duration alone predict perceptual assimilation to two-mora as opposed to one-mora Japanese categories?

B. Spectral assimilation patterns

(a) Do patterns of perceptual assimilation of the 11 AE vowels to the five J spectral categories differ systematically with the place-of-articulation (and voicing) of the preceding and following consonants? Specifically, are vowels that have no close counterpart in Japanese assimilated to *different* Japanese vowel categories as a function of consonantal context?

(b) If spectral assimilation patterns vary systematically with consonantal context, to what extent do differences in formant structure predict those variations?

II. METHOD

A. Stimulus materials

Four male monolingual speakers of AE (25–36 years old) produced the corpus. [These were the same speakers as in Strange *et al.* (1998); stimulus materials for the present study and for that study were recorded at the same time.] None of the men spoke with a strong regional accent and all four maintained the [ɑ-ɔ] contrast in their spontaneous speech. There were, however, some noticeable variations in the phonetic realization of these vowels and of /æ, ʊ/. Each speaker produced the 11 AE vowels in six CVC contexts /b-b, b-p, d-d, d-t, g-g, g-k/ in blocks of 12 sentences each, within which the vowels were randomly sequenced, but the consonant context was constant. (The 12th sentence was a repetition of the first, and was not used in the experiment in order to control for any changes in intonation associated with the last utterance of the list.) Four such blocks of 11 vowels in each consonantal context were recorded.

The first author monitored the recording sessions of all four speakers and asked the speaker to repeat any utterances which were obviously misproduced, or which contained inconsistencies in intonation, rate, or voice quality. The speakers were familiar with phonetic transcription, but were not trained speakers. Speaker JM was recorded in an anechoic chamber at ATR Human Information Processing Research Laboratories, using a DAT (SONY PCM 2500, B) recorder and condenser microphone (SONY ECM-77). The remaining three speakers were recorded in an IAC chamber at USF Speech Perception Laboratory using a DAT (SONY TDC-D10) recorder and dynamic microphone (Panasonic WM-1325).

The stimuli were digitally transferred to computer files and downsampled to 22.05 kHz for storage, analysis, and presentation to subjects. Three utterances of each vowel in each context were chosen by the first author as the ''best'' instances of the vowel. In most cases, these were taken from blocks 2–4 of the original recordings. However, in a few cases, the first repetition was considered better and was substituted. In the judgment of the first author, all productions were good instances of the intended vowels except in the case of JM's productions of [e͡ɪ] in the /g-g/ context. In this context, the contrast between [e͡ɪ-ɛ] was reduced so that one of the three ''best'' [e͡ɪ] tokens was still heard by the first author as [ɛ]-like. The final stimulus corpus consisted of 198 utterances for each speaker: 11 vowels ×6 contexts ×3 repetitions. Separate listening tests were constructed for each speaker's utterances. Within each speaker condition, the consonantal context and the vowels varied randomly from trial to trial within the block of 198 trials. Each listener completed three such blocks of 198 items from a single speaker, for a total of nine judgments of each vowel in each consonantal context.[1]

A group of 16 native AE speakers who were undergraduates at the University of South Florida were presented the listening tests for identification. Pooling over all vowels, contexts, and speakers, the vowels were identified correctly 94% of the time (where ''correct'' is defined as the vowel *intended* by the speaker). Seven vowels /i, ɪ, ɛ, æ, ʌ, o͡ʊ, u/ were identified with ⩾98% accuracy over all speakers and contexts. The vowels /ɑ/ and /ɔ/ were sometimes confused with each other, reflecting the dialect patterns of some of the listeners. Overall correct identification rates for these vowels were 81% and 83%, respectively (98% and 100% when /ɑ-ɔ/ confusions were not counted as errors). JM's productions of /e/ in /g-g/ context were often misidentified as /ɛ/; correct identification of /e/ in this context averaged across speakers was 88%. In all other contexts, correct identification was ⩾90% for all four speakers. Finally, the vowel /ʊ/ was sometimes misidentified as /ʌ/, especially in labial contexts; the overall correct identification rate was 85%.

B. Acoustical analysis procedures

Digitized stimuli were analyzed on a Power Macintosh computer using Soundscope/16 speech analysis software (GW Instruments). Vocalic duration was determined by inspection of time-aligned waveform and broadband spectrographic displays (300 Hz, 1024 FFT points, 6 dB preemphasis). Vocalic duration was defined as the interval between the onset of voicing (the first full glottal pulse following the plosive burst) and the beginning of consonant closure (indicated by a rapid decrease in waveform amplitude and the cessation of energy in upper formants). Frequencies of the first three formants were measured using LPC analysis (autocorrelation—28 coefficients). LPC spectral envelopes were superimposed on narrow-band power spectra (FFT) computed at three temporal locations centered 25%, 50%, and 75% through the vocalic duration of the syllable. Win-

dow width for the 50% point was 25 ms, while 10-ms windows were used for 25% and 75% measurements, since the latter locations included portions of the syllable in which formant transitions associated with the consonants were present, especially for the short vowels. Numerical values given by the LPC analysis were used unless clear errors (spurious formants or missed formants) were evident. To determine this, LPC formant tracks were superimposed on the broadband spectrographic display of the target syllables to assess formant continuity, and LPC spectral envelopes and FFT power spectra were compared. In those cases where LPC values were judged to be incorrect, formant frequencies were determined by manual measurement of smoothed FFT power spectra. All measurements were made by two experimenters, working separately, and discrepancies (8 ms for duration, 50 Hz for $F1$, 150 Hz for $F2$, 250 Hz for $F3$) were resolved through additional analysis by the most experienced research assistant or the first author.

C. Listeners and procedures

Native speakers of Japanese (18 to 24 years old) residing in the Kansai area volunteered to serve as subjects and were paid for their participation. Six listeners were randomly assigned to each speaker for a total of 24 subjects (13 males and 11 females) in all. All were college students at nearby universities and had received the ''standard'' English language training, which consists of six years of instruction in junior and senior high school, and some classes in college. English instruction emphasized reading and writing skills, with little or no experience listening to native English speakers. None of the subjects had lived or traveled extensively in an English-speaking country. A language-background questionnaire revealed that 19 listeners were from the Kansai region, while the remaining five listeners were from other regions of Japan (Tokyo, Nagoya, Kyushu, and Hokuriku).

All subjects were tested individually in sound-attenuated chambers at ATR Human Information Processing Research Laboratories (Kyoto, Japan) by Japanese experimenters. Perceptual assimilation and goodness ratings tasks were conducted using a NeXT cube workstation for stimulus presentation and response acquisition. Stimuli were converted to analog signals and low-pass filtered using the NeXT DSP board, and presented binaurally through earphones (Stax SR Signature Professional) at a preset comfortable listening level.

A trial consisted of two presentations of each utterance in succession. After the first presentation, the subject categorized the target syllable as ''most similar'' to 1 of 18 J response alternatives, by selecting 1 of 18 *katakana* characters displayed on the computer screen. The 18 characters represented CV syllables containing the five one-mora vowels, [a, i, u, e, o], the five two-mora vowels [aa, ii, uu, ee, oo],[2] the palatalized consonant+vowel combinations [^{j}a, ^{j}u, ^{j}o] (1-mora syllables), the palatalized consonant+two-mora vowels [^{j}aa, ^{j}uu, ^{j}oo], and the two-mora vowel combinations [ou, ei].[3] The *katakana* symbols changed from trial to trial so that they represented the appropriate initial consonant; thus the listeners knew the identity of the initial consonant before responding, but not whether the final consonant (with the same place of articulation as the initial consonant) was voiced or voiceless. Following the subject's categorization response, the same utterance was repeated and the subject rated the vowel's category goodness on a scale from 1 to 7 (7=best fit). The endpoints were labeled ''Japanese-like'' (7) and ''not Japanese-like'' (1). Subjects were permitted to repeat an utterance or change their categorization response after the second presentation, but were discouraged from doing so. A new trial began after the rating response was completed; thus, all testing was subject paced.

D. Perceptual data analysis

Categorization responses were tallied for each speaker's vowels in each consonantal context. Frequencies of each of the 18 responses were summed over listeners and converted to percentages of total opportunities. Goodness ratings assigned to each response category were tallied and the median response, summed over listeners, was computed. For each vowel, overall response patterns (frequencies of selection of each response category and overall median goodness ratings) summed over all four listener/speaker groups were also computed and are presented as summary descriptive statistics in tables shown later in this work. These overall data thus pool responses both within and across listener/speaker groups to obtain reliable estimates of how the ''average'' Japanese listener perceptually assimilates the AE vowels of the ''average'' (male) AE speaker.

To assess the effects of consonantal context, speaker differences, and their interaction on perceptual assimilation patterns, percentages of selection of the modal (most frequent) response category for each vowel, summed over listeners, were analyzed using the proportion of reduction in error (PRE) procedure (Reynolds, 1984; Wickens, 1989). This is a nonparametric procedure for use with categorical data where the effect-size measures employed with continuous data are not appropriate. If there is *systematic* variation in the assimilation pattern across different contexts and speakers, the overall modal response category will not be the best predictor of assimilation patterns across all conditions and speakers. The PRE analysis quantifies these systematic changes across conditions in the native category to which the nonnative vowel is most often assimilated. For each AE vowel, the proportion of obtained responses not predicted correctly by the overall modal response—the base rate of error—was computed. To establish the effect of context on modal assimilation patterns, the modal response for each consonantal condition (summed over speakers) was used as the predicted response; then the number of incorrectly predicted responses was determined, and the reduction in error was converted to a proportion of the base rate of error. In the same way, the speaker effect was computed by using the modal response for each vowel in each speaker condition (summed over consonantal contexts) as the predicted response; the reduction in errors, expressed as a proportion of the base rate of error, was computed. Finally, the interaction between consonantal context and speakers was assessed by using the modal response categories for each speaker/context combination as the predicted responses for the AE vowel. To the extent that the proportion of errors is *further* reduced over the reduc-

TABLE I. Overall perceptual assimilation patterns, pooling responses over six consonantal contexts and four speakers. See text for details.

AE vowel	Modal assimilation response			Second most frequent response		
	Japanese category	Consistency (%)	Goodness (median)	Japanese category	Consistency (%)	Goodness (median)
iː	ii	71	5	i	26	6
e͡ɪ	ei	77	5	ee	13	5
ɑː	aa	60	5	a	21	3
o͡ʊ	ou	57	5	oo	17	5
uː	uu	73	5	u	26	5
æː	aa	32	3	a	17	1
ɔː	oo	46	4	ou	26	4
ɪ	i	45	5	e	31	5
ɛ	e	37	5	a	22	3
ʌ	a	44	5	o	25	4
ʊ	u	50	4	uu	32	4

tions attributable to contexts and speakers separately, it can be concluded that the variation in perceptual assimilation pattern across contexts was different for different speakers. As in the previous study, a PRE of at least 0.10 was considered to reflect systematic variation in perceptual assimilation patterns. [See Strange *et al.* (1998) for further details of the logic behind these analyses and reliability estimates.] That is, when context/speaker specific response categories used to predict assimilation responses led to a 10% improvement in ''fit'' to obtained response patterns over the use of the context/speaker independent predicted response, we concluded that there were systematic context/speaker effects on perceptual assimilation patterns.

III. RESULTS

In this section, we first present descriptive data on assimilation patterns, pooled over speakers and contexts, followed by the PRE analysis of context and speaker effects on modal response patterns. Following this overall analysis, the effect of consonantal context on temporal assimilation patterns is examined in more detail. Finally, spectral assimilation patterns and how they varied with contexts and speakers are presented.

A. Overall perceptual assimilation patterns

After pooling the data across all consonantal contexts and speaker/listener groups, the overall modal J response category and the second most frequent J response category were determined for each AE vowel, as shown in Table I (columns 2 and 5). Frequencies of selection of these two responses were converted to percentages of total opportunities and are presented in columns 3 and 6 of the table. These values can be considered measures of the overall consistency with which the listeners assimilated each AE vowel to particular J categories, independent of context. In addition, the median ratings assigned to these responses are given in columns 4 and 7; these values indicate the perceived category goodness of the AE vowel as an instance of the selected J category. AE vowels are grouped such that the five (intermediate and long) vowels which are phonetically most ''similar'' to J vowels are presented in the top rows, the two ''new'' long AE vowels (those with no clear phonetic counterpart in the J inventory) are presented next, and the four short vowels (also considered ''new'') are presented in the bottom rows.

As the modal response percentages indicate, *no* AE vowel was assimilated to a *single* J category in all contexts for all speakers with extremely high consistency (values ranged from 32% to 77%). As expected, the five similar AE vowels [iː, e͡ɪ, ɑː, o͡ʊ, uː] were assimilated most often to their analogous two-mora J categories [ii, ei, aa, ou, uu], with relatively better consistency (57% to 77%) than new vowels, and relatively high goodness ratings (median=5). Note that the modal responses for the diphthongized [e͡ɪ, o͡ʊ] were the two-mora vowel combinations [ei, ou]. For these five AE vowels, the second most frequent responses were the same spectral J categories as the modal responses, but were either the one-mora cognates (for [iː, ɑː, uː]) or the two-mora monophthongal vowels [ee, oo].[3] With the exception of J [a] for AE [ɑː], the judged category goodness was also relatively high for the second most frequent responses. Collapsing over these variations in perceived temporal and diphthongal patterns, these five AE vowels were assimilated to the five J spectral categories with consistencies ranging from 78% to 99% and median ratings of 5. In general, then, these vowels were perceived as relatively good to excellent instances of native categories.

In contrast, the two new long vowels and the four new short vowels were assimilated less consistently (32% to 50%) and had lower median goodness ratings for some vowels (3–5). That is, none of these vowels was categorized more than half the time (over speakers and contexts) as most similar to a single J vowel category. When the second most frequent responses are included, it is clear that [ɔː] and [ʊ] were more consistently assimilated to mid-back and high-back spectral categories, respectively, disregarding variations in perception of temporal and diphthongal parameters. In terms of vowel quality then, these vowels were also considered relatively good instances of Japanese vowels. For the remaining three short vowels and [æː], there were clear inconsistencies in spectral assimilation patterns. For [æː], two- and one-mora J [a(a)] responses, combined, accounted for

TABLE II. Results of proportion of reduction of error (PRE) analysis performed on modal responses for each AE vowel. Meaningful effects (>0.10) are indicated in boldface.

AE vowel	Base rate of error	Content effects	Speaker effects	Context × speaker interaction
iː	28.63	0.000	0.000	**0.119**
ɛɹ	23.07	0.000	0.000	0.047
ɑː	40.20	0.000	0.000	0.006
o͡ʊ	42.67	0.000	0.000	0.040
uː	27.77	0.000	0.000	0.091
æː	67.98	0.091	0.023	**0.135**
ɔː	54.39	0.000	0.000	0.057
ɪ	54.63	**0.133**	0.052	**0.234**
ɛ	63.04	0.042	**0.114**	**0.229**
ʌ	56.48	0.000	**0.112**	0.145
ʊ	50.23	0.063	0.054	**0.204**

just less than half of the total responses (49%), and were assigned very low goodness ratings (1–3). Other responses were distributed across palatalized J categories [ʲaa, ʲa] (23%; median ratings=2 to 3) and mid-front J categories [ei, ee, e] (26%; median ratings 2 to 3). The overall response patterns for [ɪ, ɛ, ʌ] placed them as perceptually between two adjacent J categories. Surprisingly, when pooling across contexts and conditions, the judged category goodness to the modal category (median ratings=5) were not any lower for these three short vowels than for the five similar intermediate and long vowels.

To ascertain how much of the variability in modal response consistency was systematically related to context and speaker differences, PRE analyses were completed for each of the 11 AE vowels. Table II displays the results; again, AE vowels are arranged such that the five similar vowels are grouped above, while the new long and short vowels are listed below. The second column lists the base rates of error; the next three columns give the PRE values for context, speaker, and context/speaker analyses, respectively. PRE values exceeding the criterial value of 0.10 are printed in boldface.

Modal responses for the five similar vowels did not vary systematically with consonantal context, speaker, or particular context/speaker combinations, except for the high front vowel [iː]. For this vowel, assimilation to two-mora versus one-mora high front J categories did change across contexts for two of the four speakers. Variations in temporal assimilation of [uː] approached the *a priori* value set for a meaningful variation, again for the same two speaker/context conditions as for [iː] (speaker×context PRE=0.091). Thus, for these five similar vowels, the PRE analysis showed that assimilation patterns were quite stable across contexts and speakers, with only weak evidence of systematic variability in temporal assimilation (discussed further in Sec. III B).

Variability in modal responses for five of the six remaining AE vowels was attributable, in part, to particular speakers, contexts, or their interaction. For [ʊ], the variability in perceptual assimilation to two-mora versus one-mora J high back categories was due to an interaction between speakers and contexts. (These results are also described in more detail in Sec. VII B.) The remaining three short vowels and [æː] produced very complex spectral assimilation patterns (discussed further in Sec. III C). For [æː], the context/speaker PRE of 0.135 was not 0.10 greater than the main context effect (PRE=0.091), but did exceed the 0.10 criterion. Thus, we can interpret this result as primarily due to systematic variation with context (see Sec. III C).

In summary, then, the PRE analyses confirmed that inconsistant assimilation patterns for 5 of the 11 AE vowels could be attributed to *systematic differences* in perceived similarity as a function of consonant context for one or more speakers' utterances. In order to examine these variations in more detail, further analyses were conducted in which response categories were partitioned into temporal groupings (presented in Sec. III B) and spectral groupings (presented in Sec. III C).

B. Effects of context on temporal assimilation patterns

For this analysis, the first question was whether Japanese listeners distinguished AE vowels in terms of their temporal structure, as evidenced by differential assimilation of short, intermediate, and long vowels to one-mora and two-mora J categories. Thus, responses were partitioned into one-mora responses [i, e, a, o, u, ʲa, ʲu, ʲo] versus two-mora responses [ii, ee, aa, oo, uu, ei, ou, ʲaa, ʲuu, ʲoo] and percentages of each type of temporal response were computed for each AE vowel in each context within each speaker condition. On average, the five long vowels were heard as most similar to two-mora J vowels 79% of the time (range 61% to 92%), while the two intermediate vowels were assimilated to two-mora categories 74% and 73% of the time, respectively. In contrast, the four short vowels were heard as most similar to two-mora J vowels on only 27% of the trials (range 19% to 37%). Thus, on 76% of the trials, long/intermediate versus short AE vowels were ''correctly'' differentiated as (most similar to) two-mora versus one-mora J categories, respectively.

A second question was whether patterns of temporal assimilation differed systematically as a function of the voicing of the following consonant, as would be expected if Japanese listeners did not ''compensate'' for non-native allophonic variation in vocalic duration. In Fig. 1(a), percentages of two-mora responses for the five long, two intermediate and four short AE vowels in final voiced versus voiceless consonant contexts (pooled over repetitions, vowels, speakers, and place-of-articulation of consonants) are plotted. For all four speakers' utterances, there was a clear effect of final consonant voicing on temporal assimilation patterns, especially for the short and intermediate vowels. All four short vowels were assimilated to two-mora J vowel categories less often in voiceless final consonant contexts than in voiced contexts. The effect of voicing context on temporal assimilation of intermediate vowels appeared similar to that for short vowels for three speakers' utterances, but smaller for the utterances of JM. Finally, the five long vowels were assimilated to two-mora response categories in both voicing contexts at about the same levels.

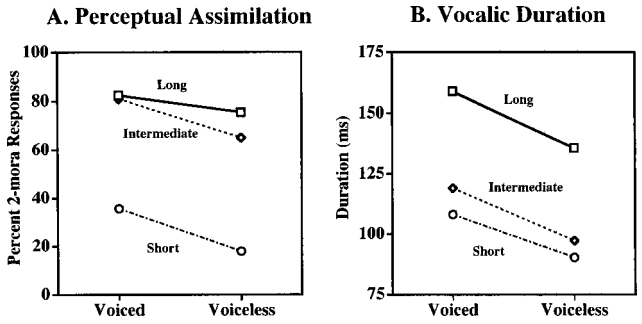


FIG. 1. (a) Consistency in perceptual assimilation (in percentage of opportunities) to two-mora Japanese categories of long [eɪ, æː, ɑː, ɔː, o͡ʊ], intermediate [iː, uː], and short [ɪ, ɛ, ʌ, ʊ] American English vowels in voiced and voiceless following consonant contexts. Data are pooled over speakers, vowels within each length group, and labial, alveolar, and velar contexts within each voicing category. (b) Vocalic duration of long, intermediate, and short vowels in voiced and voiceless following consonant contexts. Data are pooled over tokens, vowels within each length group, speakers, and labial, alveolar and velar contexts within each voicing category.

To assess the reliability across listeners of these overall trends, a repeated measures analysis of variance (ANOVA) with vowels (11) and contexts (6) nested within speaker (listener) groups was performed, using the number of two-mora responses as the score. The main effects of vowels and contexts were both highly significant, $F(10,200)=76.360$ and $F(5,100)=16.287$, respectively, both $p<0.001$. While the main effect of speakers was not significant [$F(3,20)=0.801$], there were small but significant interactions of speakers×contexts [$F(15,100)=2.219$, $p<0.02$], Speakers ×vowels [$F(30,200)=1.804$, $p<0.01$], and speakers ×vowels×contexts [$F(150,1000)=2.022$, $p<0.01$]. The vowels×contexts interaction was also highly reliable [$F(50,1000)=3.607$, $p<0.01$]. A planned comparison of the three voiced contexts versus the three voiceless contexts also yielded a statistically significant result, $F(1,100)=68.842$, $p<0.001$.

To summarize, the patterns of temporal assimilation shown in Fig. 1(a) were highly reliable across listeners. On average, vowels were judged as less similar to two-mora J categories in voiceless contexts, especially for the intermediate and short vowels. The extent of this contextual effect varied somewhat across speakers. However, despite these context/speaker effects, the four short vowels were nevertheless most often assimilated to one-mora categories, while the intermediate and long vowels were most often assimilated to two-mora categories for all speakers' vowels in both voiced and voiceless final consonant contexts.

Figure 1(b) presents the vocalic duration data for the four short, two intermediate, and five long vowels (averaged over tokens, vowels, speakers, and place-of-articulation of the consonants). For all three sets of vowels, vocalic duration varied systematically with the voicing of the final consonant. On average, vowels preceding voiced consonants were 20% longer than vowels preceding voiceless consonants. Short vowels in voiced contexts averaged 18 ms longer, intermediate vowels averaged 22 ms longer, and long vowels averaged 26 ms longer than their counterparts in voiceless contexts.

In order to assess the extent to which temporal assimilation patterns were predictable on the basis of differences in vocalic duration alone (independent of speaker and context), a Spearman rank-order correlation was performed in which the vocalic duration of each vowel (averaged over repetitions) in each consonantal context and speaker condition was the predictor variable, and the number of two-mora assimilation responses (averaged over repetitions and listeners) was the dependent variable. Results indicated a moderate but statistically reliable correlation ($\rho=0.637$, $p<0.001$). Factors other than absolute vocalic duration clearly accounted for additional variation in temporal assimilation patterns.

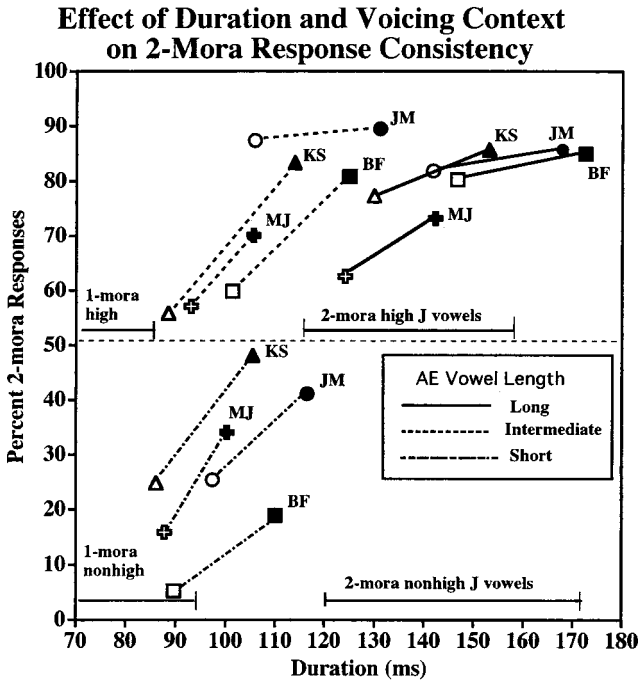


FIG. 2. Relationship between vocalic duration (abscissa) and consistency in perceptual assimilation (in percentage of opportunities) to two-mora Japanese categories. Speakers are indicated by initials and shape of symbols. Filled symbols indicate duration of vowels in voiced following consonant contexts; unfilled symbols indicate duration of vowels in voiceless contexts. Lines connect data points for each speaker's utterances: solid lines for long vowels, dashed lines for intermediate vowels; dash-dot lines for short vowels. Data are averaged over tokens (three) and vowels within each length group (four short vowels, two intermediate vowels, five long vowels). Ranges in duration of one-mora and two-mora high and nonhigh (mid and low) Japanese vowels produced in similar CVC contexts in a short sentence are indicated on the graph for comparison.

Figure 2 plots the relationship between vocalic duration and assimilation to two-mora categories for sets of short, intermediate, and long vowels produced by each speaker. Vowels in voiced contexts are indicated by filled symbols; vowels in voiceless contexts are indicated by open symbols. While this plot clearly shows a positive relationship between vocalic duration and temporal assimilation pattern, it also indicates that intermediate vowels were assimilated more often to two-mora categories than were short vowels, even when they overlapped in (absolute) vocalic duration. Vowel duration varies systematically with vowel height in Japanese, as well as in English and other languages (Fitzgerald, 1996; Peterson and Lehiste, 1960; Strange and Bohn, 1998). For comparison with the AE vowels used here, ranges in duration of Japanese vowels from utterances produced in similar

TABLE III. Overall pattern of perceptual assimilation of AE vowels to the five Japanese spectral categories, disregarding length and palatalization. Data (percent of opportunities) are pooled over all six contexts and all four speakers.

	Japanese response categories				
AE vowel	High-front i, ii	Mid-front e, ee, ei	Low a, aa ^{j}a, ^{j}aa	Mid-Back o, oo, ou, ^{j}o, ^{j}oo	High-back u, uu, ^{j}u, ^{j}uu
iː	**98**	2			
ɪ	**51**	44	1		4
e͡ɪ	5	**95**			
ɛ	2	**53**	44	<1	<1
æː	1	26	**73**		
ɑː		1	**81**	16	2
ʌ	<1	1	**51**	39	8
ɔː			13	**82**	5
o͡ʊ	<1		<1	**79**	20
ʊ			3	15	**82**
uː				<1	**99**

sentential and consonantal contexts are marked on the graph (Fitzgerald, 1996). High J vowels are indicated above the 50% point of the ordinate; mid and low (nonhigh) J vowels are indicated above the 0% point. Note the differences especially in upper boundaries of the ranges for Japanese high versus nonhigh vowels. Japanese listeners appear to have considered L1 phonetic variation in vocalic duration associated with vowel height when judging AE mid and short vowels as more similar to two-mora or one-mora vowels. They did not appear to have attributed any part of the variation in vocalic duration to the voicing of the following consonant, since such allophonic variation in vowel length is not present in Japanese. Thus, for short and intermediate AE vowels (whose durations were ambiguous in comparison with J vowels), assimilation responses differed significantly with voicing context and the relationship between absolute duration and percentage of two-mora responses appeared to be quite linear.

C. Effects of context on spectral assimilation patterns

For this analysis, the 18 response alternatives were partitioned into the five spectral categories: high-front, mid-front, low, mid-back, and high-back. That is, differences in temporal assimilation were ignored, as were differences between two-vowel sequences [ei, ou] versus double vowels [ee, oo] and palatalized versus nonpalatalized vowel responses [a(a), o(o), u(u)] versus [^{j}a(a), ^{j}o(o), ^{j}u(u)]. Table III presents the overall distribution of assimilation responses to these five categories for each AE vowel. Data were pooled over all six contexts and all four speaker conditions and converted to percentages of opportunities. The boldfaced numbers highlight the overall modal response choices for each vowel.

As these data show, there was considerable variability in the overall consistency with which particular AE vowels were assimilated to particular J spectral categories. The 11 AE vowels can be grouped into three clusters. The first group—[iː, e͡ɪ, uː]—were assimilated to their spectrally similar J counterparts with greater than or equal to 95% overall consistency. Consistency in spectral assimilation varied from 89% to 100% across the six contexts (pooling across speaker conditions). Within particular speaker/context conditions, percentages ranged from 89% to 100% for [i, u] while they were slightly more variable for [e͡ɪ] (74% to 100%).

The second group of vowels [æː, ɑː, ɔː, o͡ʊ, ʊ] were assimilated to J spectral categories with overall consistency between 73% and 82%. Across contexts (pooling over speakers) consistency ranged from 65% to 93%. However, when individual speaker/context conditions were considered, spectral assimilation patterns showed considerably more variability for these vowels than for the first group (range 31% to 100%). For [æː, ɑː, o͡ʊ, ʊ], the modal spectral response actually differed from the overall modal response in one or more speaker/context conditions.

The third group of vowels [ɪ, ɛ, ʌ] showed the least overall consistency (51% to 53%) and the greatest variability in spectral assimilation patterns across contexts (ranging from 29% to 74%). When individual speaker/context conditions were considered, the range in selection of the overall modal response was even greater (4% to 96%). For these vowels, then, no single spectral J category was chosen over 50% of the time for every speaker (pooling over contexts) or for every context (pooling over speakers).

To quantify the extent to which these variations in spectral assimilation patterns were systematic across contexts and/or speakers, new PRE analyses were conducted for the eight vowels showing less than 95% overall consistency. As before, a PRE of 10% or greater was considered a meaningful effect. These analyses indicated that variations in spectral assimilation patterns for the [æː, ɑː, ɔː, o͡ʊ, ʊ] were not related sufficiently to context and speaker variables to reach *a priori* levels of reduction in errors. That is, while consistency (and sometimes modal response choice) did change across context/speaker conditions, there was also considerable variability within conditions, such that predictions of assimilation patterns were not much improved by consideration of the context and the speaker. However, for the three short vowels [ɪ, ɛ, ʌ] the PRE analyses revealed clear systematic effects of context for one or more speakers, as shown in Table IV(a). Here again, base rates of error (column 2) and PRE values for context, speaker, and context/speaker

TABLE IV. (a) Results of proportion of reduction of error (PRE) analysis for spectral assimilation response patterns for /ɪ, ɛ, ʌ/. (b) Results of PRE analysis for ''movement'' versus ''nonmovement'' assimilation response patterns for /æ/. See text for details.

AE vowel	Base rate of error	Context effects	Speaker effects	Context × speaker interaction
(a)				
ɪ	48.69	**0.185**	**0.147**	**0.357**
ɛ	47.45	**0.115**	**0.280**	**0.380**
ʌ	48.69	0.060	**0.260**	0.281
(b)				
æː	30.40	**0.193**	0.000	**0.391**

TABLE V. Results of multiple regression analysis relating formant frequencies to Japanese spectral category assimilation responses (percent assimilated to modal spectral category).

AE vowel	Modal J category	Predictor variables (formant values)	Multiple R	Adjusted R^2	p value
/ɪ/	/i, ii/	$F1$	0.475	0.226	<0.001
		$F2$	0.212	0.031	=0.074
		$F1$ and $F2$	0.498	**0.236**	<0.001
/ɛ/	/e, ee/	$F1$	0.772	0.591	<0.001
		$F2$	0.538	0.279	<0.001
		$F1$ and $F2$	0.773	**0.586**	<0.001
/ʌ/	/a, aa, ʲa, ʲaa/	$F1$	0.333	0.098	=0.004
		$F2$	0.314	0.086	=0.007
		$F1$ and $F2$	0.482	**0.211**	<0.001

interactions are given, with values exceeding 0.10 in boldface.

While the PRE analysis of the spectral assimilation patterns for AE [æː] did not reveal a meaningful effect of context and/or speaker, an inspection of the data indicated that the proportion of assimilation responses indicating the perception of a syllable with changing spectral characteristics (i.e., assimilation to palatalized J categories or two-vowel sequences) appeared to vary systematically with contexts and speakers. Thus, an additional analysis was performed in which J responses to this vowel were partitioned into ''nonmovement'' J response categories [a, aa, e, ee] versus ''movement'' J response categories [ʲa, ʲaa, ei]. The results of a PRE analysis based on this partitioning of J response categories is shown in Table IV(b). The significant reduction in errors associated with the main effect of context and the speaker/context interaction support the conclusion that the perceptual assimilation pattern varied systematically with context for three of the speakers' utterances. Specifically, more movement assimilation responses were given to the vowel in velar consonant contexts.

To determine the extent to which systematic variation in spectral assimilation pattern as a function of speakers/contexts could be accounted for by production differences [ɪ, ɛ, ʌ], multiple regression analyses were performed relating formant values and modal response consistency (across listeners and trials/listener) for the 72 tokens of each vowel. For each analysis, the independent variables were $F1$ frequency, $F2$ frequency, and $F1/F2$ frequencies combined (values measured at the 50% duration point in each syllable), while the dependent variable was the number of modal responses (out of 18 possible; 6 listeners×3 trials) for each vowel token. Table V presents the multiple R and adjusted R-squared values for each combination of independent variables. For AE[ɪ], assimilation to high (as opposed to mid) front J vowels correlated significantly with formant frequency, especially $F1$; however, less than 24% of the variance in perceptual assimilation responses was accounted for by spectral variation. For [ɛ], $F1$ and $F2$ frequencies, both alone and in combination, showed significant correlations with perceptual assimilation to mid (as opposed to low) J vowels; together, they accounted for about 59% of the variance. Thus, vowel height of AE short front vowels, as cued by $F1$ frequency (and to a lesser extent $F2$ frequency), accounted for some of the differences in assimilation pattern across speakers and contexts. Mid-high and mid-low AE vowels with higher $F1$ values and lower $F2$ values were more likely to be assimilated to mid and low J vowels, respectively. For AE [ʌ], assimilation to low (as opposed to mid) back J vowels showed a significant, but weak relationship to formant frequencies; $F1$ and $F2$ combined accounted for only 21% of the variance.

A final analysis investigated the relationship between ''vowel-intrinsic spectral change'' and perceptual assimilation to palatalized and two-vowel J responses categories for AE [æː]. A Spearman rank-order correlation was computed in which the predictor variable was change in $F1/F2$ values (Euclidean distance in $F1/F2$ bark space) across the middle half of each syllable (from 25% to 75% durational points). The dependent variable was the number of assimilation responses to ''movement'' categories (CʲaCʲaCei). There was no overall significant relationship between extent of formant movement and assimilation to movement response categories ($\rho = -0.083$, $p = 0.492$). Thus, although perceptual assimilation to J movement categories was systematically greater in velar consonant contexts, this was not significantly related to greater diphthongization of this vowel in these contexts. It might be that perceptual assimilation to palatalized categories was associated with slower formant transitions of velar consonant-to-vowel opening and closing gestures, but this hypothesis was not testable with the acoustic measurements available for these stimuli.

In summary, the analysis of spectral assimilation patterns revealed that the short AE vowels [ɪ, ɛ, ʌ] were perceptually assimilated to *different* spectral categories depending upon the context in which the vowel was produced and presented. An analysis of the pattern of assimilation of the long AE vowel [æː] to palatalized and two-vowel versus monophthongal J vowel categories also revealed systematic effects of context for three of the four speakers' utterances. Spectral assimilation of the other new vowels [ɔː, ʊ] was more stable across contexts and speakers, with overall consistency about equal to AE vowels [ɑː, o͡ʊ] which are considered more similar to J vowels in phonological analyses. Correlations between formant cues associated with vowel height and spectral assimilation patterns showed only weak associations

between production variations and J listeners' perceptual assimilation patterns.

IV. DISCUSSION

A. Summary of results

The results of this study have shown that the 11 vowels of AE, when produced and presented to Japanese listeners in CVC syllables embedded in a carrier sentence, were perceptually assimilated to native J vowel categories with varying degrees of category goodness, as measured by consistency of assignment to a particular native category and goodness ratings. *No vowel* was perceptually assimilated to a single J response category consistently across all consonantal contexts. As might be predicted from (abstract) linguistic descriptions of the two vowel inventories, the AE vowels [iː, e͡ɪ, ɑː, o͡ʊ, uː] were most consistently assimilated to their spectrally similar counterparts in Japanese [ii, ei, aa, ou, ɯɯ], with relatively little variation across context/speaker conditions. However, there were systematic differences in temporal assimilation patterns especially for [iː, uː] as a function of the voicing of the following consonant. For five of the six ''new'' AE vowels [ɪ, ɛ, æː, ʌ, ʊ], spectral and/or temporal assimilation patterns varied systematically across context/speaker conditions.

Temporal assimilation patterns indicated that long and intermediate AE vowels were most often assimilated to two-mora J categories, while short AE vowels were more often assimilated to one-mora categories, with overall ''correct'' temporal differentiation on 76% of the trials. This is somewhat lower than reported in our earlier study (Strange *et al.*, 1998) which investigated Japanese listeners' perceptual assimilation of vowels produced in /hVb/ syllables in the same carrier sentence (84% ''correct'' temporal assimilation). However, temporal differentiation was superior to that shown for citation-form /hVbɑ/ bisyllables (53% ''correct''), suggesting that sentence context helps L2 listeners perceive temporal cues associated with segmental contrasts even when those cues are more subtle than in their native language.

The greater variability in assimilation of AE vowels to one-mora versus two-mora J categories in this mixed CVC study, in comparison with our earlier findings for /hVb/ syllables in sentences, was due primarily to the short vowels being assimilated more often to two-mora J responses categories when they were produced in final voiced consonant contexts (36% responses versus 17% in voiceless final consonants; 17% for /hVb/), and to the intermediate vowels being assimilated more often to one-mora J response categories in voiceless final consonant contexts (44% vs 19% in voiced final consonant contexts; 15% for /hVb/). The long vowels were also assimilated somewhat less consistently to two-mora J response categories in voiceless final consonant contexts (76% vs 82% in voiced contexts; 85% for /hVb/). This supports the prediction that the allophonic rule in English whereby vocalic duration differences are attributed (in part) to the voicing status of final consonants would not be ''honored'' by the Japanese listeners. The moderate correlation found between absolute vocalic duration and perceptual assimilation to two-mora responses suggests that both physical duration and native-language phonetic realization rules (variation in vocalic duration with vowel height) contributed to judged temporal similarity. That is, J listeners did not judge L2 vowel length on the basis of a very abstract (vowel and context-independent) characterization such as moraic count, nor on the basis of surface level acoustic duration (utterance-specific), but rather made L1/L2 phonetic similarity judgments which were consistent with L1, but not L2, phonetic realization rules (vowel-height-specific, but independent of following consonant context).

Spectral assimilation of the six ''new'' AE vowels also revealed differences in their judged goodness as instances of native categories not predictable from a context-independent and very abstract characterization. AE [ɔː, ʊ] were assimilated to J mid- and high-back categories, respectively, with about the same consistency as the similar AE vowels [ɑː] and [o͡ʊ] were assimilated to J low and mid-back categories, and did not show systematic context/speaker effects in spectral assimilation patterns. It is interesting to note that AE [ɑː] and [ɔː] were heard by Japanese as perceptually similar to different native categories, perhaps due to the perceived rounded quality of the latter. Thus, while AE speakers often confuse these two vowels, J L2 learners might be expected to confuse [ɔː] and [o͡ʊ] more often. [However, some dialects of Japanese are said to contrast /ɔ/ and /o/ (Sugito, 1996)]. Likewise, AE [ʊ], which was sometimes confused with [ʌ] by native speakers of AE, would be expected to be confused more with AE [uː] by Japanese learners of English. This is especially the case because of the reduced temporal differentiation of [ʊ-uː] by J listeners. In addition, the mid-high back AE [ʊ] has a relatively high *F*2, especially in alveolar contexts. Thus, it is similar acoustically to the unrounded [ɯ] of Standard (Kanto) Japanese. While most of our Japanese listeners were from the Kansai region, they apparently heard AE [ʊ] as quite similar to the Standard variant of the J high-back vowel.

The remaining new vowels [ɪ, ɛ, ʌ, æː] were considered poor instances of any one J spectral category and yielded large differences in spectral assimilation patterns as a function of context. J listeners assimilated AE [æː] in velar contexts more often to palatalized J syllables, but more often to nonpalatalized categories when in labial and alveolar contexts. The very low consistency in the overall spectral assimilation pattern for AE [ɪ, ɛ, ʌ] was attributable, in part, to systematic contextual effects on the J category chosen as the most similar. AE [ɪ] was assimilated to J [i] more often in alveolar contexts, but to J [e] in labial and velar contexts. AE [ɛ] was assimilated to J [e] most often in voiced velar contexts, and less consistently in voiced labial and alveolar contexts, whereas in voiceless contexts, it was assimilated more to J [a]. AE [ʌ] was assimilated more often to J [a] in voiceless contexts (especially g-k), but more often to J [o] in voiced labial and velar contexts. These systematic context effects were only weakly associated with differences in formant structure associated with coarticulatory and idiolectal variation in the production of these vowels.

In general, the data presented here indicate that J listeners' patterns of perceptual assimilation of AE vowels are not

TABLE VI. Consistency (in percent) of assimilation to Japanese spectral categories as a function of consonantal context, pooled over all four speakers. Entries for [æː] indicate only nonpalatalized responses.

		Consonant context					
	J spectral	Labial		Alveolar		Velar	
AE vowel	category	b-b	b-p	d-d	d-t	g-g	g-k
iː	High front /i, ii/	100	100	94	95	99	99
ɪ		49	38	73	70	50	29
uː	High back /u, uu/	100	100	96	99	100	100
ʊ		81	71	91	91	83	72
ɐɪ	Mid front /e, ee, ei/	100	100	89	91	97	94
ɛ		56	44	57	45	74	39
oʊ	Mid back /o, oo, ou/	77	81	71	66	93	88
ɔː		83	86	77	74	87	88
æː	Low central /a, aa/	60	65	64	60	25	20
ɑː		75	78	77	81	84	93
ʌ		39	52	50	54	45	67

highly predictable either from cross-language descriptions at the abstract (context-independent) phonological feature level, nor from surface level acoustic parameters associated with vowel identity. Thus, we will argue that cross-language phonetic similarity must be defined at an intermediate level of abstraction, in which context-specific phonetic realization rules in L1 are brought to bear on cross-language judgments of similarity. The most accurate way of establishing cross-language perceptual similarities is by direct assessment, using stimulus materials in which phonetic variation is present and tasks such as the ones employed here.

Current theories of L2 speech learning are concerned with how perceptual assimilation patterns might be used to predict better the perceptual difficulties adult L2 learners face in acquiring new phonetic contrasts. The data presented here might be used to predict relative difficulties of Japanese learners acquiring AE vowel distinctions. Specifically, the results indicate that perceptual difficulties will vary significantly with the phonetic context in which the vowel pairs are presented.

B. Predicted perceptual difficulty of AE vowel pairs for Japanese listeners

Using Best's perceptual assimilation model (1994, 1995), the perceptual assimilation patterns revealed in the present study would predict that seven pairs of AE vowels will be difficult for Japanese listeners to perceptually differentiate because they reflect single category or category goodness assimilation patterns in at least some contexts: /i-ɪ, u-ʊ, e-ɛ, o-ɔ, æ-ɑ, ɑ-ʌ, ɪ-ɛ/. However, all but one of these pairs, /o-ɔ/, include an AE vowel which differed in modal category assignment as a function of context. Table VI summarizes the spectral assimilation data across contexts for five sets of vowels which, in some contexts, at least, were assimilated to the same native category.

Consider first the two vowel contrasts between high and mid-high AE vowels: [iː-ɪ] and [uː-ʊ], shown in the first two boxed rows. While these two pairs constitute the ''same'' non-native phonological feature contrast (height or, alternatively, tenseness), the perceptual assimilation data would predict differences in relative perceptual difficulty both within and across pairs of vowels. Overall, these pairs constitute a category goodness (CG)-type assimilation pattern: the high AE vowels are excellent (spectral) exemplars, whereas the mid-high AE vowels are poorer exemplars of J high vowels. However, context-specific assimilation patterns predict that both pairs of vowels would be more difficult to differentiate in alveolar context than in labial and velar contexts, at least for the AE speakers' utterances used in this study. Indeed, for [uː-ʊ] in alveolar contexts, the spectral assimilation pattern may be considered a single category (SC) type, with both AE vowels being considered very good instances of J[u(u)]. For the front vowel pair in labial and velar contexts, the spectral assimilation pattern is more a two-category (TC) type, with AE [ɪ] being more often assimilated to J [e], whereas the back vowel pair in these contexts reflect a CG pattern in these contexts. Differentiation on the basis of temporal assimilation pattern would also be difficult for both these vowel pairs, since assimilation to one-mora versus two-mora categories for all four of these vowels differed significantly with consonantal context. Thus, in Best's taxonomy, this feature contrast shows a TC pattern in some contexts, a CG pattern in others, and a SC pattern in still other contexts. Difficulty in learning to differentiate these vowels, then, would be predicted to vary from very easy to quite difficult, depending upon context.

Consider next the two mid/mid-low AE vowel pairs [ɐɪ-ɛ] and [oː-ɔː] (shown in the next two rows of Table VI). The front pair constitutes a CG-type spectral assimilation pattern (/eɪ /an excellent exemplar of J/ei/ or /ee/, /ɛ/ a poor exemplar of J/e(e)/) or a TC pattern (/ɛ/ an exemplar of J/a/), depending upon context. Again, differential temporal assimilation patterns would be of only marginal help to J listeners in mixed consonantal contexts, since [ɛ] was often assimilated to two-mora categories in voiced contexts. The back pair would be predicted to be even more difficult to differentiate in all contexts. Both [o͡ʊ] and [ɔː] were assimilated primarily to two-mora native categories and were heard as good exemplars of mid-back J vowels [oo,ou]; i.e., the pair constitutes a SC-type pattern. However, response patterns did suggest that Japanese listeners perceptually differentiated the diphthongal pattern in [o͡ʊ] from the monophthongal pattern of [ɔː]. If this were the case, then discrimination would be predicted to be easier for these vowels in contexts in which the AE mid vowels are more diphthongized (e.g., preceding voiced consonants and in open syllables).

The low and mid-low AE vowels [æː, ɑː, ʌ] were all assimilated most often to the low J vowels; however, they differed considerably in perceived category goodness. AE [ɑː] was judged a relatively good exemplar of J [a(a)], especially in velar contexts, while [ʌ] was judged a marginal exemplar in voiceless contexts and more like J [o(o)] in voiced labial and velar contexts. AE [æː] was a poor exemplar of J a(a) in labial and alveolar contexts and more like

$g^j a(a)$ in velar contexts. Thus, the [ɑː-æː] contrast constitutes a CG-type pattern in labial and alveolar contexts (or possibly categorizable-uncategorizable), but a TC pattern in velar contexts. The [ɑː-ʌ] contrast constitutes a CG type in some contexts and a TC type in others; furthermore, differentiation would be expected to be helped by perceived temporal differences. Thus, [ɑː-ʌ] should be easier to discriminate than [ɑː-æː] except in velar contexts.

Finally, the AE pair [ɪ-ɛ] (not shown in adjacent rows in Table VI) were both assimilated to the J [e] category in some contexts for some speakers. Since these vowels were both heard as most similar to one-mora J vowels, we would predict on the basis of spectral assimilation patterns that they would be most difficult to differentiate in labial contexts (CG pattern) and easiest in alveolar contexts (TC pattern) for the utterances used in this study.

The patterns of temporal and spectral assimilation of AE vowels to native J vowel categories reported above suggest generally that all 11 AE vowels are perceptually distinct for Japanese L2 learners of English in at least some CVC contexts. That is, it was never the case that two vowels were equally well assimilated to the same J vowel category in all context/speaker conditions. However, predicted perceptual difficulties based on the overall assimilation patterns, or any particular context condition, would not be as accurate as predictions which take contextual variables into account. In addition to differences which can be attributed to task effects, L2 experience effects, and other methodological variables, context effects such as those demonstrated here might account for the lack of consistency in previous studies of L2 vowel perception outlined in the introduction. In future research on cross-language vowel perception, it will be necessary to consider the prosopic and allophonic details of the stimulus materials in interpreting the outcomes of perceptual assimilation and L2 discrimination performance. On the practical side, results such as the ones reported here can be used to generate ''graded'' stimulus materials (Jamieson and Morosan, 1986) for training studies of non-native vowel contrasts. Finally, the results reported here supply additional evidence that an adequate description of the effects of the first language phonology on the perception (and production) of L2 phonetic segments by adult second language learners will require cross-language comparisons of the phonetic details of both L1 and L2 systems. Neither very abstract phonological descriptions of phoneme inventories nor acoustic comparisons of specific realizations of phoneme categories will be adequate in predicting cross-language *perceptual* similarities. Rather, an intermediate level of abstraction, in which language-specific ''rules'' about phonetic variation are brought to bear on categorization processes, appears to be the appropriate level of analysis of L1/L2 phonetic relationships.

ACKNOWLEDGMENTS

The authors wish to express their thanks to Takahiro Adachi for programming the experiments, to David H. Thornton, Robin L. Rodriguez, and Amanda Grainger for their assistance in data analysis, and to James J. Jenkins for his comments on the manuscript. This research was supported by a grant to the first author from the NIH (Grant No. NIDCD-00323). Requests for reprints should be addressed to the first author: PhD Program in Speech and Hearing Sciences, CUNY-Graduate School.

[1]In this study, it was impractical to use a completely repeated measures design, because of the large number of trials. In the previous study (Strange *et al.*, 1998) using subjects drawn from the same population, we found that perceptual assimilation patterns were quite stable across independent groups of listeners. In contrast, we found systematic differences in perceptual assimilation patterns across speakers for the same listeners. In the present study, speakers and listeners are confounded variables. However, the main variable of interest (i.e., consonantal context) was a within-listeners variable. Thus, one can think of this design as consisting of four replications of an investigation of context effects, using different stimuli (speakers) and different listeners.

[2]In this article, two-mora monophthongal Japanese vowels will be indicated in IPA transcriptions by two identical symbols

[3]There is some ambiguity with respect to whether the Katakana symbols [ei] and [ou] represent a two vowel sequence. In Hiragana (the syllabary which is used for words of Japanese origin, as opposed to loan words) the monophthongal two-mora vowels [ee] and [oo] are often written as [ei] and [ou], respectively.

Andruski, J. E., and Nearey, T. M. (**1992**). ''On the sufficiency of compound target specifications of isolated vowels and vowels in /bVb/ syllables,'' J. Acoust. Soc. Am. **91**, 390–410.

Beckman, M. (**1982**). ''Segment duration and the 'mora' in Japanese,'' Phonetica **39**, 1113–135.

Best, C. T. (**1994**). ''The emergence of native-language phonological influences in infants: A perceptual assimilation model,'' in *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, edited by J. Goodman and H. C. Nusbaum (MIT, Cambridge, MA), pp. 167–224.

Best, C. T. (**1995**). ''A direct realist view of cross-language speech perception,'' in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 171–204.

Best, C. T., Faber, A., and Levitt, A. (**1996**). ''Assimilation of non-native vowel contrasts to the American English vowel system,'' J. Acoust. Soc. Am. **99**, 2602 (abstract).

Best, C. T., McRoberts, G. W., and Sithole, N. M. (**1988**). ''Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English speaking adults and infants,'' J. Exp. Psychol. **14**, 345–360.

Bohn, O.-S. (**1995**). ''Cross-language speech perception in adults: First language transfer doesn't tell it all,'' in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 279–304.

Campbell, W. N., and Sagisaka, Y. (**1991**). ''Moraic and syllable-level effects on speech timing,'' Technical Report of IEICE **SP90-107**, pp. 35–40.

Crystal, T. H., and House, A. S. (**1988a**). ''Segmental duration in connected-speech signals: Current results,'' J. Acoust. Soc. Am. **83**, 1553–1573.

Crystal, T. H., and House, A. S. (**1988b**). ''Segmental duration in connected-speech signals: Syllabic stress,'' J. Acoust. Soc. Am. **83**, 1574–1585.

Fitzgerald, B. H. (**1996**). ''Acoustic analysis of Japanese vowels produced in multiple consonantal contexts,'' unpublished Master's thesis, University of South Florida.

Flege, J. E. (**1992**). ''Speech learning in a second language,'' in *Phonological Development: Models, Research, and Applications*, edited by C. Ferguson, L. Menn, and C. Stoel-Gammon (York, Timonium, MD), pp. 565–604.

Flege, J. E. (**1995**). ''Second language speech learning: Theory, findings, and problems,'' in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 233–277.

Flege, J. E., and Eefting, W. (**1987**). ''The production and perception of English stops by Spanish speakers of English,'' J. Phonetics **15**, 67–83.

Flege, J. E., and Hillenbrand, J. (**1987**). ''Differential use of closure voicing and release burst as cues to stop voicing by native speakers of French and English,'' J. Phonetics **15**, 203–208.
Fourakis, M. (**1991**). ''Tempo, stress, and vowel reduction in American English,'' J. Acoust. Soc. Am. **90**, 1816–1827.
Gottfried, T. L. (**1984**). ''Effects of consonant context on the perception of French vowels,'' J. Phonetics **12**, 91–114.
Gottfried, T., and Beddor, P. S. (**1988**). ''Perception of temporal and spectral information in French vowels,'' Lang. Speech **31**, 57–75.
Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.
Homma, Y. (**1981**). ''Durational relationship between Japanese stops and vowels,'' J. Phonetics **9**, 273–281.
Homma, Y. (**1992**). *Acoustic Phonetics in English & Japanese* (Yamaguchi, Kyoto, Japan).
Huang, C. (**1991**). ''An acoustic and perceptual study of vowel formant trajectories in American English,'' Report 563, Research Laboratory in Electronics, MIT, Cambridge, MA.
Jamieson, D. G., and Morosan, D. E. (**1986**). ''Training non-native speech contrasts in adults: Acquisition of the English /ð/-/θ/ contrast by francophones,'' Percept. Psychophys. **40**, 205–215.
Keating, P. A., and Huffman, M. K. (**1984**). ''Vowel variation in Japanese,'' Phonetica **41**, 191–207.
Klatt, D. (**1976**). ''Linguistic uses of segmental duration in English: Acoustic and perceptual evidence,'' J. Acoust. Soc. Am. **59**, 1208–1221.
Kohler, K. J. (**1981**). ''Contrastive phonology and the acquisition of phonetic skills,'' Phonetica **38**, 213–226.
Kuhl, P. K., Williams, L. A., Lacerda, F., Stevens, K. N., and Linblom, B. (**1992**). ''Linguistic experience alters phonetic perception in infants by 6 months of age,'' Science **255**, 606–608.
Lively, S. E., Logan, J. S., and Pisoni, D. B. (**1993**). ''Training Japanese listeners to identify English /r/ and /l/ II: The role of phonetic environment and talker variability in learning new perceptual categories,'' J. Acoust. Soc. Am. **94**, 1242–1255.
Mochizuki, M. (**1981**). ''The identification of /r/ and /l/ in natural and synthesized speech,'' J. Phonetics **9**, 283–303.
Moon, S-J., and Lindblom, B. (**1994**). ''Interaction between duration, context, and speaking style in English stressed vowels,'' J. Acoust. Soc. Am. **96**, 40–55.
Morosan, D. E., and Jamieson, D. G. (**1989**). ''Evaluation of a technique for training new speech contrasts: Generalization across voices, but not word position or task,'' J. Speech Hear. Res. **32**, 501–511.
Nearey, T., and Assmann, P. (**1986**). ''Modeling the role of inherent spectral change in vowel identification,'' J. Acoust. Soc. Am. **80**, 1297–1308.
Peterson, G. E., and Barney, H. L. (**1952**). ''Control methods used in the study of vowels,'' J. Acoust. Soc. Am. **24**, 175–184.
Peterson, G. E., and Lehiste, I. (**1960**). ''Duration of syllable nuclei in English,'' J. Acoust. Soc. Am. **32**, 693–703.
Polka, L. (**1991**). ''Cross-language speech perception in adults: Phonemic, phonetic, and acoustic factors,'' J. Acoust. Soc. Am. **89**, 2961–2977.
Polka, L. (**1992**). ''Characterizing the influence of native language experience on adult speech perception,'' Percept. Psychophys. **52**, 37–52.
Polka, L. (**1995**). ''Linguistic influences in adult perception of non-native vowel contrasts,'' J. Acoust. Soc. Am. **97**, 1286–1296.
Polka, L., and Bohn, O-S. (**1996**). ''A cross-language comparison of vowel perception in English-learning and German-learning infants,'' J. Acoust. Soc. Am. **100**, 577–592.
Reynolds, H. T. (**1984**). *Analysis of Nominal Data (Second Edition)* (Sage, Beverly Hills).
Schmidt, A. M. (**1996**). ''Cross-language identification of consonants: Part 1. Korean perception of English,'' J. Acoust. Soc. Am. **99**, 3201–3211.
Sheldon, A., and Strange, W. (**1982**). ''The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception,'' Appl. Psycholing. **3**, 243–261.
Shibatani, M. (**1990**). *The Languages of Japan* (Cambridge U. P., Cambridge, UK).
Stevens, K. N., and House, A. S. (**1963**). ''Perturbation of vowel articulations by consonantal context: An acoustical study,'' J. Speech Hear. Res. **6**, 111–128.
Strange, W. (**1995**). ''Cross-language studies of speech perception: A historical review,'' in *Speech Perception and Linguistic Experience*, edited by W. Strange (York, Timonium, MD), pp. 3–45.
Strange, W., and Bohn, O-S. (**1998**). ''Dynamic specification of coarticulated German vowels: Perceptual and acoustical studies,'' J. Acoust. Soc. Am. **104**, 488–504.
Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., Nishi, K., and Jenkins, J. J. (**1998**). ''Perceptual assimilation of American English vowels by Japanese listeners,'' J. Phonetics **26**, 311–344.
Sugito, M. (**1996**). *Nihongo onsei no kenkyuu (Studies of Japanese speech sounds). Series 3, Nohongo no Otto (Japanese Speech Sounds)*.
Wickens, T. D. (**1989**). *Multiway Contingency Tables Analysis for the Social Sciences* (Erlbaum, Hillsdale, NJ).
Wolfram, W. (**1991**). *Dialects of American English* (Prentice–Hall, Englewood Cliffs, NJ).
Yamada R. A., and Tohkura Y. (**1992a**). ''Perception of English /r/ and /l/ by native speakers of Japanese,'' in *Speech Perception, Production and Linguistic Structure*, edited by Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka (OHM, Tokyo, Japan), pp. 155–174.
Yamada, R. A., and Tohkura, Y. (**1992b**). ''The effect of experimental variables on the perception of American English /r/ and /l/ by Japanese Listeners,'' Percept. Psychophys. **52**, 376–392.

Evidence for an analytic perception of multiharmonic sounds in the bat, *Megaderma lyra*, and its possible role for echo spectral analysis

Katrin Krumbholz[a] and Sabine Schmidt[b]
Zoologisches Institut der Ludwig-Maximilians-Universität München, Luisenstrasse 14, 80333 München, Germany

(Received 8 March 2000; revised 2 January 2001; accepted 8 January 2001)

For echolocation, the gleaning bat *Megaderma lyra* relies on short and broadband calls consisting of multiple harmonic components, each of which is downward frequency modulated. The harmonic components in *M. lyra*'s calls have a relatively small frequency excursion and do not overlap spectrally. Broadband calls of other bat species, on the other hand, often consist of only a few harmonics which are modulated over broad and sometimes overlapping frequency ranges. A call consisting of narrow and nonoverlapping harmonic components may provide a less complete representation of target structure than a call which consists of broadly modulated components. However, a multiharmonic call may help the bats to perceive local spectral changes in the echo from shifts in the peak frequencies of single harmonics, and thereby to extract additional information about the target. To assess this hypothesis, the accuracy with which *M. lyra* can analyze frequency shifts of single partials in multiharmonic complex tones was investigated. A two-alternative, forced-choice behavioral task was used to measure *M. lyra*'s frequency discrimination threshold for the third partial in complex tones whose spectral composition resembled that of the bat's sonar calls. The discrimination threshold for the third partial in a 21.5-kHz harmonic tone amounted to about 2% and was similar to the bat's pure-tone discrimination threshold at 64.5 kHz. Discrimination performance was essentially unaffected by random frequency changes of the other partials and by reducing stimulus duration from 50.5 to 1.5 ms. Both findings are in accordance with predictions made on the basis of the shape of *M. lyra*'s cochlear filters. The comparison between the observed frequency discrimination performance and a computational estimate of the expected frequency shift in the third harmonic of an echo reflected by a simple, two-front target showed that *M. lyra*'s frequency resolution is sufficient for analyzing the target-specific information conveyed by shifts in the peak frequency of single echo components.
[DOI: 10.1121/1.1354198]

PACS numbers: 43.80.Ka, 43.80.Lb, 43.66.Gf, 43.66.Fe [WA]

I. INTRODUCTION

Bats have evolved an active auditory orientation system. They acquire information about their environment by emitting calls and analyzing the returning echoes. By exploiting the time delay between the emission of the call and the return of the echo, bats are able to estimate the distance to the reflecting target (Simmons, 1973). Bats can also identify the shape and the texture of targets (see, e.g., Simmons *et al.*, 1974; Schnitzler and Henson, 1980; Habersetzer and Vogler, 1983; Schmidt, 1988a). A spatially extended target rarely reflects a single echo. Rather, the different surfaces of the target reflect a multitude of partial echoes with slightly differing delays (Schmidt, 1992; Simmons *et al.*, 1995). If the delay differences are shorter than the impulse responses of the relevant auditory filters, the partial echoes temporally merge and interfere in the bat's inner ear. The bat then perceives a complex echo whose spectral composition is changed relative to that of the call (Beuter, 1980). The frequency transfer function of the reflection is determined by the temporal relations between the overlapping partial echoes as well as by their relative levels. A number of behavioral studies have shown that bats actually use the characteristic *spectral* changes in the echoes to identify targets with different *temporal* relations between the partial echoes (Schmidt, 1988b, 1992; Modgans and Schnitzler, 1990; Simmons *et al.*, 1990).

When bats are inspecting the shape and texture of targets, they commonly emit downward frequency-modulated, broadband echolocation calls (Neuweiler, 1990). The use of a broadband call is necessary for the bat to be able to analyze the transfer function of the reflection over a large frequency range. The actual frequency–time structure of the broadband portions of sonar calls, however, differs considerably across different bat species (Simmons *et al.*, 1979). It still remains largely unknown, which *perceptual* qualities may be involved in the analysis of the broadband spectral characteristics of echoes and how the perception of echoes is constrained by the specific frequency-time structure of the echolocation calls.

[a]Author to whom correspondence should be addressed: Centre for the Neural Basis of Hearing, Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, United Kingdom. Electronic mail: katrin.krumbholz@mrc-cbu.cam.ac.uk
[b]Present address: Institut für Zoologie, Tierärztliche Hochschule Hannover, Bünteweg 17, 30559 Hannover, Germany.

The calls of many bat species, for example *Eptesicus fuscus*, consist of one to maximally three harmonic frequency components which are modulated over very broad and often partially overlapping frequency regions (Simmons *et al.*, 1995). The frequency excursion of the first harmonic component in the sonar calls of *E. fuscus* discriminating between targets typically amounts to about 75% of its center frequency. The Megadermatid (Schmidt *et al.*, 2000; Schmidt, 1992) as well as Phyllostomid (Simmons *et al.*, 1979; Thies *et al.*, 1998) bats, on the other hand, rely exclusively on short and broadband echolocation calls which consist of multiple harmonic components, each of which is modulated with a comparatively small frequency excursion. In *Megaderma lyra*, the Indian False Vampire bat, approach-phase calls typically consist of up to six linearly downward-modulated, harmonic components (Schmidt *et al.*, 2000; Schmidt, 1992). The frequency of the first harmonic may sweep from about 23 to 18 kHz. Thus, the frequency excursion of this component corresponds to only about 24% of its center frequency. The individual harmonic components do not overlap spectrally and appear as distinct peaks in the composite frequency spectrum of *M. lyra*'s calls. With the multiharmonic calls, the bat is able to resolve target texture differences in the millimeter range (Schmidt, 1988a). This ability may be a prerequisite for *M. lyra*'s pursuit of ground-dwelling prey. After the bat has detected a prey by passively listening for its faint rustling noises (Fiedler, 1979; Marimuthu and Neuweiler, 1987), the bat closely approaches the prey and remains, up to a few seconds, hovering almost motionless while inspecting the prey's shape with a series of echolocation calls (Schmidt *et al.*, 2000; Marimuthu and Neuweiler, 1987). Thereby, *M. lyra* is able to distinguish between prey and an unpalatable counterfeit that closely resembles the shape of the item of prey (Schmidt *et al.*, 2000). Some nectar-eating Glossophagine bats have been reported to show a similarly remarkable performance in discriminating the shape of flowers (von Helversen and von Helversen, 1999).

What, however, is the adaptive value of *multiharmonic* sonar calls for target structure identification? At first glance, it would seem that a sonar call consisting of a few broadly modulated frequency components would be best suited for analyzing spectral changes in the echo. Such a call would sample the spectral transfer function of the target over wide, coherent frequency regions. *M. lyra*'s multiharmonic calls exhibit pronounced spectral notches between the harmonic components. Thus, each call echo provides a less complete representation of the target transfer function. However, the multicomponent frequency structure of sonar calls may facilitate the perception of *narrow-band* spectral changes in the echoes. Assume that a reflection from a structured target causes a characteristic narrow spectral notch in a restricted frequency region of the sonar call. If the frequency excursion of the relevant harmonic component in the sonar call is not too large with respect to the width of the notch, the notch will change that component's average frequency[1] by cutting out a part of its frequency range (Krumbholz and Schmidt, 1999). If we assume that the components of the sonar call can be perceived separately, the bat may detect the notch as a pitch shift in the respective echo component.

The shift in the average frequency of the component would be larger and thus *more easily perceptible* the smaller the component's frequency excursion in relation to the width of the notch. In a comparative study with human listeners, Krumbholz and Schmidt (1998) measured the detectability and the discriminability of narrow spectral notches in very short (1.018 ms) frequency sweeps. The sweeps were linearly downward frequency modulated with a center frequency of 6 kHz and a frequency excursion of either 3.5 or 5.5 kHz, so their frequency-modulation rates were comparable to the modulation rate of the first harmonic component in *M. lyra*'s sonar calls. The width of the spectral notch was always equal to 810 Hz. In a first experiment, we measured the detection threshold for a notch with a fixed center frequency in terms of notch depth; in the second experiment, we measured the discrimination threshold for the center frequency of a notch with a fixed depth of -60 dB. The minimum depth required to detect a notch in the 3.5-kHz sweep was smaller by about 10 dB than in the 5.5-kHz sweep. Moreover, the discrimination threshold for the notch center frequency was by a factor of 3 smaller for the 3.5-kHz sweep than for the 5.5-kHz sweep.

These data suggest that a narrow spectral notch might be easier to detect and more readily discriminable in the echo of a multicomponent sonar call whose harmonic components are modulated with a small frequency excursion than in the echo of a sonar call which consists of a few broadly modulated components. Krumbholz and Schmidt (1999) showed that a spectral notch caused by the interference of two partial echoes with a delay difference of only a few microseconds can shift the average frequency of the third harmonic component in a typical approach-phase echolocation call of *M. lyra* by as much as 10%. If *M. lyra* could analyze such shifts in the average frequency of individual echo components, the bat might be able to perceive certain narrow-band details in the reflection transfer function that may be imperceptible with a sonar call consisting of broadly modulated components. However, is *M. lyra*'s frequency resolution fine enough to detect small frequency shifts in individual echo components? And, are the bats able to analyze the frequency of any particular echo component independent of the frequency of the other components?

The present study investigates the extent to which *M. lyra* is able to make use of the information conveyed by shifts of the average frequency of single spectral components in the echoes. For that, we examined the accuracy with which this bat species is able to analyze the frequency of a single partial in a multicomponent complex tone with a similar frequency composition as its sonar calls. Further, we investigated whether the perception of the frequency of a single component in *M. lyra*'s sonar calls is independent of the frequencies of the other components by testing how the frequency discrimination performance for a single partial in a complex tone is influenced by random frequency shifts of the other partials. In addition, we examined the effect of stimulus duration on frequency discrimination performance. The experimental frequency discrimination thresholds are compared to predictions that can be made on the basis of existing

behavioral data on the auditory filter shapes in *M. lyra*. Further, *M. lyra*'s frequency discrimination threshold for a partial in a complex tone is compared to a computational estimate of the expected frequency shift in the corresponding echo component caused by a reflection from a simple, two-front target.

II. MATERIALS AND METHODS

A. Experimental setup

All experiments were carried out in a sound-insulated room (2.2×3.4×2.2 m) whose walls were coated with sound-absorbing wedges. The bat's starting position was a perch in the rear upper part of the room. The loudspeaker was mounted at a distance of about 2 m away from and 1.5 m below the starting position and was directed to the bat's head. Two feeding dishes were mounted symmetrically on either side of the loudspeaker. The experimenter was seated about 1 m behind the loudspeaker. The room was dimly illuminated by a 15-W bulb. The transmission characteristics of the setup were examined regularly using a B&K $\frac{1}{4}$-in. microphone and measuring amplifier (type 2610). For calibration, the microphone was mounted at the position of the bat's head when hanging from the perch.

All stimuli were generated digitally on a DSP board (AP2, Tucker Davis Technologies). They were digital-to-analog converted with a 16-bit amplitude resolution and a sampling rate of 250 kHz (DA3-2, Tucker Davis Technologies), and low-pass filtered with a cutoff frequency of 100 kHz (Krohn-Hite 3323). The low-pass-filtered signals were attenuated (Elementa), amplified (Harman/Kardon HK6100), and finally fed to the magnetostatic loudspeaker (Technics EAS 10 TH800C). The frequency transfer function of the loudspeaker was linear within a ±3-dB range between 3 and 100 kHz.

B. General properties of signals

Test stimuli were pure tones or complex tones consisting of four partials. The stimuli were gated with Gaussian on- and offset ramps with a −6-dB duration of 0.332 ms. The steady-state portion of the stimuli had a duration of either 50 or 1 ms. Thus, their equivalent rectangular duration[2] (ERD) amounted to either 50.5 or 1.5 ms. During the initial training, the test stimuli were preceded by a reference tone with a fixed frequency of 64.5 kHz. The reference tone had the same temporal envelope as the respective test stimuli. The silent gap between reference tone and test stimuli was 200 ms (between the 0–V points) and the interpair interval was 700 ms. *M. lyra* is known to have an excellent absolute pitch (Schmidt *et al.*, 1995). Therefore, the reference tone was omitted after the bats had memorized its frequency. All data of the experiments described below were collected without presentation of the 64.5-kHz reference tone. Throughout data collection, the 50.5- and 1.5-ms test stimuli were repeated with a silent gap of 700 and 400 ms and presented at an overall level of 55 and 60 dB SPL (measured at the position of the bat's head), respectively.

C. Procedure

Frequency discrimination thresholds for pure tones (experiment 1) and for the third frequency component (partial) in four-component complex tones (experiments 2 and 3) were measured using a two-alternative, forced-choice behavioral paradigm and the method of constant stimuli. At the beginning of each trial, the experimenter triggered signal presentation via a computer keyboard. The stimuli were presented repetitively until the bat responded by flying either to the right or to the left feeding dish. Signal presentation was stopped after the bat had made its choice and the response was stored in the computer memory. The animals were trained to choose the right side if the frequency of the relevant signal, i.e., the pure tone in experiment 1 and the third partial in the complex tone in experiments 2 and 3, was higher than the reference frequency of 64.5 kHz. Conversely, the bat had to choose the left side if the relevant signal frequency was lower than 64.5 kHz. The bat was rewarded with a mealworm if it had chosen the feeding dish on the correct side. If the bat had chosen the wrong side in response to a certain stimulus, the same stimulus was presented in the following trial. The results of these correction trials were discarded. For a given trial, the experimenter could either choose a specific signal frequency, or invoke the pseudorandom function of the measuring program to pick one of the possible signal frequencies randomly. Signal frequencies above and below the reference were presented with equal probability, and the experiments were double-blind, i.e., the experimenter was only informed about the rewarded side after the bat had made its choice. Moreover, not more than four signal frequencies above, or, respectively, below the reference were presented in a row to avoid the bats developing side preferences.

D. Data analysis

The bats' responses to the test stimuli were expressed as the percentage of choices to the right, P_R, and plotted as a function of the logarithmic frequency separation (expressed in cent[3]), δF, between the respective test frequency and the reference frequency of 64.5 kHz in order to obtain the psychometric function for the frequency discrimination task. In experiments 1, 2 A, and 3, frequency discrimination thresholds were derived from these psychometric functions. For that, sigmoid functions were fitted to the measured P_R values. The sigmoid functions were defined by the general formula[4]

$$P_R(\delta F) = \frac{b}{\exp(-\lambda(\delta F - a)) + 1} + c. \quad (1)$$

The parameters λ and a determine the slope of the function and its displacement relative to the zero point on the axis of frequency separations, δF. These two parameters were fitted to the experimental data using the Gauss–Newton method[5] (Schwarz, 1988). The parameters b and c define the upper and lower asymptotes of the function, respectively, and should ideally be equal to 100% and 0%. However, the bats' responses rarely reached these ideal values even for suprathreshold frequency separations. Thus, b and c were set to

appropriate values between 90% and 100% and 0% and 10%, respectively, and remained fixed during the iterations of the Gauss–Newton fit.[6]

The frequency discrimination thresholds were calculated as the mean of the frequency separations $\delta F_{25} = P_R^{-1}(25\%)$ and $\delta F_{75} = P_R^{-1}(75\%)$ corresponding to the 25% and 75% values on the fitted sigmoid functions. The error of the discrimination thresholds is the square root of the mean of the squared error of δF_{25} and δF_{75}, which was calculated by multiplying the mean deviation between the measured and the fitted P_R values with the slope of the inverse of the fitted function, P_R^{-1}, in the 25% and 75% points, respectively.[7]

E. Animals

Four individuals of the bat species *M. lyra* (one male and three females) took part in the experiments. One female bat only completed the frequency discrimination for pure tones (experiment 1). All four bats had participated in earlier psychoacoustic experiments. The bats were fed on the mealworms that they received as rewards for correct decisions during the experimental sessions. In addition, a small house mouse was supplied every 2 weeks. The bats' weight was monitored daily. The animals were kept in a dimly illuminated flight room at 28 °C, where they had free access to water.

The experiments in this study were noninvasive and therefore did not require an animal experimentation approval. Permission to keep and breed *M. lyra* in fulfillment of Sec. 11, Abs. 1, S. 1, No. 1 Tiersch G has been given to the Zoologisches Institut der Universität München by the Städtisches Veterinäramt der Landeshauptstadt München, dated 6th November 1992.

III. EXPERIMENT 1

A. Stimuli

First, the frequency discrimination threshold for pure tones at 64.5 kHz was measured. The pure tones had an ERD of 50.5 ms. During the initial training, the frequency separations between the test tones and the 64.5-kHz reference were well above the anticipated discrimination threshold. When the bats responded to the suprathreshold test frequencies correctly in more than 90% of the cases, psychometric functions were measured with a set of 14 test frequencies, seven of which were higher and the other seven of which were lower than 64.5 kHz. Now, the frequency separations between test tones and reference comprised both subthreshold and suprathreshold values. Each two corresponding frequencies above and below 64.5 kHz had the same frequency separation, δF, to 64.5 kHz on a logarithmic scale. Figure 1 (experiment 1) schematically shows the spectrograms of two pure tones whose frequencies are by $\delta F = 58.7$ cent lower or higher than 64.5 kHz, respectively. In the first five to ten trials of each experimental session, only the largest frequency separations were presented. In the subsequent trials, the signal frequency was chosen randomly from the set of 14 test frequencies. Thereby, the easier test frequencies were usually chosen with a proportionately higher probability, in order not to discourage the bats with too frequent unrewarded decisions. A minimum of 31 and up to 110 responses were collected for each bat and test frequency to establish the percentage value of choices to the right side, P_R, for that frequency.

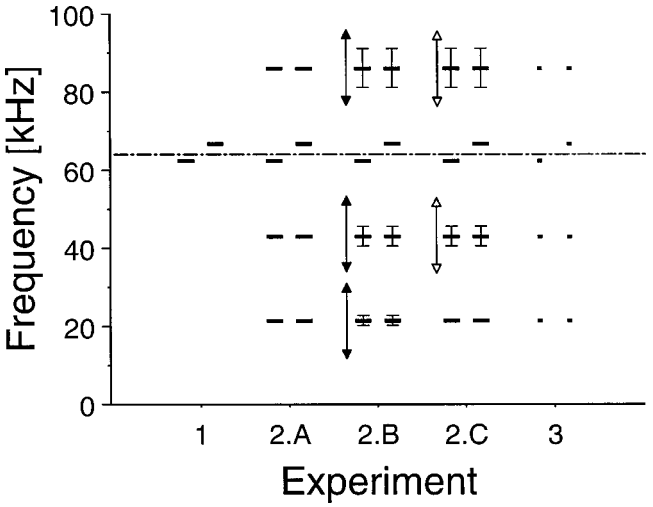


FIG. 1. Spectral composition of experimental stimuli. Each column shows one low and one high test stimulus with the same frequency separation, $\delta F = 58.7$ cent, to 64.5 kHz for each of the experiments in this study. The dash–dotted line represents the reference frequency of 64.5 kHz. The short, solid lines show the presented frequency components. Their lengths indicate the duration of the stimuli (not to scale). The filled arrows indicate that the respective partials were randomized coherently by shifting their fundamental frequency (experiment 2 B); the open arrows indicate that only one partial's frequency was randomly shifted in a given trial (experiment 2 C). The vertical bars delimit the frequency range within which the partials were randomized.

B. Results

After about 1 month of training, the bats had learned to correctly classify pure tones according to frequency relative to the reference tone. The four panels of Fig. 2 show the individual responses of four bats to the 14 test frequencies. The solid lines show the sigmoid functions that were fitted to the P_R values according to Eq. (1). Each of the fitted functions in Fig. 2 accounts for at least 93% of the variance in the experimental data. The individual frequency discrimination thresholds were calculated from the fitted functions as described in Sec. II D and are given in the top left corners of the graphs. The errors are given in brackets. On average, the four bats correctly classified the pure tones in 75% of the cases when the frequency separation from 64.5 kHz amounted to 35.3 (± 4.2) cent, corresponding to a linear frequency difference of 1.32 kHz and a Weber fraction[8] of 2.04%.

IV. EXPERIMENT 2

A. Stimuli

1. Condition A

In experiment 2, three of the four bats that had participated in experiment 1 were trained to discriminate the frequency of the third partial in four-component complex tones. In condition A, the ''irrelevant'' first, second, and fourth partials were harmonics of 21.5 kHz. Thus, their frequencies

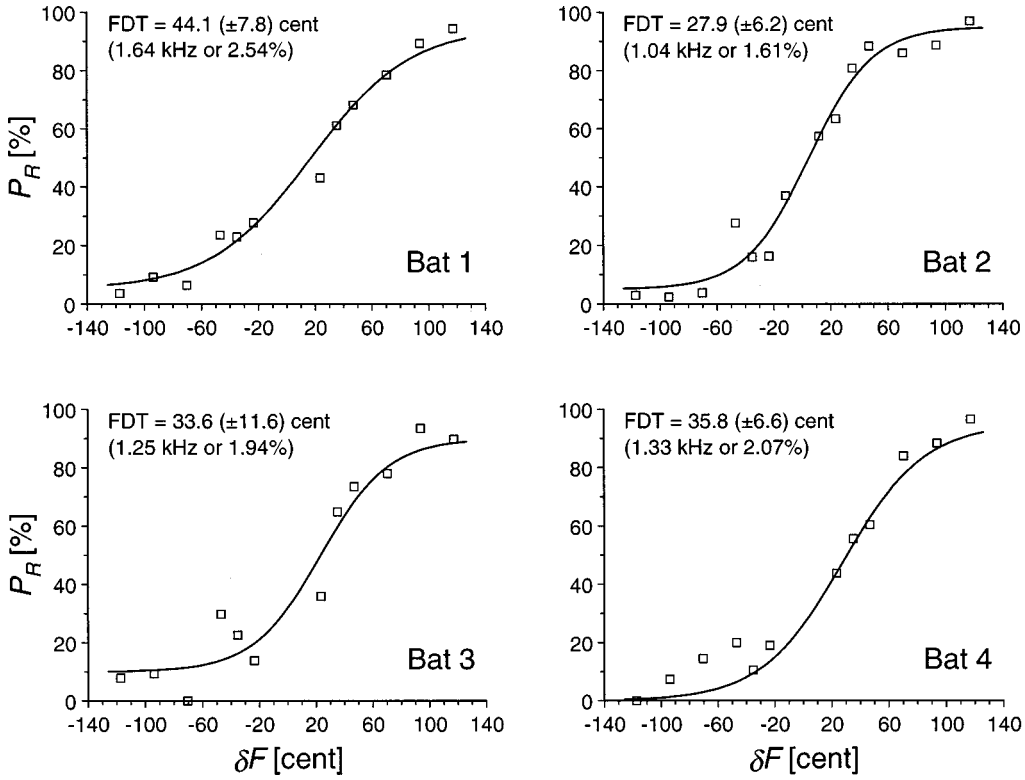


FIG. 2. Frequency classification performance of four bats for 50.5-ms pure tones at 64.5 kHz (experiment 1). The open symbols show the percentage of responses to the right side, P_R, as a function of the logarithmic frequency separation between the respective test frequency and 64.5 kHz, δF. The sigmoid curves (solid lines) were fitted according to Eq. (1). The frequency discrimination threshold (FDT) and the error for each bat is given in the top left corner of the graphs.

were fixed at 21.5, 43, and 86 kHz. The bats had to chose the left or right feeding dish, if the frequency of the ''relevant'' third partial in the presented complex tone was lower or higher than the reference frequency of 64.5 kHz, respectively.[9] Figure 1 (experiment 2 A) schematically shows the spectral composition of two complex tones whose third partial is by 58.7 cent lower or higher than 64.5 kHz. Initially, the levels of the irrelevant partials were attenuated by 25 dB relative to the level of the relevant third partial. During training, their levels were gradually increased in 5-dB steps until the level profile of the four partials in the complex tones approximately corresponded to the average level profile of the harmonic frequency components in *M. lyra*'s typical approach-phase echolocation calls; the first partial was attenuated by 10 dB relative to the second, third, and fourth partial which all had the same level. The first harmonic component in *M. lyra*'s echolocation calls is usually attenuated by at least 10 dB and sometimes by as much as 30 dB relative to the levels of the higher components. The relative levels of the higher components can vary considerably between different calls. A study by Schmidt *et al.* (2000) showed that the levels of the higher components systematically depend on the different phases during a prey pursuit. The complex tones had an ERD of 50.5 ms. In experiment 2 A, psychometric functions were measured with the same set of 14 test frequencies as used for the pure tones in experiment 1. The percentage of choices to the right side, P_R, for each test frequency was established from a minimum of 32 and up to 97 responses from each bat.

2. Condition B

In experiment 2 B, the frequencies of the irrelevant partials were no longer fixed. Their frequencies were still harmonically related but their fundamental frequency was now randomly shifted either above or below 21.5 kHz in each trial. In Fig. 1 (experiment 2 B), which schematically shows two complex tones whose third partial is by 58.7 cent lower or higher than 64.5 kHz, the filled arrows indicate the random frequency variations of the irrelevant partials. The magnitude of the random frequency shifts was uniformly distributed between 25 and 100 cent. The frequency of the relevant third partial was independent of the random frequency shifts of the other partials. The percentage value of choices to the right side, P_R, was determined for a total of eight test frequencies of the third partial, four above and four below 64.5 kHz. Each data point was established from a minimum of 35 and up to 114 presentations of the respective test frequency.

3. Condition C

The irrelevant partials in the complex tones used in conditions A and B of experiment 2 were harmonics of a common fundamental frequency. Only the relevant third partial was mistuned with respect to the harmonic relation of the other three partials. In human listeners, a mistuned partial in an otherwise harmonic complex tone can perceptionally stick out of the tone complex and be perceived as a separate tone (Moore *et al.*, 1985; Hartmann *et al.*, 1990; Demany *et al.*, 1991). Thus, the bats may have possibly heard out the rel-

evant partial as a separate tone in the complex tones used in parts A and B of the experiment, which might have helped them to classify its frequency. In condition C, the frequency relation of the irrelevant first, second, and fourth partials was no longer harmonic: next to the relevant third partial, one of the irrelevant partials was also mistuned from harmony. In a given trial, either the second or the fourth partial was mistuned with equal probability. The frequency of the selected partial was shifted either above or below its nominal harmonic frequency (43 and 86 kHz for the second and fourth partial, respectively). The frequency of the first partial was fixed at 21.5 kHz. The magnitude of the random frequency shifts of the second or fourth partial was uniformly distributed between 30 and 100 cent. In Fig. 1 (experiment 2 C), the open arrows indicate the random frequency variations of the second or fourth partial. The percentage values of choices to the right, P_R, were measured with the same set of eight test frequencies of the third partial as used in experiment 2 B. Each data point resulted from a minimum of 32 and up to 58 presentations of the respective test frequency.

B. Results

None of the bats managed to spontaneously apply the frequency classification of pure tones acquired during experiment 1 to the complex tones used in experiment 2 A. The bats were probably confused by the fact that all of the presented stimuli now contained frequencies both below and above 64.5 kHz. After about 1 month of training, the bats had learned to classify the complex tones according to the frequency of their third partial. Conditions B and C of experiment 2, then, did not require any additional training. The open squares in Fig. 3 show the animals' responses to the complex tones used in experiment 2 A (fixed irrelevant partials) as a function of the logarithmic frequency separation, δF, between the relevant third partial and the reference frequency of 64.5 kHz. The solid lines show the sigmoid curves that were fitted to the psychometric functions from experiment 2 A. The fit curves account for more than 95% of the variance in the data. The three bats' average frequency discrimination threshold for the third partial in a complex tone with fixed irrelevant partials amounted to 36.4 ($\pm$5.1) cent. That corresponds to a linear frequency difference of 1.36 kHz and a Weber fraction of 2.1%. Thus, the mean frequency discrimination thresholds measured in experiments 1 and 2 A are essentially equal. The difference between the individual discrimination thresholds is quite large for bats 1 and 2. However, due to the rather large errors, none of the differences between the individual discrimination thresholds from experiments 1 and 2 A was significant under the assumption that the difference between two thresholds is normally distributed with a variance that is equal to the sum of their squared errors (Krengel, 1990; bat 1: $z=1.211$, $p=0.226$; bat 2: $z=1.657$, $p=0.0975$; bat 3: $z=0.192$, $p=0.848$).

The responses to the complex tones used in conditions B and C of experiment 2 are shown by the filled symbols in Fig. 3. The filled circles show the responses to the complex tones whose irrelevant partials were still harmonic but had a randomized fundamental frequency (condition B). The triangles show the responses to the complex tones in which only either the second or the fourth partial was randomly mistuned from its nominal harmonic frequency (condition C). In the case of bats 1 and 2, the data points from conditions B and C can be accounted for by the same sigmoid function that was fitted to the data points from condition A (open squares). In the case of bat 3, the random frequency changes had a slight detrimental effect on classification performance. However, according to an F test with 6 and 12 degrees of freedom[10] (Linder, 1964), the mean deviation between the fit curve obtained for condition A and the data points from conditions B and C was not significantly greater

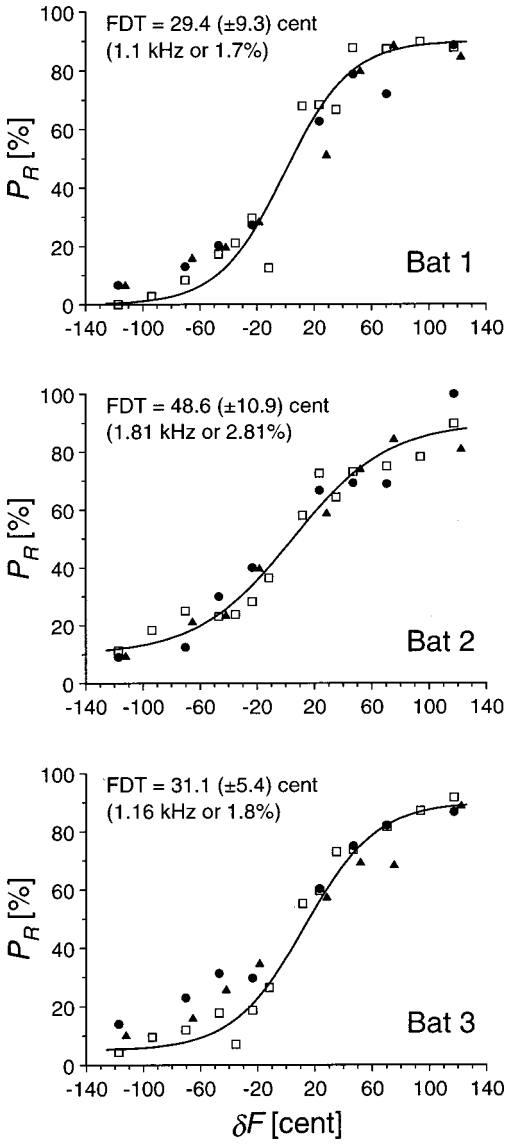


FIG. 3. Frequency classification performance of three bats for the third partial in four-component complex tones with a duration of 50.5 ms (experiment 2), plotted as a function of the frequency separation, δF, between the third partial and 64.5 kHz. The open symbols show the percentage of choices to the right side, P_R, for harmonic tones with a fixed fundamental of 21.5 kHz (condition A). The sigmoid functions (solid lines) were fitted to the data points from condition A. The filled circles show the responses to complex tones with a randomized fundamental frequency (condition B). The filled triangles show the responses to complex tones whose second or fourth partial was randomly shifted out of the harmonic relation (condition C). The individual frequency discrimination thresholds (FDT) for condition A are given in the top left corner of the graphs.

for any of the bats than the deviation between the fit curve and the data points from condition A (condition B, bat 1: $f=1.022$, $p=0.457$; bat 2: $f=1.515$, $p=0.254$; bat 3: $f=2.394$, $p=0.093$; condition C, bat 1: $f=1.133$, $p=0.400$; bat 2: $f=0.787$, $p=0.597$; bat 3: $f=2.476$, $p=0.086$). Thus, neither the presence nor the random frequency changes of the first, second, and fourth partials significantly affected the bats' frequency discrimination thresholds for the third partial.

V. EXPERIMENT 3

A. Stimuli

Experiment 3 was almost identical to experiment 2 A, apart from the different stimulus duration that was used: the bats had to discriminate the frequency of the third partial in a four-component complex tone whose first, second, and fourth partials were harmonics of 21.5 kHz. During the training for experiment 3, the duration of the complex tones was gradually reduced from 50.5 to 1.5 ms. In order to give the stimuli the same energy, the level of the 1.5-ms stimuli would have had to increase by 15 dB relative to the 50.5-ms stimuli. However, in order to avoid an effect of possible level-dependent changes of the cochlear filter characteristics on frequency discrimination performance, the level of the shorter stimuli was only 5 dB higher than the level of the longer stimuli. The psychometric function was measured with a set of 12 test frequencies of the third partial, six above and six below 64.5 kHz. Each data point was established from a minimum of 30 and up to 131 presentations of the respective test frequency.

B. Results

The bats' classifications for the 1.5-ms complex tones are shown in Fig. 4 (open symbols). The fitted sigmoid curves (solid lines) account for more than 93% of the variance in the data. The bats correctly classified the third partial in the 1.5-ms complex tones in 75% of the cases when its frequency separation from 64.5 kHz amounted to an average of 31.8 (± 7.1) cent. That corresponds to a linear frequency difference of 1.18 kHz and a Weber fraction of 1.84%. None of the individual frequency discrimination thresholds from experiment 3 differed significantly from the discrimination thresholds measured in experiment 2 A under the assumption that the difference between two thresholds is normally distributed with a variance that is equal to the sum of their squared errors (bat 1: $z=0.035$, $p=0.972$; bat 2: $z=0.742$, $p=0.458$; bat 3: $z=0.260$, $p=0.795$).

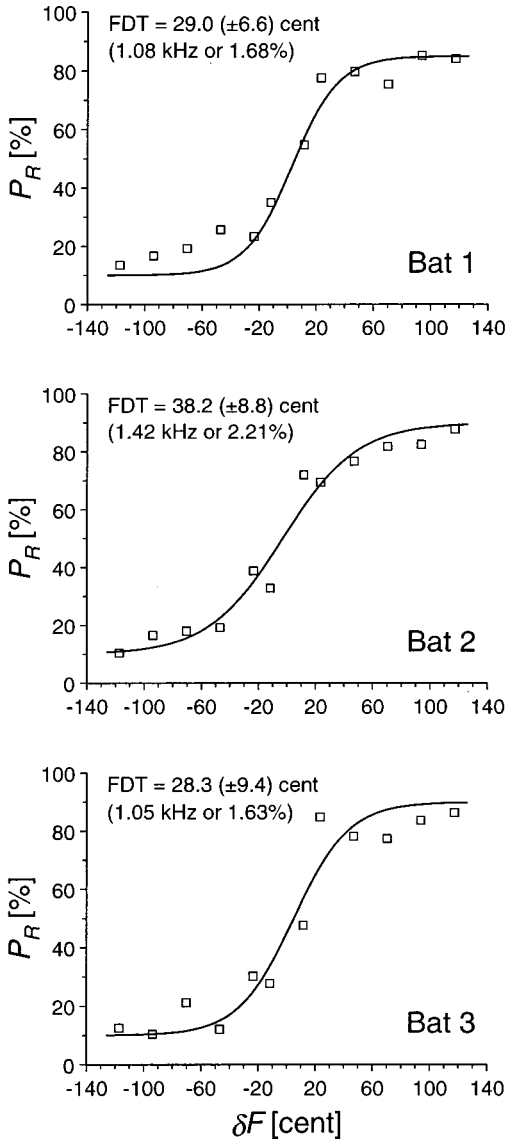


FIG. 4. Frequency classification performance (open symbols) of three bats for the third partial in a complex tone with a fixed fundamental frequency of 21.5 kHz and a duration of 1.5 ms (experiment 3), plotted as a function of the frequency separation, δF, between the third partial and 64.5 kHz. The sigmoid curves (solid lines) were fitted according to Eq. (1). The individual frequency discrimination thresholds (FDT) are given in the top left corner of the graphs.

VI. DISCUSSION

Experiment 1 has shown that *M. lyra*'s relative frequency discrimination threshold for pure tones at 64.5 kHz, with a duration of 50.5 ms and a level of 55 dB SPL, amounts to about 2%. von Stebut (1999) measured *M. lyra*'s pure-tone frequency discrimination threshold at 24 kHz and found a slightly smaller Weber fraction between 1% to 1.2% at this frequency. The difference between von Stebut's and our results might be due to methodical differences; von Stebut used a slightly different experimental paradigm and presented the stimuli at a higher overall level of 70 dB SPL. However, it is more likely that the threshold difference is brought about by the nonlinear frequency dependence of auditory filter shape in *M. lyra*. Schmidt *et al.* (1992, 1995) showed that the width of the critical band is proportional to an expansive power of frequency with an exponent of about 1.71. Moreover, Schmidt *et al.* (1995) found that the slope of the flanks of the critical band is inversely proportional to frequency raised to a power of 1.71. Under the assumption that the absolute frequency discrimination threshold (in Hz) is a constant fraction of the critical bandwidth at the respective frequency, the relative frequency discrimination threshold at 64.5 kHz is expected to be about twice as large as at 24 kHz. This is in good agreement with the observed threshold differences.

M. lyra's pure-tone frequency discrimination threshold at 64.5 kHz is essentially equal to the discrimination threshold for the third partial in a complex tone with a fundamental frequency of 21.5 kHz. Moreover, the frequency classification of the relevant third partial was not severely impaired by random frequency variations of the irrelevant partials (conditions B and C of experiment 2). This shows that the bats had not based their classifications on the overall pitch of the complex tones, and that their perception of the third partial was not significantly influenced by the frequencies of the other partials. The bats' similar performance across the three conditions of experiment 2 suggests that they were able to attend exclusively to the pitch of the relevant partial. Finally, *M. lyra*'s frequency discrimination threshold for the third partial in a complex tone with a fundamental frequency of 21.5 kHz did not deteriorate noticeably when the duration of the complex tones was decreased from 50.5 to 1.5 ms. The similarity between pure-tone frequency discrimination and frequency discrimination in a complex tone and the independence of frequency discrimination performance of stimulus duration down to the millisecond range can be accounted for by the shape of the cochlear filters in *M. lyra*. This is discussed in the next section.

A. Frequency discrimination of a single partial and cochlear filter shape in *M. lyra*

In the auditory system, different frequencies give rise to activity or excitation at different places. The excitation pattern of an auditory stimulus may be defined as the internal distribution of the stimulus-evoked activity as a function of ''place'' [for a detailed discussion of the concept of the excitation pattern see Moore (1997) and Moore and Glasberg (1990)]. Owing to the one-to-one correspondence of place and frequency, the excitation at a given frequency may be considered as the total response of all neurons with characteristic frequencies (CF) equal to the frequency. Moore (1997) suggested measuring excitation physiologically, in terms of the effective input to neurons at a given CF. In psychophysical terms, the shape of the masked audiogram of a given sound reflects the shape of that sound's excitation pattern (Zwicker, 1970). Moore and Glasberg (1983) defined the excitation pattern in terms of the cochlear filter shape, which they derived from psychophysical masking data. Apart from small systematic differences attributable to species differences (Evans and Wilson, 1973; Evans *et al.*, 1992) and fitting technique (Evans, 1995, 2000), psychophysical and physiological measures of frequency selectivity are essentially identical.

According to the place model of pitch, the perception of a change in the frequency of a pure tone or a partial in a complex tone is equivalent to the perception of a shift in the pattern of excitation that the stimulus elicits along the tonotopic axis of the cochlea (Moore *et al.*, 1984; Hesse, 1987; Schorer, 1989). Thus, a frequency change would be just perceptible if it caused a just-perceptible change in the response strength of at least one cochlear filter. The frequency change would be detected at the flanks of the excitation pattern elicited by the stimulus. Thus, the minimum detectable frequency change would depend upon the slope of the steepest portion of the excitation pattern as well as the size of the minimum detectable change in excitation level.

The flanks of the excitation pattern elicited by a single partial in a complex tone can be partially masked by the excitation caused by the adjacent partials. Therefore, human listeners' frequency discrimination thresholds for a single partial in a harmonic complex tone tend to be larger than the respective pure-tone discrimination thresholds (Moore *et al.*, 1984). This is especially true for higher harmonic numbers where the spectral resolution of individual partials is poor. The similarity between *M. lyra*'s frequency discrimination threshold for the third partial in a complex tone with a fundamental frequency of 21.5 kHz (experiment 2 A) and the pure-tone discrimination threshold at 64.5 kHz (experiment 1) shows that the flanks of the third partial's excitation pattern are not substantially masked by the adjacent partials, i.e., the partials must be spectrally well resolved in *M. lyra*'s cochlear filters. This finding is in agreement with existing psychophysical data on the width and the shape of the critical bands in *M. lyra* (Schmidt *et al.*, 1992, 1995).

Schmidt *et al.* (1992, 1995) estimated the width of *M. lyra*'s critical bands by measuring the masked threshold for pure tones in a wideband noise masker as a function of pure-tone frequency. The ratio between the pure tone's power at masked threshold and the noise power density is usually known as the critical ratio (CR). The CR of *M. lyra* as a function of frequency, *f*, is reasonably well described by the analytical expression

$$\mathrm{CR}(f) = 4.9 \cdot (f/50)^{1.71}. \tag{2}$$

At a selected frequency of 64.5 kHz, Schmidt *et al.* (1992, 1995) compared the critical bandwidth (CB) obtained through measurement of the CR with the CB obtained through the method of band widening (see, e.g., Moore, 1997, pp. 90–93) and confirmed that both methods yielded essentially the same estimate for the CB. Schmidt *et al.* (1992) characterized the shape of the critical bands in *M. lyra* by measuring the spectral masking patterns of narrowband noise maskers whose bandwidths approximately corresponded to the CBs at their respective center frequencies. The masking patterns were measured for pure-tone signals within a certain frequency range around the masker passband. Masking patterns were measured at masker center frequencies of 23, 30, 40, and 64.5 kHz and at several different masker levels. Schmidt *et al.* showed that in contrast to the human critical bands, *M. lyra*'s critical bands exhibit symmetric low- and high-frequency flanks with a slope of 40 to 100 dB/CB depending on masker level.

The data on the width and the shape of *M. lyra*'s critical bands were used to calculate the excitation pattern for the 21.5-kHz complex tone used in the present experiments (Fig. 5). The equivalent rectangular bandwidth (ERB) of the excitation patterns associated with each of the individual partials in Fig. 5 corresponds to the CB at the respective harmonic frequency [calculated according to Eq. (2)]. The low- and high-frequency flanks of the excitation patterns are symmetric with a slope of about 50 dB/CB according to the medium presentation levels used in the current experiments. Figure 5 shows that the widths of the excitation patterns elicited by

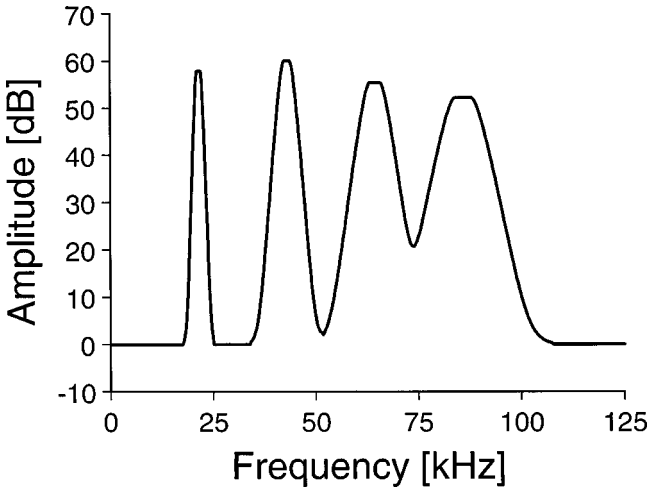


FIG. 5. Excitation pattern of a four-component harmonic complex tone with a fundamental frequency of 21.5 kHz. The excitation pattern was calculated on the basis of the psychophysically measured shapes of the critical bands in *M. lyra* (Schmidt *et al.*, 1992, 1995).

the lower partials are much smaller than the frequency separation between adjacent partials, i.e., 21.5 kHz, and so the lower partials are spectrally almost completely resolved. Accordingly, the frequency of the third partial in the complex tone with a fundamental frequency of 21.5 kHz should be represented with a similar accuracy as the frequency of a pure tone at 64.5 kHz. The frequency discrimination thresholds measured in the present study confirm this expectation.

The width of a pure tone's power spectrum as well as the slope of its flanks are reciprocally related to the tone's duration (see, e.g., Ambardar, 1995). Thus, the flanks of a short tone's power spectrum can be shallower than the flanks of the critical band centered at its frequency. In this case, the slope of the flanks of the tone's excitation pattern would be limited by the slope of the flanks of its power spectrum rather than by the frequency selectivity of the cochlear. The similarity between the frequency discrimination thresholds for the third partial at 64.5 kHz in complex tones with a duration of 1.5 ms (experiment 3) and 50.5 ms (experiment 2 A) implies that the flanks of the partial tone's power spectrum in the 1.5-ms stimulus are still considerably steeper than the flanks of *M. lyra*'s critical band at 64.5 kHz, so that the slope of the flanks of the excitation pattern would still determined by the slope of the flanks of the critical band. The envelope of the 1.5-ms complex tone can be roughly approximated by a rectangular envelope with a duration of 1.5 ms. Apart from a constant factor, the power spectrum of a rectangularly gated tone burst with frequency F and duration t_0 is given by the formula

$$\frac{\sin(\pi t_0(f-F))^2}{(\pi t_0(f-F))^2} \tag{3}$$

(Bronstein and Semendjajew, 1989), where f means frequency. For $t_0 = 1.5$ ms, the flanks of the main peak of this power spectrum have an average slope of about 20 dB/kHz. At an input level of 60 dB SPL, the flanks of *M. lyra*'s critical band at 64.5 kHz have a slope of approximately 50 dB/CB corresponding to 6.6 dB/kHz. Thus, the flanks of the power spectrum of the 1.5-ms partial at 64.5 kHz are still considerably steeper than the flanks of *M. lyra*'s critical band at 64.5 kHz, and so the slope of the flanks of the excitation pattern elicited by the 1.5-ms partial should still be primarily determined by the slope of the flanks of the critical band, and the frequency discrimination thresholds for the 1.5-ms partial should be similar to the discrimination threshold for the 50.5-ms partial, in accordance with the results of experiments 2 A and 3. In human listeners, on the other hand, the frequency discrimination threshold for pure tones increases by a factor of about 10 when the pure-tone duration is decreased from 50 to 5 (Fastl and Hesse, 1984). The difference between the effect of duration on frequency discrimination in humans and in bats is due to the fact that the bat's hearing range spans much higher frequencies than the human hearing range, and so the absolute widths of the cochlear filters are much larger in bats than in humans. Decreasing the duration of a pure tone, and thereby increasing the width of its power spectrum, has a much smaller relative effect if the width of the relevant cochlear filter is large.

B. Influence of random frequency shifts of the "irrelevant" partials

Experiment 2 has shown that M. *lyra*'s frequency discrimination threshold for the third partial in a 21.5-kHz complex tone is essentially unaffected by random frequency changes of the other partials. In contrast, Moore and Glasberg (1990) found that human listeners cannot ignore the frequencies of irrelevant partials when discriminating the frequency of a single partial in complex tones. In their experiment, listeners had to compare the frequency of the lowest partial in two complex tones which had either the same or different fundamental frequencies. The complex tones contained only higher harmonics with harmonic numbers greater than 6. In this case, the frequency of the lowest partial is especially well discriminated and the discrimination of the pitch of the complex tone as a whole is largely determined by the lowest partial. Moreover, the harmonics immediately above the lowest harmonics were omitted so as to further enhance the discriminability of the lowest partial. When the two complex tones whose lowest partials were to be discriminated had the same fundamental frequency, the discrimination threshold amounted to about 0.2% to 0.5%. Thus, in this case, the discrimination threshold for the lowest partial is of the same order of magnitude as the pure-tone discrimination threshold in the relevant frequency region. When, on the other hand, the complex tones had different fundamental frequencies, the discrimination threshold for the lowest partial was up to a factor of 10 larger than in the first case. This shows that the human listeners in Moore and Glasberg's study were unable to attend solely to the pitch of the lowest component, and thus to ignore the differences in the fundamental frequency of the complex tones, even though the lowest component was spectrally reasonably well isolated. Thus, the different performance in *M. lyra* and in humans cannot be attributed to the spectral resolution of partials in the two experiments and may reflect species-specific differences in frequency processing.

C. Implications for spectral echo analysis by *M. lyra*

The experimental results and the considerations of the previous paragraphs show that *M. lyra*'s auditory system is well suited for the perception of brief sounds consisting of multiple frequency components, like the bat's echolocation calls. *M. lyra*'s auditory filters are narrow enough to spectrally resolve the harmonic frequency components in the sonar calls and their echoes. On the other hand, the slopes of *M. lyra*'s auditory filters are not steeper than is needed for analyzing the spectral characteristics of sounds with a duration in the millisecond range. Narrower auditory filters would only facilitate frequency discrimination for signals with long durations. The frequency discrimination for very brief signals, like the frequency components in *M. lyra*'s sonar calls, would not be enhanced, because it would be limited by their broad power spectra with shallow flanks. Due to the reciprocal relation between the width of the peripheral auditory filters and the duration of their impulse response (filter ringing), more narrowly tuned filters would imply a poorer temporal resolution. Thus, *M. lyra* can exploit a maximum of the available auditory information if the widths of the bat's auditory filters are not too narrow with respect to the widths of the power spectra of the brief echolocation calls.

Experiment 3 has shown that *M. lyra* discriminates the frequency of a single partial in a harmonic complex tone which has a similar fundamental frequency and duration as the typical approach-phase echolocation calls with an accuracy of about 2%. The bat can probably discriminate the average frequency of a single frequency component in the echoes with a similar accuracy:[11] a comparative study with human listeners has shown that the frequency discrimination threshold for the average frequency of a very short frequency glide comprising relatively high frequencies (above the phase-locking limit) is comparable to the discrimination threshold for an equally short pure tone at the glide's center frequency (Krumbholz and Schmidt, 1998). On the other hand, Krumbholz and Schmidt (1999) have shown that the reflection from a structured target can shift the average frequency of a single component in the sonar call by up to 10%. The upper panel of Fig. 6 shows two reflection transfer functions that are just discriminable by *M. lyra* (Schmidt, 1992; adapted from Fig. 11 in Krumbholz and Schmidt, 1999). The transfer functions correspond to a perpendicular reflection from a target consisting of two ideally plain surfaces with a distance of 1.3 and 1.5 mm, respectively. The solid line shows an idealized power spectrum of a typical approach-phase echolocation call of *M. lyra* (Schmidt *et al.*, 2000; Schmidt, 1992). The call consists of four linearly downward-modulated, harmonic components with relative levels of −25, −15, 0, and −9 dB, respectively. The first harmonic component has a frequency excursion of 6 kHz around a center frequency of 20 kHz. The two lower panels of Fig. 6 show the power spectra of the respective echoes, obtained by multiplying the power spectrum of the call with the two transfer functions. The two vertical lines indicate the respective average frequencies of the third spectral component in the echoes. The average frequency for the 1.3-mm target is smaller by 6.9 kHz than for the 1.5-mm target because the

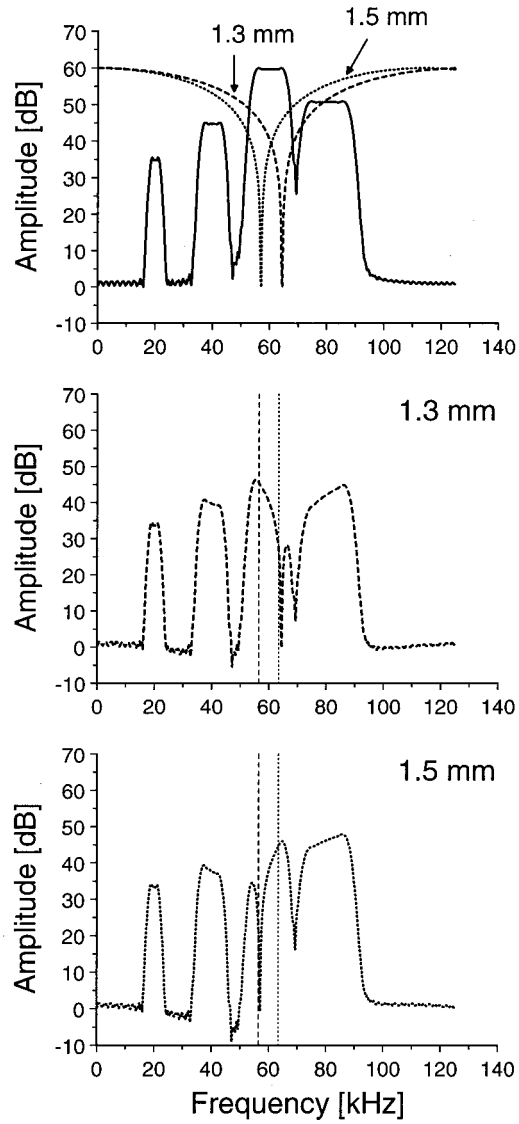


FIG. 6. Upper panel: two just-discriminable reflection transfer functions corresponding to targets consisting of two ideally flat surfaces with a distance of 1.3 (dashed line) and 1.5 mm (dotted line). The solid line shows an idealized frequency spectrum of a typical approach-phase sonar call of *M. lyra* (adapted from Krumbholz and Schmidt, 1999). Middle and lower panels: frequency spectra of the resulting echoes. The vertical dashed and dotted lines denote the spectral center of gravity of the third frequency component in the 1.3-mm echo and the 1.5-mm echo, respectively.

respective notch in its transfer function attenuates higher frequency portions of the third harmonic of the call. The difference between the average frequencies in Fig. 6 amounts to about 10%. According to the results of the present experiments, this difference should be well discriminable by *M. lyra*. Clearly, the situation depicted in Fig. 6 is highly idealized. In the real echolocation situation, the derived difference between the average frequencies would be partially blurred by a number of factors, such as environmental noise. However, the example gives an impression of the magnitude of the realistic frequency shifts that can be caused by a reflection from a structured or spatially extended target. It also shows that *M. lyra*'s frequency discrimination performance is sufficient to enable the bats to analyze the target-specific information conveyed by these frequency shifts.

A reflection from a structured target also changes the

relative levels of the harmonic components in the echo. Adjacent components in *M. lyra*'s calls have a frequency separation of about 20 kHz. Thus, the relative levels of the harmonic components are determined by the relations between widely separated frequency regions of the reflection transfer function. The average frequencies of the echo components, on the other hand, are determined by the local characteristics of the transfer function in the frequency regions covered by the harmonic components in the sonar call. Thus, *M. lyra* could acquire complementary information about the transfer function by evaluating the reflection-induced changes in the relative levels, and the average frequencies of the frequency components in the echoes: the relative levels mediate information about the broadband characteristics of the transfer function, whereas the average frequencies convey information about its local features in the frequency regions covered by the individual components. The analysis of frequency shifts of the echo components would therefore not only enable *M. lyra* to perceive some of the details in the transfer function that would not be detectable with a call consisting of a few broadly frequency-modulated components; it would also enable the bat to probe the transfer function independently in different frequency regions. The results of conditions B and C of experiment 2 show that the perceptual requirement for such a frequency-specific analysis of the reflection transfer function is met: *M. lyra* can analyze the frequency of a single spectral component in a multicomponent sound, like the bat's echolocation calls and the call echoes, independent of the frequencies of the other components. Thus, each echo component can convey independent information about target structure.

The average frequencies of the spectral components in the echoes, together with their broadband level profile, define the spectral signature of the reflecting target. *M. lyra* may analyze a target by comparing the spectral features of the reflected echoes with previously learned patterns. The bat could identify target structure on the basis of typical spectral features which are characteristic of certain geometrical aspects of the target. Due to the multicomponent structure of the sonar calls, the local and the global spectral characteristics of the reflection transfer function may be translated into different perceptual qualities, mediated by the average frequencies and the overall level profile of the frequency components in the echoes.

ACKNOWLEDGMENTS

This study was supported by the Deutsche Forschungsgemeinschaft, SFB 204 ''Gehör,'' München. The authors would like to thank Professor G. Neuweiler for his support.

[1]The intensity-weighted average of the instantaneous frequency of a frequency-modulated signal, shortly referred to as the average frequency, is equal to the signal's spectral center of gravity (Anantharaman and Krishnamurthy, 1993).

[2]The ERD of a stimulus gated with a smooth envelope is the duration of a rectangular envelope which has the same maximum amplitude and the same energy as the smooth envelope (see Zwicker and Feldkeller, 1967).

[3]It has proven useful to express frequency separations on a logarithmic ratio scale comparable to the decibel scale for sound intensity, because subjective relations between musical notes usually correspond to frequency ratios instead of linear frequency differences. The ratio measure of the separation between two frequencies f_1 and f_2 is defined by $\delta F = \log_2(f_2/f_1)\text{oct} = 12 \cdot \log_2(f_2/f_1)\text{st} = 1200 \cdot \log_2(f_2/f_1)$ cent, whereby $\log_2$ is the base 2 logarithm (e.g., Terhardt, 1998). Thus, 1 octave (oct) corresponds to a frequency ratio of $f_2/f_1 = 2$ which is equal to 12 semitones (st) or 1200 cent.

[4]The data were fitted with a sigmoid instead of a cumulative Gaussian function because it was computationally simpler to fit an analytic instead of a tabulated function and the difference between the derived threshold estimates was anticipated to be insignificant. We also tried converting the P_R values to *z* scores (Macmillan and Creelman, 1991) and fitting them with a linear regression line. However, the thresholds derived from the linear fits tended to be slightly worse than the threshold estimates from the sigmoid fits and their errors were larger, probably due to the large variance of the *z* scores for P_R values near 0 or 100%.

[5]The Gauss–Newton method is one of the two existing numerical methods for approximating the solution of a redundant system of nonlinear equations according to the least-squares principle, in which the system of nonlinear equations is approximated with a linear system in each iterative step.

[6]The parameters *b* and *c* were selected ''by eye.'' Their exact values did not affect the thresholds estimates to any significant degree.

[7]The mean deviation between the measured and the fitted P_R values, *s*, was calculated according to the formula $s = \sqrt{\Sigma_1^N (P_R^{\text{meas}} - P_R^{\text{fit}})^2/\text{df}}$, where *N* is the number of test frequencies and df is the number of the degrees of freedom. df was set to the number of test frequencies minus the number of parameters that were fitted with the data, i.e., $\text{df} = N - 2$.

[8]The Weber fraction is defined as the ratio dF/F between the just discriminable linear frequency difference dF and frequency F.

[9]64.5 kHz is also the nominal harmonic frequency of the third partial in a harmonic complex tone with a fundamental frequency of 21.5 kHz.

[10]The degrees of freedom, df_1 and df_2, of the *F* distribution were set to the number of test frequencies used in condition A ($N_1 = 14$) and in condition B or C ($N_2 = 8$) minus the number of parameters that were fitted with the data, i.e., $\text{df}_i = N_i - 2$ for $i = 1,2$.

[11]So far, conclusive behavioral experiments revealing differences in frequency discrimination during echolocation versus passive listening are still missing. (Siewert *et al.*, 2000) measured detection thresholds for probe stimuli whose presentation was triggered by the emission of sonar calls and found a pronounced forward masking of the probe by the preceding sonar call. If we assume that frequency discrimination in bats improves with increasing sensation level, frequency discrimination may deteriorate somewhat in an echolocation situation with short echo delays.

Ambardar, A. (**1995**). *Analog and Digital Signal Processing*, Foundations in Engineering Series (PWS, Boston).

Anantharaman, J. N., and Krishnamurthy, A. K. (**1993**). ''Intensity-weighted average of instantaneous frequency as a model for frequency discrimination,'' J. Acoust. Soc. Am. **94**, 723–729.

Beuter, K. J. (**1980**). ''A new concept of echo evaluation in the auditory system of bats,'' in *Animal Sonar Systems*, edited by R.-G. Busnel and J. F. Fish (Plenum, New York), pp. 747–761.

Bronstein, I. N., and Semendjajew, K. A. (**1989**). *Taschenbuch der Mathematik* (Harry Deutsch, Thun).

Demany, L., Semal, C., and Carlyon, R. P. (**1991**). ''On the perceptual limits of octave harmony and their origin,'' J. Acoust. Soc. Am. **90**, 3019–3027.

Evans, E. F. (**1995**). ''Comparison between physiological and behavioural frequency selectivity in the guinea pig: Effect of fitting Gaussian and rounded exponential filter shapes to the physiological data,'' Br. J. Audiol. **29**, 44.

Evans, E. F. (**2000**). ''Latest comparisons between physiological and behavioural frequency selectivity,'' in *Physiological and Psychophysical Bases of Auditory Function*, edited by A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven (Shaker BV, Maastricht) (in press).

Evans, E. F., Pratt, S. R., Spenner, H., and Cooper, N. P. (**1992**). ''Comparisons of physiological and behavioural properties: auditory frequency selectivity,'' in *Auditory Physiology and Perception*, edited by Y. Cazals, K. Horner, and L. Demany (Pergamon, Oxford), pp. 159–169.

Evans, E. F., and Wilson, J. P. (**1973**). ''Frequency selectivity of the cochlea,'' in *Basic Mechanisms in Hearing*, edited by A. R. Møller (Academic, New York), pp. 519–551.

Fastl, H., and Hesse, A. (**1984**). ''Frequency discrimination of pure tones at short durations,'' Acustica **56**, 41–47.

Fiedler, J. (**1979**). ''Prey catching with and without echolocation in the Indian False Vampire (*Megederma lyra*),'' Behav. Ecol. Sociobiol. **6**, 155–160.
Habersetzer, J., and Vogler, B. (**1983**). ''Discrimination of surface-structured targets by the echolocating bat *Myotis myotis* during flight,'' J. Comp. Physiol. **152**, 275–282.
Hartmann, W. M., McAdams, S., and Schmith, B. K. (**1990**). ''Hearing a mistuned harmonic in an otherwise periodic complex tone,'' J. Acoust. Soc. Am. **88**, 1712–1724.
Hesse, A. (**1987**). ''A model of spectral pitch of pure tones,'' Acustica **63**, 1–16.
Krengel, U. (**1990**). *Einführung in die Wahrscheinlichkeitstheorie und Statistik* (Vieweg, Braunschweig).
Krumbholz, K., and Schmidt, S. (**1998**). ''The role of the multi-component structure of sonar calls for the spectral analysis of echoes in the bat, *Megaderma lyra*,'' in *Göttingen Neurobiology Report 1998*, edited by N. Elsner and R. Wehner (Georg Thieme, Stuttgart).
Krumbholz, K., and Schmidt, S. (**1999**). ''Perception of complex tones and its analogy to echo spectral analysis in the bat, *Megaderma lyra*,'' J. Acoust. Soc. Am. **105**, 898–911.
Linder, A. (**1964**). *Statistische Methoden* (Birkhäuser, Basel).
Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press, Cambridge).
Marimuthu, G., and Neuweiler, G. (**1987**). ''The use of acoustical cues for prey detection by the Indian False Vampire Bat, *Megaderma lyra*,'' J. Comp. Physiol. **160**, 509–515.
Modgans, J., and Schnitzler, H.-U. (**1990**). ''Range resolution and the possible use of spectral information in the echolocating bat, *Eptesicus fuscus*,'' J. Acoust. Soc. Am. **88**, 754–757.
Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing* (Academic, San Diego).
Moore, B. C. J., and Glasberg, B. R. (**1983**). ''Suggested formulas for calculating auditory-filter bandwidths and excitation patterns,'' J. Acoust. Soc. Am. **74**, 750–753.
Moore, B. C. J., and Glasberg, B. R. (**1990**). ''Frequency discrimination of complex tones with overlapping and nonoverlapping harmonics,'' J. Acoust. Soc. Am. **87**, 2163–2177.
Moore, B. C. J., Glasberg, B. R., and Shailer, M. J. (**1984**). ''Frequency and intensity difference limens for harmonics within complex tones,'' J. Acoust. Soc. Am. **75**, 550–561.
Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1985**). ''Thresholds for the detection of inharmonicity in complex tones,'' J. Acoust. Soc. Am. **77**, 1861–1967.
Neuweiler, G. (**1990**). ''Auditory adaptations for prey capture in echolocating bats,'' Physiol. Rev. **70**, 615–641.
Schmidt, S. (**1988a**). ''Discrimination of target surface structure in the echolocating bat, *Megaderma lyra*,'' in *Animal Sonar Systems*, edited by P. E. Nachtigall and M. P. W. B. (Plenum, New York), pp. 507–512.
Schmidt, S. (**1988b**). ''Evidence for a spectral basis of texture perception in bat sonar,'' Nature (London) **331**, 617–619.
Schmidt, S. (**1992**). ''Perception of structured phantom targets in the echolocating bat, *Megaderma lyra*,'' J. Acoust. Soc. Am. **91**, 2203–2223.
Schmidt, S., Gerber, P., and Pötke, M. (**1992**). ''Psychoacoustical characterization of auditory filter shape in the bat, *Megaderma lyra*,'' in *Abstracts of the 15th annual meeting of the European Neuroscience Association*, Supplement No. 5, Eur. J. Neurosci., 141.
Schmidt, S., Hanke, S., and Pillat, J. (**2000**). ''The role of echolocation in the hunting of terrestrial prey—new evidence for an underestimated strategy with the gleaning bat, *Megaderma lyra*,'' J. Comp. Physiol. A (in press).
Schmidt, S., Preisler, A., and Sedlmeier, H. (**1995**). ''Aspects of pitch perception in the ultrasonic range,'' in *Advances in Hearing Research*, edited by G. A. Manley, G. Klump, C. Köppl, H. Fastl, and H. Oeckinghaus (World Scientific, Singapore), pp. 374–382.
Schnitzler, H.-U., and Henson, O. W. (**1980**). ''Performance of airborne animal sonar systems. I. Microchiroptera,'' in *Animal Sonar Systems*, edited by R.-G. Busnel and J. F. Fish (Plenum, New York), pp. 695–714.
Schorer, E. (**1989**). ''Ein Funktionsschema eben wahrnehmbarer Frequenz- und Amplitudenänderungen,'' Acustica **68**, 268–287.
Schwarz, H. R. (**1988**). *Numerische Mathematik* (B. G. Teubner, Stuttgart).
Siewert, I., Schillinger, T., and Schmidt, S. (**2000**). ''Forward masking and the consequences on echo perception in the gleaning bat, *Megaderma lyra*,'' in *Echoloaction in Bats and Dolphins*, edited by J. Thomas, C. F. Moss, and M. Vater (University of Chicago Press, Chicago) (in press).
Simmons, J. A. (**1973**). ''The resolution of target range by echolocating bats,'' J. Acoust. Soc. Am. **54**, 157–173.
Simmons, J. A., Fenton, M. B., and O'Farell, M. (**1979**). ''Echolocation and pursuit of prey by bats,'' Science **203**, 16–21.
Simmons, J. A., Ferragamo, M. J., Saillant, P. A., Haresign, T., Wotton, J. M., Dear, S. P., and Lee, D. N. (**1995**). ''Auditory dimensions of acoustic images in echolocation,'' in *Hearing by Bats*, edited by R. N. Popper and R. R. Fay (Springer, New York), pp. 146–190.
Simmons, J. A., Lavender, W. A., Lavender, B. A., Doroshaw, C. A., Kiefer, S. W., Livingston, R., Scallet, A. C., and Crowley, D. E. (**1974**). ''Target structure and echo spectral discrimination by echolocating bats,'' Science **186**, 1130–1132.
Simmons, J. A., Moss, C. F., and Ferragamo, M. (**1990**). ''Convergence of temporal and spectral information into acoustic images of complex sonar targets perceived by the echolocating bat, *Eptesicus fuscus*,'' J. Comp. Physiol. A **166**, 449–470.
Terkhardt, E. (**1998**). *Akustische Kommunikation* (Springer, Berlin).
Thies, W., Kalko, E. K. V., and Schnitzler, H.-U. (**1998**). ''The role of echolocation and olfaction in two Neotropical fruit-eating bats *Carollia perspicillata* and *castanea*, feeding on Piper,'' Behav. Ecol. Sociobiol. **42**, 397–409.
von Helversen, D., and von Helversen, O. (**1999**). ''Acoustic guide in bat-pollinated flower,'' Nature (London) **398**, 759–760.
von Stebut, N. B. (**1999**). ''Spektrale Muster als Informationsträger: spezifische Übertragungsleistungen der Echoortungslaute von *Eptesicus fuscus* und *Megaderma lyra*,'' Doctoral dissertation, Ludwig-Maximilians-Universität München, Germany.
Zwicker, E. (**1970**). ''Masking and psychological excitation as consequences of the ear's frequency analysis,'' in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden).
Zwicker, E., and Feldkeller, R. (**1967**). *Das Ohr als Nachrichtenempfänger* (S. Hirzel, Stuttgart).

Assessment of dolphin (*Tursiops truncatus*) auditory sensitivity and hearing loss using jawphones

Randall L. Brill[a] and Patrick W. B. Moore
Biosciences Division, SPAWAR Systems Center San Diego, Code D351, 49620 Beluga Road, San Diego, California 92152-6506

Lois A. Dankiewicz
Biosciences Division, SPAWAR Systems Center San Diego, Code D351, 49620 Beluga Road, San Diego, California 92152-6506 and Science Applications International Corporation, Maritime Services Division, 10260 Campus Point Drive, San Diego, California 92121-1578

(Received 14 February 2000; revised 22 January 2001; accepted 24 January 2001)

Devices known as jawphones have previously been used to measure interaural time and intensity discrimination in dolphins. This study introduces their use for measuring hearing sensitivity in dolphins. Auditory thresholds were measured behaviorally against natural background noise for two bottlenose dolphins (*Tursiops truncatus*); a 14-year-old female and a 33-year-old male. Stimuli were delivered to each ear independently by placing jawphones directly over the pan bone of the dolphin's lower jaw, the assumed site of best reception. The shape of the female dolphin's auditory functions, including comparison measurements made in the free field, favorably matches that of the accepted standard audiogram for the species. Thresholds previously measured for the male dolphin at 26 years of age indicated a sensitivity difference between the ears of 2–3 dB between 4–10 kHz, which was considered unremarkable at the time. Thresholds for the male dolphin reported in this study suggest a high-frequency loss compared to the standard audiogram. Both of the male's ears have lost sensitivity to frequencies above 55 kHz and the right ear is 16–33 dB less sensitive than the left ear over the 10–40 kHz range, suggesting that males of the species may lose sensitivity as a function of age. The results of this study support the use of jawphones for the measurement of dolphin auditory sensitivity. [DOI: 10.1121/1.1356704]

PACS numbers: 43.80.Ka, 43.80.Jz, 43.66.Gf [WA]

I. INTRODUCTION

Johnson (1967) produced the first detailed behavioral auditory thresholds for a bottlenose dolphin (*Tursiops truncatus*) that have become the standard for the species (Fig. 1). His subject was a 10-year-old male dolphin trained to detect pure tone, free-field stimuli using an up/down staircase procedure (cf. Moore and Pawloski, 1990) and a ''go/no-go'' response paradigm (cf. Brill *et al.*, 1992). Based on Johnson's results, bottlenose dolphins are generally assumed to have a functional hearing range of 100 Hz to 150 kHz, above which sensitivity falls off at 495 dB per octave, with the best sensitivity between 15 and 110 kHz. Audiograms for several other small odontocete cetaceans show similar characteristics, though best frequencies vary across species (Andersen, 1971; Hall and Johnson, 1972; Jacobs and Hall, 1972; White *et al.*, 1978; Thomas *et al.*, 1988).

The hypothesis that the dolphin's lower jaw is integrally involved in the reception of high-frequency biosonar signals has also become generally accepted. Norris (1964, 1968, 1969) first suggested that the flared posterior ends of the dolphin's lower jaw, specifically the areas over the characteristically thin pan bones, were primary reception sites for returning echolocation signals rather than the external auditory meatus and auditory canal. His model hypothesized that specialized fat bodies filling the lower jaw and connecting to the tympanoperiotic bone provided an efficient acoustic pathway from the pan bone to the tympanoperiotic bone. Electrophysiological data reported by Bullock *et al.* (1968) and McCormick *et al.* (1970, 1980) support Norris' theory. Varanasi and Malins (1971, 1972) reported that the fat found in the lower jaw and melon, which they called ''acoustic fat,'' was composed of lipids with densities close to sea water, which would be expected to facilitate sound conduction. By applying closed-cell neoprene rubber over the lower jaw of their dolphin, Brill *et al.* (1988) behaviorally demonstrated that the attenuation of returning signals at the lower jaw of an actively echolocating dolphin prevented the animal from successfully detecting targets.

Given the evidence for sound reception at the dolphin's lower jaw, Moore and Pawloski (1993) designed and introduced the use of ''jawphones.'' To test their dolphin's use of interaural time and intensity cues, they embedded small transducers in suction cups formed from degassed silicone rubber. Using these jawphones, they obtained evidence of interaural time and intensity discrimination in a dolphin superior to the performance of any mammal previously measured (Moore *et al.*, 1995). We measured baseline auditory thresholds masked by the ambient noise of San Diego Bay on two dolphin subjects: a young female and an older male. Using jawphones placed over the pan bone to deliver the acoustic stimuli, masked thresholds for the left and right ears (referred to as ''jawphone thresholds'') were independently measured for each animal. For comparison, we also measured masked free-field thresholds for the female dolphin.

[a]Electronic mail: brill@spawar.navy.mil

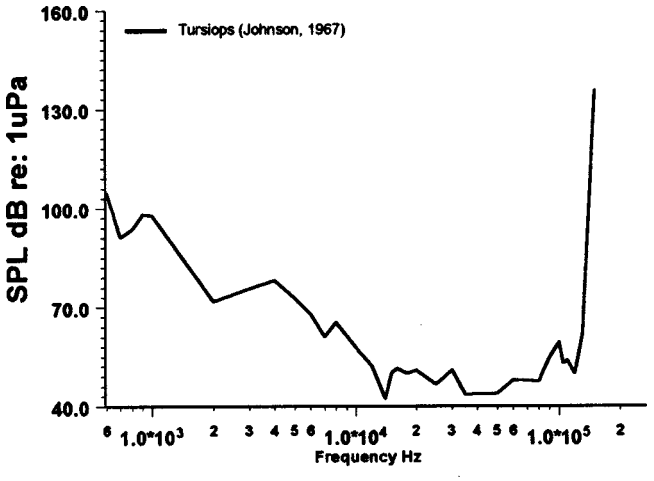


FIG. 1. The standard audiogram for the bottlenose dolphin, *Tursiops truncatus* (after Johnson, 1967).

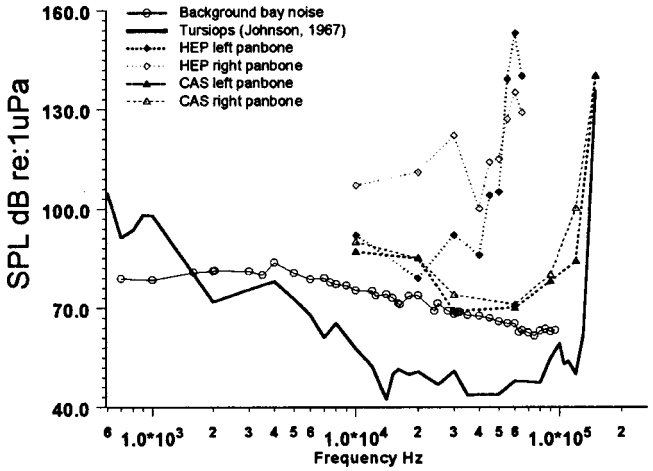


FIG. 2. Ambient noise measured at the test site in San Diego Bay, jawphone thresholds measured for the male dolphin (HEP), and jawphone and free-field thresholds measured for the female dolphin (CAS) are compared to the standard.

II. MATERIALS AND METHODS

A. Subjects

The subjects were two Atlantic bottlenose dolphins (*Tursiops truncatus*), a 14-yr-old female, and a 33-yr-old male, housed and tested in a floating pen complex located in San Diego Bay, San Diego, California. Medical records for both animals suggested that their hearing capabilities were normal. The female (CAS) had not been tested previously. The male (HEP) is the same animal that participated in the study of free-field sound localization reported by Renaud and Popper (1975) as well as the studies of interaural capabilities conducted by Moore and Pawloski (1993) and Moore *et al.* (1995). No gross hearing abnormalities were noted for the male in the earlier studies.

B. Ambient noise in San Diego Bay

Au *et al.* (1985) described the ambient noise in San Diego Bay as having a relatively constant spectral level from 10–60 kHz averaging 40 dB *re*: 1 μPa^2/Hz with a sharp decline above 60 kHz. For this study, ambient noise levels at our test site in San Diego Bay were measured using a B&K 8103 hydrophone. The hydrophone output was amplified and filtered from 2 Hz to 200 kHz using a B&K 2635 charge amplifier and input to a HP3561A digital signal analyzer. A Hanning window was employed to reduce spectral leakage. The noise spectral density shown in Fig. 2 was the result of 40 averages; the frequency bandwidth is 375 Hz.

The noise level in San Diego Bay has apparently increased over the years; most likely due to increased biological load, boat traffic, and other activities. While the spectral levels are still relatively flat over the range of 1 to 90 kHz, they gradually roll off from 80 to 65 dB.

C. Equipment and procedure for jawphone measurements

The basic construction of a jawphone is illustrated in Fig. 3. Three different jawphones were used to test the animals' threshold limits. Each contained only one transducer and was constructed by hand from a room temperature vulcanizing (RTV) silicone rubber compound (Visilox V-1063) manufactured by the Loctite Corporation which has a rho-c value close to that of water. The uncured mixture was degassed under a vacuum of 30 in. of Hg or greater for 30 min to remove any air bubbles introduced during the mixing process. Once cured, the material had a tensile strength of 550 psi and a specific gravity of 1.1.

For the 10 and 20 kHz test frequencies, a small subminiature earphone element encapsulated in an air-filled chamber inside one jawphone was used as the stimulus transducer. For frequencies of 30, 60, 90, and 120 kHz, an Edo Western 6600, 10-mm spherical transducer embedded in a second jawphone was used to transmit the test tones. A third jawphone contained a Bruel & Kjaer (B&K) 8103 miniature hydrophone used at 120 kHz as a comparison to the Edo Western 6600. Only the B&K 8103 was used for testing at 150 kHz. Combined, the three jawphones provided an effective sound source range of 10–150 kHz.

Each of the jawphones was wrapped with closed-cell neoprene to restrict sound transmission from any direction other than the suction cup end. This had the effect of limiting the area of ensonification to the area of the device that at-

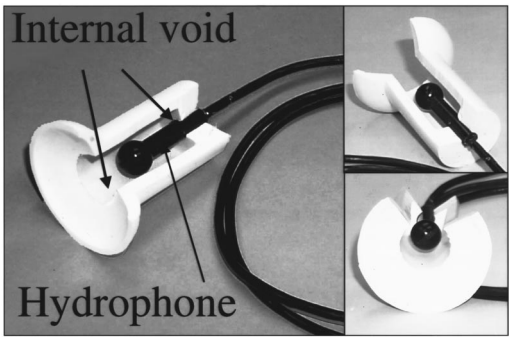


FIG. 3. A cutaway view of a jawphone. The cylindrical internal void is filled with degassed Visolox (the same material from which the suction cup is made) and combines the hydrophone and suction cup into a solid unit.

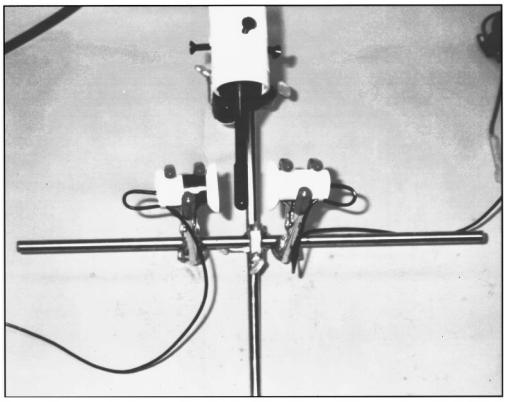


FIG. 4. An armature was used to hold jawphones equidistant from the measuring hydrophone for calibration while submerged in a tank filled with artificial seawater.

tached to the animal. Because this also affected the transmitting response due to the closed sides of the jawphones, the devices were measured exactly at each test frequency along the maximum response axis, which was the long axis of the jawphones terminating in the suction cup attachment (the direction of sound conduction into the dolphin's head).

After potting the transmitting element, each jawphone was calibrated for each specific test frequency for which it would be used. These elements have a nonuniform transmitting response across the frequency range used, so a precise measurement of the transmitting frequency was required. This was especially true for the earphone element encapsulated in the air-filled chamber. Each jawphone was calibrated using a special calibration armature (Fig. 4) that secured both the jawphone and a measuring hydrophone (Edo Western 6660 for 150 kHz and an ITC-6069 for 10–120 kHz). A fixed separation distance of 430 mm was set between the jawphone and the measuring hydrophone acoustic axes. The two were directly in line with both the long axis of the jawphone being calibrated and the transmitting element. This same orientation occurred when the jawphone was attached to the animal's head. The armature was submerged in a tank of water. Measured levels were then used to calculate the animals' thresholds. Both dolphins were trained to accept a jawphone over either pan bone of their lower jaw (Fig. 5). Consistent placement from session to session was achieved

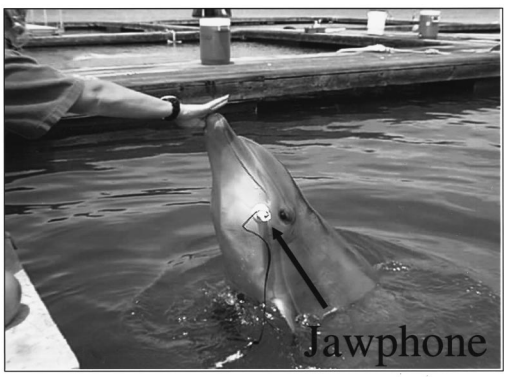


FIG. 5. The jawphone positioned over the dolphin's left pan bone or ''acoustic window.''

by the use of external anatomical landmarks on the individual dolphin and photographic records made for comparison. The animals could dislodge the device at will.

Pure tone stimuli were produced by a BK Precision 5 MHz, Model 4011 Function Generator. The output of the function generator was fed to an RLC Electronics model AT201 attenuator and then to custom equipment built at our laboratory that controlled stimulus duration (1 s) and rise/fall time (20 ms to eliminate onset/offset transients). This shaped output was fed directly to an additional attenuator (either a Hewlett-Packard model 350D or a Pasternack Enterprises PE 7008-2, depending on the sound source level during testing) and then to a jawphone. A Hafler DH-2200 stereo amplifier was used for mid- to high-range (60–150 kHz) test frequencies.

Similar to the procedures used by Johnson (1967), the estimated jawphone thresholds for our dolphins were measured using a derivation of an up/down staircase stimulus presentation method (see Moore and Schusterman, 1987) and a ''go/no-go'' response paradigm (see Brill *et al.*, 1988). A trial began with the dolphin wearing a jawphone and stationed in front of the trainer. On the presentation of a hand cue, the dolphin moved to an underwater test station positioned 50 cm below the surface, where it was required to remain for two seconds before the stimulus was presented. If a tone was presented within the following three seconds (the ''go'' condition), the dolphin left its station immediately to touch a nearby response paddle just above the water's surface. If no tone was presented (the ''no-go'' condition), it remained at its station until a bridging stimulus was delivered and the dolphin returned to the trainer. All correct responses were reinforced with a fish reward. The trial type was determined by a Gellerman series (Gellerman, 1933) modified to produce a 0.5 first-order conditional probability of ''go'' following a ''no-go'' or the reverse.

Sessions consisted of ten warm-up trials followed by a sufficient number of test trials to produce a threshold defined as ten reversals (see below). Based on prior threshold bracketing data, starting stimuli were presented at levels easily audible to the dolphin. To estimate the threshold for each frequency tested, the test stimulus was attenuated from the warm-up level in decrements of 2 dB until the dolphin failed to respond correctly. The first miss constituted the beginning of a run and the SPL was increased in 1-dB steps until the animal again made a correct response to a signal trial (hit). A run was defined as a series of attenuator changes in one direction (i.e., increments or decrements contingent on the animal's response to signal trials). For the rest of the session, 1-dB steps were used for ascending and descending runs. False alarms and correct rejections did not modify signal levels. Thresholds were estimated by averaging the attenuator value of the ten reversal points and subtracting that from the calibrated stimulus level. This procedure yields an estimated 50% threshold value.

D. Equipment and procedure for free-field measurements

As a comparison for the jawphone thresholds, auditory thresholds for the female dolphin were measured in the free

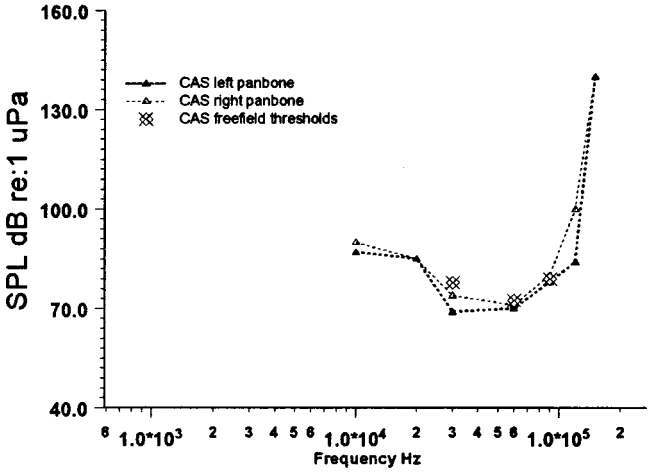


FIG. 6. The jawphone and free-field thresholds for the female dolphin (CAS).

field at 30, 60, and 90 kHz using a B&K 8103 hydrophone as the transmitter. Free-field source levels were measured using an Edo Western 6069 calibrated hydrophone fixed at 30 cm from the acoustic axis of the transmitter, the same position described above.

Thresholds for the dolphin's left and right ears were measured individually by positioning the hydrophone 30 cm away with the acoustic axis directly in line with the animal's pan bone. The testing procedures were the same as those described above.

III. RESULTS

A. Audiograms for the young female dolphin

The jawphone thresholds for the female dolphin measured between 10 and 120 kHz and the free-field thresholds measured at 30, 60, and 90 kHz are shown in Fig. 6. Sensitivity in the two ears is well matched. As would be anticipated, the audiograms are typically mammalian and parallel the standard for this species measured by Johnson (1967). The free-field thresholds are in agreement as well.

B. Audiograms for the male dolphin

Jawphone thresholds for the male dolphin are shown in Fig. 2. Previously, this dolphin was a subject in validation tests for the use of jawphones and had thresholds that closely matched the standard audiogram for the test range of 4–10 kHz (Moore and Pawloski, 1993). The 2–3-dB difference in sensitivity between the two ears for thresholds taken at age 26 was considered unremarkable at the time, and it was assumed that this dolphin's hearing was normal over the range of untested frequencies. However, our jawphone thresholds measured for the same dolphin seven years later at the age of 33, revealed substantial hearing deficits when compared to both the standard and the female dolphin's audiograms at high frequencies and for the right ear versus the left ear at mid-range frequencies. The right ear appears to be 16–33 dB less sensitive as compared to the left ear between 10 and 40 kHz. Sensitivity in both ears falls off rapidly above 55 kHz, suggesting a high-frequency cutoff of about 70 kHz lower in comparison to both the standard (Johnson, 1967) and the female dolphin in this study.

IV. DISCUSSION

Figure 2 compares Johnson's (1967) audiogram to our measurements of the ambient noise at the test site and all estimated thresholds for both dolphins. Since our female dolphin and Johnson's male were healthy and close in age, we expected their audiograms to compare favorably. Our measurements are obviously limited by the noise in San Diego Bay, resulting in a 27–30 dB difference in sensitivity compared to the standard. However, the shapes of the resulting auditory functions for both jawphone and free-field measurements clearly parallel the standard, indicating normal hearing for the young female dolphin.

The jawphone data for the older male, however, clearly show hearing losses in comparison to both the female and to the standard.

Ridgway and Carder (1993, 1997) first reported hearing deficits in older bottlenose dolphins observed in the conduct of an acoustic response time study. Three older males (23, 26, and 34 years of age) exhibited abrupt or progressive hearing losses between 70 and 120 kHz at 111–135 dB *re*: 1 μPa while an older female (33 years of age) exhibited a loss of sensitivity to frequencies above 100 kHz. In contrast, four other dolphins in the same study (three females of ages 11, 32, and 35, and a 7-year-old male) exhibited no evidence of hearing loss to the same stimuli. All of the subject animals were considered to be of normal health without any history of auditory irregularities.

Our male dolphin's bilateral loss of sensitivity to frequencies above 55 kHz at the age of 33 is consistent with the findings of Ridgway and Carder (1993, 1997). These data considered collectively support the notion that male dolphins of this species lose sensitivity to high frequencies as a function of age; a condition that may be similar to presbycusis (Gulick *et al.*, 1989) in humans. The distinct loss of sensitivity in our male dolphin's right ear, however, suggests that variables other than or in addition to age are involved. Ketten *et al.* (1997) reported the clinical aspects of the hearing loss in this animal as well as on those of the Ridgway and Carder (1993, 1997) study and discussed the implications for hearing loss mechanisms.

There is a good deal of data that indicate the plasticity of and the dolphin's control over the parameters of outgoing echolocation signals (see Moore and Pawloski, 1990). Brill *et al.* (1992) reported that a false killer whale (*Pseudorca crassidens*) adjusted the source level, bandwidth, and repetition rate of its outgoing signals as functions of target shape and strength, task specificity, and learning experience. Likewise, Au *et al.* (1995) described four basic echolocation signals used by a false killer whale that had a linear relationship between the center frequency and the peak-to-peak sound level. Brill and Harder (1991) reported alterations in a dolphin's outgoing signals in response to the attenuation of returning echoes. Houser *et al.* (1997) evaluated longitudinal data on our male dolphin's outgoing echolocation signals

and found signal variations over time that are consistent with the decrease in sensitivity that we measured in its hearing.

Unlike the human auditory system, there is no quantified model of dolphin auditory pathways from the surface of the head or jaw to the inner ear. Any assumptions concerning the sound conducting properties of a suspected path are unwarranted. Lacking any scientifically credible model, therefore, we assume that the calibrated sound levels we measured were the same levels that reached the animal's inner ear. This assumption produces a conservative threshold estimate, as there are many interfaces between the various fat bodies and bony surfaces that lay between the surface of the head and the inner ear that potentially produce acoustic losses. Also, any effects of loading that may have occurred due to the attachment of the jawphone to the dolphin's head were ignored, as the amount of suction required was minimal.

The observations that the free-field thresholds and the jawphone thresholds for 30, 60, and 90 kHz are approximately the same validate our assumption that the jawphone simply serves to place the sound source in closer proximity to the animal and concentrates the source in a small, localized area. Since the dolphin ear is water adapted, the pinna and meatal canal are either nonexistent or nonfunctional (McCormick *et al.*, 1970), resulting in the loss of the transformer action required of an air adapted mammalian ear. The dolphin's tissue and blubber are transparent to sound underwater and the ''direct path'' is via the water through the animal's tissue to the tympanoperiotic bone. The fat bodies in the lower jaw and on the side of the dolphin's head observed by Ketten *et al.* (1997) may act to enhance sound conduction to the tympanoperiotic bone based on frequency. However, the role of these fat channels is speculative. We assume that jawphones do not act like headphones do in humans since humans have an air-adapted ear with an ear canal, an air–tissue boundary (eardrum), and a middle ear transformer composed of the occicular chain. The dolphin's hearing, however, is tissue bound and relies on translational and compressional bone conduction (McCormick *et al.*, 1980).

The observation that the shape of functions we generated for the female dolphin with both jawphones and in the free field are nearly identical to that measured by Johnson (1967) supports the notion that the use of jawphones for auditory assessment is a useful method for the presentation of auditory stimuli. Given the sensitivity to interaural cues that has been measured with jawphones (Moore and Pawloski, 1993; Moore *et al.*, 1995), and the fact that the jawphone thresholds for these two dolphins bracket the established standard, our data support the use of jawphones for measurements of dolphin auditory sensitivity as a legitimate method that facilitates the use of unilateral or bilateral stimuli.

Sensitivity to an acoustic stimulus along the lower jaw appears to be a function of position. Møhl *et al.* (1999) used a contact sound source, a piezoelectric transducer embedded in a suction cup, and Brill *et al.* (2000) used jawphones as described in this paper to investigate relative sensitivity on the heads and lower jaws of their dolphin subjects. Both reported greater sensitivity forward of the panbone area for higher frequencies. Since the data reported here were collected prior to the Møhl *et al.* (1999) and Brill *et al.* (2000), the pan bone was chosen as the site of best sensitivity as described above. Had our measurements been taken elsewhere, we would have observed some difference in sensitivity.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research, Contract No. N001498WX20086, Code 032 US, Nancy Harned. The authors are grateful to Debra Skaar and Linda Williams, who assisted in the dolphin training and data collection, to Dr. Dorian Houser and Dr. James Finneran for their assistance in making noise measurements, and Dr. David Helweg for his helpful comments on this manuscript.

Andersen, S. (**1971**). ''Auditory sensitivity of the harbor porpoise *Phocoena phocoena*,'' in *Investigations on Cetacea* (University of Berne, Berne), Vol. III, pp. 255–259.

Au, W. W. L., Carder, D. A., Penner, R. H., and Scronce, B. L. (**1985**). ''Demonstration of adaptation in beluga whale echolocation signals,'' J. Acoust. Soc. Am. **77**, 726–730.

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.

Brill, R. L., Moore, P. W. B., Helweg, D. A., and Dankiewicz, L. A. (**2000**). ''Mapping acoustic sensitivity about the dolphin's head: A look at the peripheral hearing system,'' J. Acoust. Soc. Am. **107**, 2786.

Brill, R. L., and Harder, P. J. (**1991**). ''The effects of attenuating returning echolocation signals at the lower jaw of a dolphin (*Tursiops truncatus*),'' J. Acoust. Soc. Am. **89**, 2851–2857.

Brill, R. L., Sevenich, M. L., Sullivan, T. J., Sustman, J. D., and Witt, R. E. (**1988**). ''Behavioral evidence of hearing through the lower jaw by an echolocating dolphin (*Tursiops truncatus*),'' Marine Mammal Sci. **4**, 223–230.

Brill, R. L., Pawloski, J. L., Helweg, D. A., Au, W. W., and Moore, P. W. B. (**1992**). ''Target detection, shape discrimination, and signal characteristics of an echolocating false killer whale (*Pseudorca crassidens*),'' J. Acoust. Soc. Am. **92**, 1324–1330.

Bullock, T. H., Grinnell, A. D., Ikezono, E., Kameda, K., Nomoto, M., Sato, O., Suga, N., and Yanagisawa, K. (**1968**). ''Electrophysiological studies of the central auditory mechanisms in cetaceans,'' Z. Vergleich. Physiol. **59**, 117–156.

Gellerman, L. W. (**1933**). ''Chance orders of alternative stimuli in visual discrimination experiments,'' J. Gen. Psychol. **42**, 206–208.

Gulick, W. L., Gescheider, G. A., and Frisina, R. D. (**1989**). *Hearing: Physiological Acoustics, Neural Coding, and Psychoacoustics* (Oxford University Press, New York).

Hall, J. D., and Johnson, C. S. (**1972**). ''Auditory thresholds of a killer whale *Orcinus orca* Linnaeus,'' J. Acoust. Soc. Am. **51**, 515–517.

Houser, D. S., Helweg, D. A., and Moore, P. W. B. (**1997**). ''Classification of dolphin echolocation clicks by energy and frequency distributions,'' J. Acoust. Soc. Am. **102**, 3124 (abstract).

Jacobs, D. W., and Hall, J. D. (**1972**). ''Auditory thresholds of a freshwater dolphin *Inia geoffrensis* Blaineville,'' J. Acoust. Soc. Am. **51**, 530–533.

Johnson, C. S. (**1967**). ''Sound detection thresholds in marine mammals,'' in *Marine Bioacoustics*, edited by W. Tavolga (Pergamon, New York), pp. 247–260.

Ketten, D. K., Moore, P. W. B., Dankiewicz, L. A., Brill, R. L., and Van Bonn, W. (**1997**). ''The slippery slope of a Johnsonian ear: Natural variability versus natural loss,'' J. Acoust. Soc. Am. **102**, 3101 (abstract).

McCormick, J. G., Wever, E. G., Palin, J., and Ridgway, S. H. (**1970**). ''Sound conduction in the dolphin ear,'' J. Acoust. Soc. Am. **48**, 1418–1428.

McCormick, J. G., Wever, E. G., Ridgway, S. H., and Palin, J. (**1980**). ''Sound reception in the porpoise as it relates to echolocation,'' in *Animal Sonar Systems*, edited by R.-G. Busnel and J. F. Fish (Plenum, New York), pp. 449–467.

Møhl, B., Au, W. W. L., Pawloski, J., and Nachtigall, P. E. (**1999**). ''Dolphin hearing: Relative sensitivity as a function of point of application of a contact sound source in the jaw and head region,'' J. Acoust. Soc. Am. **105**, 3421–3424.

Moore, P. W. B., and Pawloski, D. A. (**1990**). ''Investigations on the control of echolocation pulses in the dolphin (*Tursiops truncatus*),'' in *Sensory Abilities of Cetaceans*, edited by J. Thomas and R. Kastelein (Plenum, New York), pp. 305–316.

Moore, P. W. B., and Pawloski, D. A. (**1993**). ''Interaural time discrimination in the bottlenose dolphin,'' J. Acoust. Soc. Am. **94**, 1829.

Moore, P. W. B., Pawloski, D. A., and Dankiewicz, L. A. (**1995**). ''Interaural time and intensity difference thresholds in the Bottlenose dolphin (*Tursiops truncatus*),'' in *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (De Spil, Woerden, The Netherlands), pp. 11–23.

Moore, P. W. B., and Schusterman, R. J. (**1987**). ''Audiometric assessment of Northern fur seals, *Callorhinus ursinus*,'' Marine Mammal Sci. **3**, 31–53.

Norris, K. S. (**1964**). ''Some problems of echolocation in cetaceans,'' in *Marine Bioacoustics*, edited by W. N. Tavolga (Pergamon, New York), pp. 317–336.

Norris, K. S. (**1968**). ''The evolution of acoustic mechanisms in odontocete cetaceans,'' in *Evolution and Environment*, edited by E. T. Drake (Yale University Press, New Haven, CT), pp. 297–324.

Norris, K. S. (**1969**). ''The echolocation of marine mammals,'' in *The Biology of Marine Mammals*, edited by H. T. Andersen (Academic, New York, NY), pp. 391–423.

Renaud, D. L., and Popper, A. N. (**1975**). ''Sound localization by the bottlenose porpoise *Tursiops truncatus*,'' J. Exp. Biol. **63**, 569–585.

Ridgway, S. H., and Carder, D. A. (**1993**). ''High-frequency hearing loss in old (25+ years-old) male dolphins,'' J. Acoust. Soc. Am. **94**, 1830.

Ridgway, S. H., and Carder, D. A. (**1997**). ''Hearing deficits measured in some *Tursiops truncatus*, and discovery of a deaf/mute dolphin,'' J. Acoust. Soc. Am. **101**, 590–594.

Thomas, J., Chun, N., and Au, W. (**1988**). ''Underwater audiogram of a false killer whale (*Pseudorca crassidens*),'' J. Acoust. Soc. Am. **84**, 936–940.

Varanasi, U., and Malins, D. C. (**1971**). ''Unique lipids of the porpoise (*Tursiops gilli*): differences in triacylglycerols and wax esters of acoustic (mandibular and melon) and blubber tissues,'' Biochem. Biophys. Acta **231**, 415–418.

Varanasi, U., and Malins, D. C. (**1972**). ''Triacylglycerols characteristic of porpoise acoustic tissues: molecular structures of diisovaleroylglycerides,'' Science **176**, 926–928.

White, M. J., Norris, J., Ljungblad, D., Baron, K., and Di Sciara, G. (**1978**). ''Auditory thresholds of two beluga whales,'' Hubbs/Sea World Research Inst. Tech. Rep. No. 78–109.

Comparison between visual and passive acoustic detection of finless porpoises in the Yangtze River, China

Tomonari Akamatsu[a)]
National Research Institute of Fisheries Engineering, Ebidai, Hasaki, Kashima, Ibaraki 314-0421, Japan

Ding Wang,[b)] Kexiong Wang, and Zhou Wei
Institute of Hydrobiology, The Chinese Academy of Sciences, Wuhan 430072, People's Republic of China

(Received 8 February 2000; accepted for publication 24 January 2001)

Recently, sonar signals and other sounds produced by cetaceans have been used for acoustic detection of individuals and groups in the wild. However, the detection probability ascertained by concomitant visual survey has not been demonstrated extensively. The finless porpoises (*Neophocaena phocaenoides*) have narrow band and high-frequency sonar signals, which are distinctive from background noises. Underwater sound monitoring with hydrophones (B&K8103) placed along the sides of a research vessel, concurrent with visual observations was conducted in the Yangtze River from Wuhan to Poyang Lake in 1998 in China. The peak to peak detection threshold was set at 133 dB *re* 1 μPa. With this threshold level, porpoises could be detected reliably within 300 m of the hydrophone. In a total of 774-km cruise, 588 finless porpoises were sighted by visual observation and 44 864 ultrasonic pulses were recorded by the acoustical observation system. The acoustic monitoring system could detect the presence of the finless porpoises 82% of the time. A false alarm in the system occurred with a frequency of 0.9%. The high-frequency acoustical observation is suggested as an effective method for field surveys of small cetaceans, which produce high-frequency sonar signals. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1356705]

PACS numbers: 43.80.Ka, 43.80.Jz, 43.66.Gf [WA]

I. INTRODUCTION

Acoustical survey is an effective method to detect cetaceans. Submarine detection systems have recorded minke whale (*Balaenoptera acutorostrata*) sounds (Nishimura and Conlon, 1994) and located blue whales (*Balaenoptera musculus*) by monitoring their low frequency calls (Stafford *et al.*, 1998). The distance and depth of Hawaiian spinner dolphins (*Stenella longirostris*) were measured using a one-hydrophone system (Aubauer *et al.*, 2000). Atlantic bottlenose dolphins (*Tursiops truncatus*) were localized with a passive acoustical observation system (Freitag and Tyack, 1993). The sperm whale's (*Physeter macrocephalus*) pulse signals can be used for measurement of body size (Goold, 1996). Using the delay of surface reflection from the direct path signal, the diving depth of a sperm whale was obtained (Bondaryk *et al.*, 1999). A sonobuoy system used for other cetacean observations (van Parijs *et al.*, 1998), for example, humpback whales (*Megaptera novaeangliae*) (Norris *et al.*, 1998), bottlenose dolphins (Furusawa, 1998) and bowhead whales (*Balaena mysticetus*) (Cummings and Holliday, 1985). The search effort and target range of porpoises indicated by echolocation events and pulse intervals were observed in the wild (Akamatsu *et al.*, 1994, 1998).

Acoustical surveys of cetaceans have several advantages compared with visual surveys. The acoustical survey can be operated automatically and maintain an identical detection threshold over long times. The acoustical survey enables all day observation even nocturnally. In contrast, the advantages of the visual surveys are that they can recognize different species and count numbers of animals directly. An abundance of populations can be estimated using line transect surveys (Buckland *et al.*, 1993; Shirakihara *et al.*, 1994).

So far, the detection probability of cetacean by acoustical surveys could not be compared with visual findings quantitatively, except for several baleen whales (McDonald and Fox, 1999; Clark and Fristrup, 1997), because the performance of an acoustical observation system is highly variable. The detection performance depends on vocalization rates, distances of animals and background noise levels. To determine the reliability of acoustical detection, a concomitant visual survey is needed.

Finless porpoises have good echolocation abilities (Nakahara *et al.*, 1997). They produce echolocation signals every three to four seconds (Akamatsu *et al.*, 2000). The echolocation signals of this species have a relatively narrow band spectrum (Kamminga *et al.*, 1986). These characteristics encourage a good signal-to-noise ratio in a noisy environment. The dive time of finless porpoises is usually no longer than two minutes (Zhang *et al.*, 1996), which makes visual detection possible within a couple of minutes by experienced observers, even in the muddy water environment. A passive acoustical monitoring of the finless porpoises, concurrent with visual observations, was conducted in the Yangtze River, China, to determine the detection probability of the acoustical observation system on the presence of animals.

II. MATERIALS AND METHODS

A. Visual surveys

The research vessel surveyed from Wuhan to the mouth of Poyang Lake downstream. The upstream operation caused

[a)]Electronic mail: akamatsu@nrife.affrc.go.jp
[b)]Electronic mail: wangd@ns.ihb.ac.cn

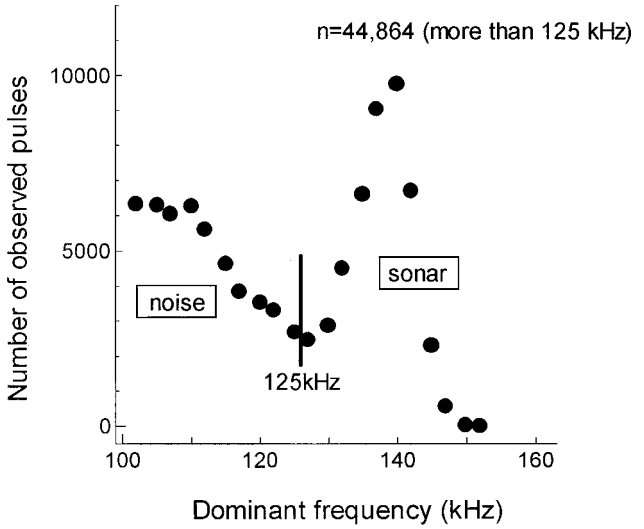


FIG. 1. Dominant frequency distribution of ultrasonic pulses detected by the acoustical observation system. To exclude the noise, only the pulses having peak frequency more than 125 kHz were analyzed.

intense noise and made acoustical monitoring difficult. The relative speed of the research vessel with water flow was fixed at 2.8 m/s, which was similar to the maximum swimming speed (2.9 m/s) of a free-ranging finless porpoise (Akamatsu *et al.*, 2000). The ground speed was 2.9 m/s to 4.0 m/s depending on the current speed of the Yangtze River. Two visual observers on the top of the research vessel (6.25-m high from the water surface) searched the left and right sides of the frontal area, respectively. When an observer found a group of finless porpoises, the minimum number, the distance and direction from the research vessel and behavioral remarks of the animals were recorded. The minimum group size was defined as the number of the animals surfaced at the same time. After the determination of the minimum group size, the observer switched the observation area to the downstream area below the last observed animals. This protocol prevented double counting, because the vessel speed is likely larger than the swim speed of the animal. The visual detection was compared later to the recordings of the acoustical observation system to determine the detection probability of the system on the presence of animals.

B. Acoustical surveys

Two broad band hydrophones (B&K 8103, sensitivity -211 dB *re* 1 V/μPa $+2/-9$ dB, up to 180 kHz) were deployed at a depth of 0.8 m, 6.3 m apart from each other on two sides of the vessel (Kekao No. 1 of the Institute of Hydrobiology, The Chinese Academy of Sciences). A preamplifier with 10-kHz high pass filter (OKI ST-80B) eliminated the low frequency noises caused by water flow and engine operation. A digital data recorder (SONY PCHB 244, sampling rate of 384 kHz, dc to 147 kHz within 3 dB) allowed only one channel broad band recording at a time. Hence the channel on the appropriate side where animals frequently appeared was recorded accordingly. The calibration of the present system was done in a test tank of System Giken Co. by projecting 100 kHz and 10-cycle tone bursts.

All the pulses more than 133 dB peak-to-peak *re* 1 μPa were digitized by using an analogu digital converter (System Design Service, DASS BOX 1000). A data analysis program on MATLAB® developed for this study calculated dominant frequency (an analysis bandwidth of 2.49 kHz) and the pulse interval of each pulse. Most of the low frequency and high intensity pulse noises were caused by water flow around the hydrophone. Since the dominant frequencies of finless porpoise sonar signals were usually higher than 110 kHz (Kamminga *et al.*, 1986; Akamatsu *et al.*, 1998), only the pulses having more than a 100-kHz dominant frequency processed. The irregular (two times or more and half or less) changes of successive pulse intervals were also considered to be noises, because the pulse intervals of the finless porpoise echolocation signals were usually changed smoothly (see Akamatsu *et al.*, 1998 for details). Additionally, the maximum sound pressure levels in a click train detected on the animals at 100 m from the research vessel were averaged.

III. RESULTS

In total, 588 finless porpoises were sighted visually during 774-km cruise along the Yangtze River, in November 1998. During 1,835 minutes effective recording time, echolocation clicks of finless porpoises were present and recorded during 214 minutes.

Highly endangered baiji (*Lipotes vexillifer*), which share the same river, are quite different from the porpoise. For example, baiji have a dorsal fin and emit broad band sonar signals (Akamatsu *et al.*, 1998). These features make the discrimination between two species quite easy, both visually and acoustically. No baiji were observed by both a visual and an acoustic method during the entire cruise.

The acoustical observation system detected 93,418 ultrasonic pulses with dominant frequencies higher than 100 kHz. The dominant frequency distribution of these pulses had a local peak at 140 kHz and local minimum at 125 kHz as shown in Fig. 1. While higher than 125 kHz, the pulses showed typical narrow band characters of sonar signals of the finless porpoises. Pulses with dominant frequencies of less than 125 kHz showed broadband and long duration characters, which were considered to be noise. Therefore, only pulses with dominant frequencies higher than 125 kHz ($n = 44{,}864$) were analyzed hereafter.

Numbers of pulses and finless porpoises observed in every minute are presented in Fig. 2, showing 500 minutes (about two days cruise) effective recording time. A clear coincidence between visual and acoustic detection can be seen.

The correct detection by the acoustical observation system is defined as the detection of more than a cutoff number of pulses (15 pulses/minute) as depicted by the horizontal bar in Fig. 3 within a time window. The window width was set as ± 2 minutes from the moment of visual sighting of the animals. A miss is defined as less than 15 pulses captured in a minute within the time window. Correct rejection is defined as less than the cutoff number of pulses associated with no visual findings within ± 5 minutes. A false alarm is defined as more than the cutoff number of pulses associated without visual finding within ± 5 minutes.

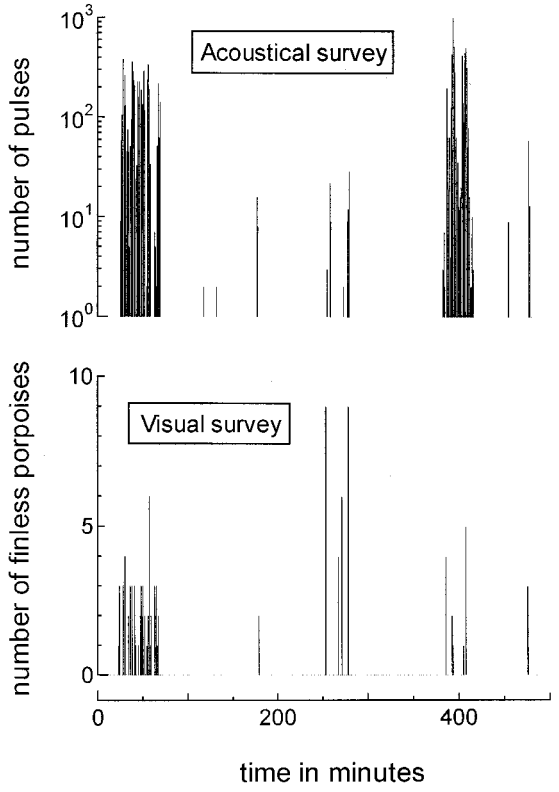


FIG. 2. Comparisons between the number of finless porpoises and pulse signals every minute during part of the observation (500 min). A clear coincidence between visual and acoustical detection is indicated.

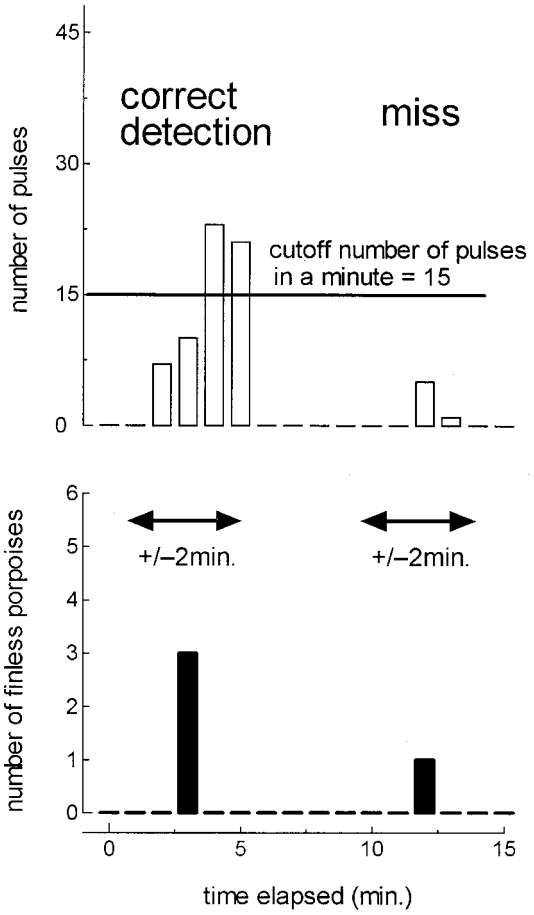


FIG. 3. Schematic of correct detection and miss of the acoustical observation system. A time window centered at the visual detection was depicted as two arrows. The acoustical correct detection of the animals was defined that the number of the observed pulses in a minute within the time window was more than 15. The acoustical miss was defined as the lesser number of pulses in a minute within the time window.

The receiver operating characteristic (ROC) curve according to the cutoff number of pulses in a minute were shown in Fig. 4. At the cutoff number of pulses (15), reliable correct detection (82%) and small false alarm levels (0.9%) were indicated. This was the reason to choose 15 as the cutoff number of pulses.

The miss rate according to the distance from the research vessel is depicted in Fig. 5. The porpoises could be observed visually up to 600 m at maximum. Out of the 300-m range, the miss rate of the acoustical observation system was higher than 45%, whereas it was less than 25% within a 300-m range. The average of the maximum sound pressure level corresponding with visual detection at 100 m was 137.4 dB p–p re 1 μPa.

IV. DISCUSSION

This study showed that a high frequency acoustical survey of finless porpoises has a reliable performance to detect echolocating cetaceans with a small false alarm rate. However, the acoustic detection probability of the finless porpoises depends on the source level of the echolocation signal, directionality of the beam pattern and vocalization rate. The evaluation of these factors will help to understand the high detection probability of the present system.

The averaged sound pressure level of the finless porpoises' signal was 137.4 dB at 100 m apart from the observer. In the shallow water system, the sounds propagate spherically up to the depth of the water, then become a cylindrical propagation. Assuming the depth of the Yangtze

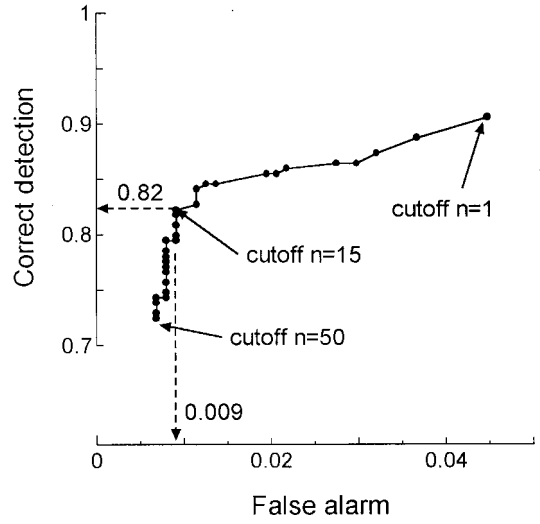


FIG. 4. ROC curve according to the cutoff number of pulses. The false alarm decreased until up to 15 cutoff number of pulses and was stable more than 15.

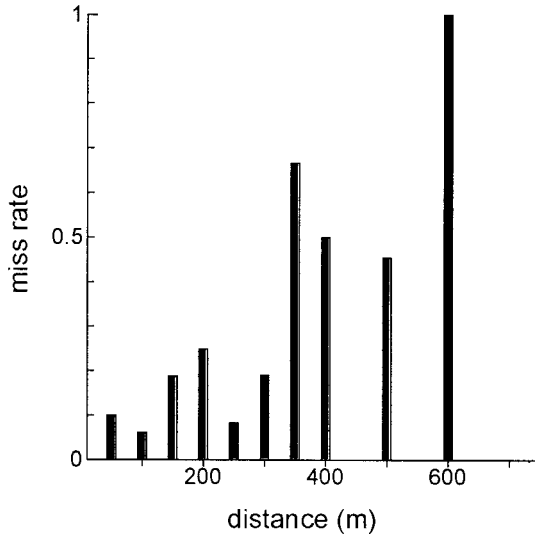


FIG. 5. Miss rate according to the distance from the research vessel. The larger miss rate was observed at more than 300-m distance.

River is 10 m, the sound produced from a porpoise propagates spherically up to 10 m with −20 dB attenuation. Then, the sound propagates cylindrically and has another −10-dB attenuation up to 100 m. Therefore the source level of the finless porpoises is calculated to be 167.4 dB. Using this source level and a 133-dB detection threshold level, the effective acoustic detection distance is calculated as 275 m. This is consistent with the higher missing rate of more than 300 m (Fig. 5). If we exclude the data out of the 300-m range, the correct detection rate will increase to be 88%.

The two-minute time window used in the definition of correct detection and miss was chosen from the 300-m effective acoustical detection distance estimated above. During two minutes the research vessel proceeded 330 m. A porpoise might produce sound before or after visual finding. The two-minute time window is adequate to detect a sound produced within the acoustic detection distance. On the other hand, the five-minute time window used for the definition of a false alarm and correct rejection was chosen to exclude any possibility of porpoise existence around the research vessel. Within the ±5 minute cruise, the vessel proceeds 1660 m, which is approximately five times longer than the acoustical detection range. If there was no visual finding with no acoustical detection, it is considered to be correct rejection. If there was no visual finding with more than the cutoff number of pulses recording, it should be treated as a false alarm.

Dolphin sonar signals are directional and the sound pressure levels change by 0 dB to −20 dB depending on the relative angle of a bottlenose dolphin toward a hydrophone (Au, 1993). Au *et al.* (1999) reported that the 3-dB beam width of a harbor porpoise (*Phocoena phocoena*) was approximately 16.5 degrees. This is wider than other species, for example, 9.7 degrees for bottlenose dolphins (Au, 1993) and 6.5 degrees for beluga (*Delphinapterus leucas*) (Au *et al.*, 1987). According to the acoustic datalogger recording of free-ranging finless porpoises, the 120 degrees off-axis sonar signals still had a 160-dB peak-to-peak sound pressure level at one meter from the animal (Akamatsu *et al.*, 2000). In this case, the detection range of the present system is expected to be 50 m even if the animals are 120 degrees off the direction to the hydrophone.

Two free-ranging finless porpoises in clearer water of the semi-natural reserve (an oxbow of the Yangter River) produced 14.5 to 19.1 click trains in a minute (Akamatsu *et al.*, 2000). This suggests that a sufficient number of pulses can be recorded when the animals were around the hydrophone since a click train consists of several tens of ultrasonic pulses. Especially, in the muddy water of the Yangtze River, finless porpoises are thought to use echolocation much more frequently, because of the limited visual sense.

The fresh water population of finless porpoise in the Yangtze River is now endangered because of the expansion of human activities, such as fisheries, water transportation and pollution (Wang *et al.*, 2000). For the conservation of this population, the distribution along the entire river should be known. Once the detection performance of an acoustical observation system has been evaluated, the presence of animals can be observed acoustically. For example, when underwater sound monitoring systems are deployed at several points along the Yangtze River, it seems possible to observe the movement of finless porpoise passing by the points. Results may provide the dinural or seasonal migration pattern of this species. The finless porpoises have no clear natural marks to identify individuals, thus photo identification does not work. The acoustical observation method seems quite helpful for the Yangtze River finless porpoise conservation and other small cetaceans as well.

ACKNOWLEDGMENTS

We wish to thank the Institute of Hydrobiology, The Chinese Academy of Sciences and the crew of the research vessel KEKAO No. 1, Xiaoqiang Wang, Xinan Kuang, En Kawa, Kotoe Sasamori, and Hirokazu Ichikawa worked with us in the field and greatly supported this experiment. Ryoji Shimizu (OKI Sea Tech), Kenji Kato (Shizuoka OKI Electric), Tetsuo Fukuchi and Akira Kojima (System Giken) developed the acoustical observation system. Thomas Jefferson and Kathleen Dudzinski provided constructive criticism on this manuscript. This research was supported by the Science and Technology Agency of Japan, National Research Institute of Fisheries Engineering, the Chinese Academy of Sciences (KZ951-A1-105 and 1999051), and the National Natural Science Foundation of China (39670105), Grant-in-Aid for Scientific Research (B) from the Ministry of Education, Science and Culture of Japan (B09450172 and 10460086).

Akamatsu, T., Wang, D., Wang, K., and Naito, Y. (**2000**). ''A method for individual identification of echolocation signals in free-ranging finless porpoises carrying data loggers,'' J. Acoust. Soc. Am. **108**, 1353–1356.

Akamatsu, T., Wang, D., Nakamura, K., and Wang, K. (**1998**). ''Echolocation range of captive and free-ranging baiji (*Lipotes vexillifer*), finless porpoise (*Neophocaena phocaenoides*) and bottlenose dolphin (*Tursiops truncatus*),'' J. Acoust. Soc. Am. **104**, 2511–2516.

Akamatsu, T., Hatakeyama, Y., Kojima, T., and Soeda, H. (**1994**). ''Echolocation rates of two harbor porpoises (*Phocoena phocoena*),'' Marine Mammal Sci. **10**, 401–411.

Au, W. W. L., Kastelein, R. A., Rippe, H. T., and Schooneman, N. (**1999**). ''The transmission beam pattern and echolocation signal of the harbor porpoise (*Phocoena phocoena*),'' J. Acoust. Soc. Am. **106**, 3699–3705.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer-Verlag, New York).

Au, W. W. L., Penner, R. H., and Turl, C. W. (**1987**). ''Propagation of beluga echolocation signals,'' J. Acoust. Soc. Am. **82**, 807–813.
Aubauer, R., Lammers, M. O., and Au, W. W. L. (**2000**). ''One-hydrophone method of estimating distance and depth of phonating dolphins in shallow water,'' J. Acoust. Soc. Am. **107**, 2744–2749.
Bondaryk, J. E., D'Amico, A., and Portunato, N. (**1999**). ''Techniques for passive detection and localization of marine mammals using wide aperture arrays,'' J. Acoust. Soc. Am. **106**, 2188.
Buckland, S. T., Anderson, D. R., Burnham, K. P., and Laake, J. L. (**1993**). *Distance Sampling Estimating Abundance of Biological Populations* (Chapman & Hall, London).
Clark, C. W., and Fristrup, K. M. (**1997**). ''Whales '95: A combined visual and acoustic survey of blue and fin whales off southern California,'' Rep. Int. Whal. Commn. 0 (47), 583–600.
Cummings, W. C., and Holliday, D. V. (**1985**). ''Passive acoustic location of bowhead whales in a population census off Point Barrow, Alaska,'' J. Acoust. Soc. Am. **78**, 1163–1169.
Freitag, L. E., and Tyack, P. L. (**1993**). ''Passive acoustic localization of the Atlantic bottlenose dolphin using whistles and echolocation clicks,'' J. Acoust. Soc. Am. **93**, 2197–2205.
Furusawa, M. (**1998**). ''Acoustical observation of bottlenose dolphins around Mikura Island, Tokyo, Japan,'' Report of dolphin watching in Mikura Island, Tokyo Metropolis (in Japanese).
Goold, J. C. (**1996**). ''Signal processing techniques for acoustic measurement of sperm whale body lengths,'' J. Acoust. Soc. Am. **100**, 3431–3441.
Kamminga, C., Kataoka, T., and Engelsma, F. J. (**1986**). ''Investigations on cetacean sonar VII. Underwater sounds of *Neophocaena phocaenoides* of the Japanese coastal population,'' Aquatic Mammals **12**, 52–60.
McDonald, M. A., and Fox, C. G. (**1999**). ''Passive acoustic methods applied to fin whale population density estimation,'' J. Acoust. Soc. Am. **105**, 2643–2651.
Nakahara, F., Akira, T., Koido, T., and Hirud, H. (**1997**). ''Target discrimination by an echolocating finless porpoise, *Neophocaena phocaenoides*,'' Marine Mammal Sci. **13**, 639–649.
Nishimura, C. E., and Conlon, D. M. (**1994**). ''IUSS dual use: monitoring whales and earthquakes using SOSUS,'' Mar. Technol. Soc. J. **27**, 13–21.
Norris, T. F., McDonald, M., and Barlow, J. (**1998**). ''Acoustic detections of singing humpback whales (*Megaptera novaeangliae*) in the eastern North Pacific during their northbound migration,'' J. Acoust. Soc. Am. **106**, 506–514.
Shirakihara, M., Shirakihara, K., and Takemura, A. (**1994**). ''Distribution and seasonal density of the finless porpoise *Neophocaena phocaenoides* in the coastal waters of western Kyushu, Japan,'' Fisheries Sci. **60**, 41–46.
Stafford, K. M., Fox, C. G., and Clark, D. S. (**1998**). ''Long-range acoustic detection and localization of blue whale calls in the northeast Pacific Ocean,'' J. Acoust. Soc. Am. **104**, 3616–3625.
van Parijs, S. M., Thompson, P. M., Hastie, G. D., and Bartels, B. A. (**1998**). ''Modification and deployment of a sonobuoy for recording underwater vocalizations from marine mammals,'' Marine Mammal Sci. **14**, 310–315.
Wang, D., Liu, R., Zhang, X., Yang, J., Wei, Z., Zhao, Q., and Wang, X. (**2000**). ''Status and conservation of the Yangtze finless porpoise,'' Occasional Papers of IUCN/SSC 23, 81–85.
Zhang, X., Wang, D., Yang, J., Wei, Z., Wang, K., and Würsig, B. (**1996**). ''Study on radio-tracking finless porpoise *Neophocaena phocaenoides*, at the Yangtze River,'' Acta Ecol. Sin. **16**, 490–496.

The acoustic calls of blue whales off California with gender data

Mark A. McDonald
2535 Sky View Lane, Laramie, Wyoming 82070

John Calambokidis
Cascadia Research, 218½ West Fourth Avenue, Olympia, Washington 98501

Arthur M. Teranishi and John A. Hildebrand
Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0205

(Received 10 February 2000; revised 20 December 2000; accepted 5 January 2001)

The acoustic calls of blue whales off California are described with visual observations of behavior and with acoustic tracking. Acoustic call data with corresponding position tracks are analyzed for five calling blue whales during one 100-min time period. Three of the five animals produced type A-B calls while two produced another call type which we refer to as type D. One of the animals producing the A-B call type was identified as male. Pauses in call production corresponded to visually observed breathing intervals. There was no apparent coordination between the calling whales. The average call source level was calculated to be 186 dB *re*: 1 μPa at 1 m over the 10–110-Hz band for the type B calls. On two separate days, female blue whales were observed to be silent during respective monitoring periods of 20 min and 1 h. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1353593]

PACS numbers: 43.80.Lb [WA]

I. INTRODUCTION

The blue whale (*Baleanoptera musculus*) call repertoire includes intense, long-duration continuous calls, having the greatest acoustic power of any animal (Aroyan *et al.*, 2000). Little is understood of the behavioral function/role of blue whale calls. Possible functions include sensing the environment, prey detection, and communication. The behavioral significance of communication is typically determined by observing natural interactions and quantifying changes in behavior following a call (Tyack, 2000). Among the baleen whales, the behavioral function of acoustic calls is best studied for the humpback (*Megaptera noveangliae*), somewhat studied for the right (*Eubalaena glacialis* and *Eubalaena australis*), bowhead (*Balaena mysticetus*), and fin (*Balaenoptera physalus*) whales, and less well studied for the other species (Edds-Walton, 1997). For humpback whales there is an established gender bias in calling, where males produce the songs (Tyack, 1998).

To better understand the acoustic calls of blue whales, this study was undertaken to collect visual behavioral observations, photo-identification history, and gender for whales which were monitored acoustically for call behavior. The acoustic monitoring system allowed simultaneous tracking of multiple calling blue whales, some of which were not observed visually. To test for any communication context associated with a whale call we looked for either consistent behavioral interactions associated with the call, or an acoustic response (Edds-Walton, 1997). An important parameter in understanding blue whale call context would be the determination of any gender bias in the types of calls produced.

The blue whale population, which occupies the west coast of North America, numbers about 2000 animals (Barlow, 1995), of which more than 1100 individuals have been documented by photo-identification (Calambokidis *et al.*, 1999). This may be the largest population of blue whales in the world, and is among the best studied. These animals are believed to range from the Queen Charlotte islands of British Columbia to the Costa Rica Dome based on photo-identification, satellite tagging, and acoustic recordings (Calambokidis *et al.*, 1999; Mate *et al.*, 1999; Stafford *et al.*, 1999b). The most commonly recorded blue whale call type is that which we refer to as broadcast calls, typically 15 to 20 s each part, produced repeatedly with a nearly fixed interval for long periods of time, hours to days (Tyack, 1998; Watkins *et al.*, 2000). We believe the term broadcast call is appropriate because of the apparent lack of acoustic counter-calling and the long duration of these call series. We do not intend the term broadcast call to imply a specific function. There is now extensive remote monitoring of the broadcast-type blue whale calls (Clark and Charif, 1998; Curtis *et al.*, 1999; Stafford *et al.*, 1999a; Watkins *et al.*, 2000), but few recordings have been made together with behavioral observations, and none, previous to this study, with genetic material from the calling animals. A better understanding of these acoustic calls will contribute to the acoustic monitoring efforts.

II. DATA COLLECTION

A. Effort

The 53.3-m National Oceanic and Atmospheric Administration (NOAA) ship *McArthur* was used to survey whales off the coast of Oregon and California from 30 September through 16 October 1997. There were 172 h of survey effort for the *McArthur* during 16 days. Procedures and equipment on the flying bridge of the *McArthur* were similar to National Marine Fisheries Service (NMFS) line transect surveys, including computer logging of data with an integrated global positioning system (GPS) and pedestal-mounted 25 power binoculars with range finding reticles (Fiedler *et al.*, 1998).

At least three observers were always on watch on the *McArthur* flying bridge. Two rigid hull inflatable boats (RHIBs) sometimes ran ahead and to the side of the *McArthur*, weather permitting, to increase the chance of encountering whales for biopsy, photo-identification, and acoustic monitoring, but these vessels were not directly involved in line-transect survey efforts.

The primary goal of the survey was to collect photo-identification of blue and humpback whales with a secondary goal of collecting blue whale acoustic recordings. When a blue whale was sufficiently separated from other animals to be certain acoustic recordings could be correlated to a specific animal, there was an additional goal of collecting genetic material for gender determination. The cruise focused on areas where blue whales were likely to be found but remained relatively far offshore where whales would be difficult to study using small vessels alone.

B. Methods

1. Sonobuoy calibration and recording systems

Broadband sonobuoys of type 57B, and directional sonobuoys (DIFAR) of types 53B and 53D (Richardson *et al.*, 1995) were used to obtain the recordings described here. Sonobuoy signals were received on five radios specially modified and calibrated by GreeneRidge Sciences, Inc. (Goleta, CA) to an accuracy of 0.1 dB. Recording systems included two Sony TCD-D8 stereo digital recorders sampling at 48 kHz. A second system simultaneously recorded all five sonobuoy channels using a National Instruments A/D board sampling each channel at 1 kHz through custom-built active anti-aliasing filters, each calibrated to less than 0.3 dB. Data recorded on the National Instruments system were used for quantitative analyses of received levels and for time delay sound source tracking methods, while data from the Sony recorders were used for DIFAR processing of bearings to calling whales. The National Instruments recording system and the Sony recording system each time stamp the acoustic data with internal clock time which was synchronized to GPS time to within 1-s accuracy. The time jitter between channels on the National Instruments system was less than 1 ms. The frequency response of the DIFAR sonobuoys differs from that of the omni-directional sonobuoys and appropriate corrections were made in the frequency domain to flatten the total frequency response of all the sonobuoys and the frequency response of the active filters. The greatest potential error in our sound level measurements is the sonobuoy manufacturer specification which is ± 2 dB for type 57 buoys and ± 3 dB for type 53 buoys. Some reduction in overall error was obtained by averaging ambient noise spectrum levels on each buoy in an array during time periods where ship noise did not appear to cause a bias across the array. Sonobuoys are expected to always have a self-noise below ocean ambient noise levels.

A verification of the sound pressure level calibration was obtained for the 15 October data by comparing observed ambient ocean noise levels at 500 Hz, largely above shipping noise frequencies, with expected ocean ambient levels. The 15 October recordings were made during Beaufort zero conditions. A 500-Hz spectrum level of 42 dB *re*: 1 μPa^2/Hz was observed, corresponding to a Knudsen curve level of 46.5 dB (Urick, 1983). More recent work suggests the Beaufort zero Knudsen levels to be too high for the Pacific (Chapman and Cornish, 1993), indicating our sound pressure level calculations are in agreement with expected ambient noise levels. Four of the five sonobuoys used on this occasion showed the same levels within our 2–3-dB measurement accuracy. The signal recorded from the fifth sonobuoy (a type 53D) showed a totally unexpected frequency response as judged against the ambient noise spectrum measured by the other sonobuoys and was not used for source level estimates, though it was used for direction finding. The 0–500-Hz noise spectra was examined for each measurement of whale call received level, and when the sonobuoy or some other component of the recording system appeared to be overloaded by ship or whale sound, these data were not used.

2. Photographic identification of individual blue whales

Identification photographs were taken using standard procedures employed in past research off California and Washington (Calambokidis *et al.*, 1990). Both the right and left sides of blue whales in the vicinity of the dorsal fin or hump were photographed as well as the ventral surface of the flukes. Identification photographs were first compared to others from the same time period and then compared to a catalog of 1070 blue whales identified along the California coast from 1975 to 1997 (Calambokidis *et al.*, 1999). Individual whales that did not match past years and that were of suitable quality were assigned unique identification numbers and added to the blue whale catalog.

3. Biopsies of blue whales

Skin samples were collected to determine gender of individual whales (Baker *et al.*, 1991). Biopsy samples were collected from whales using a biopsy dart system (Lambertsen, 1987). The biopsy dart consisted of an aluminum crossbow bolt (arrow) and a stainless steel biopsy punch, which has a flange or ''stop'' to prevent its penetrating too deeply. The punch is 7 to 9 mm in diameter and 2 to 5 cm in length and is fitted with two or three internal pins to secure the sample. A hole drilled transversely through the punch and just distal of the flange prevents pressure buildup inside the punch as it penetrates the skin. The dart was fired from a commercially available crossbow having a 125- to 150-lb draw. The recoil from the bolt stop striking the whale dislodges it from the whale and the free-floating bolt is retrieved by hand. Dart recoveries were sometimes aided by luminescent dyes added to the bolts. Gender determinations were made by the molecular genetics laboratory at Southwest Fisheries Science Center.

4. Tracking methods

On several occasions when whales were sighted, line transect efforts were terminated, allowing both the *McArthur* and the RHIBs to deploy an array of sonobuoys in the vicinity of the whales. Visual whale positions were maintained

using bearings and ranges from the pedestal-mounted binoculars and an integrated GPS recording system when a RHIB was not available to directly record the GPS position of the final flukeprint of a surfacing series. Sonobuoy deployment positions were recorded using GPS. Personnel on the *McArthur* flying bridge watched for previously unseen whales and coordinated the RHIBs.

The sonobuoy array allowed us to determine the location of calling whales for correlation with visual positions of whale surfacings. Both time delay localization methods and DIFAR processing of bearing angle from each sonobuoy were applied to the data (D'Spain, 1994; D'Spain *et al.*, 1992). The DIFAR bearing errors were found to have a standard deviation of two degrees (McDonald, unpublished data). DIFAR localization of calls was found to provide more accurate positions than time delay localization and was used exclusively in the acoustic tracks presented here. Sonobuoy drift corrections were made with visual/GPS buoy drift information and with acoustic location surveys using weighted light bulbs as sound sources (Heard *et al.*, 1997) processed with root mean square (rms) residual grid search localization methods (Wilcock and Toomey, 1991).

5. Source level estimation

Received levels were converted to source levels using only spherical spreading losses (Urick, 1983). Source levels were calculated only for the one animal for which a GPS position was available for every surfacing. The location of each call from that animal was interpolated by the time of the call between consecutive surface positions. Only calibrated receivers at greater than 2.5 km were used because range errors become less significant to the source level computation at greater ranges. Sonobuoy calibration is lost for very short range recordings because the received amplitude exceeds the dynamic range of the sonobuoy. Only calls free from interference were used for these measurements, and many of these calls were recorded without interference on four calibrated sonobuoys, each at a different range.

A detailed propagation model was not considered to be viable due to the many poorly known variables including sound speed profiles, bathymetry, seafloor characteristics, and depth of the calling animals. Several hundred propagation loss models were computed with parabolic equation methods using a range of best estimates for each variable. The average of these models suggests propagation losses slightly higher on average than spherical spreading, but a spherical spreading model was chosen for simplicity.

Fine scale variability in propagation loss was examined by comparing computed source levels over the 10–110-Hz band from the same call to different sonobuoys and observed to be ± 3 dB. To determine if source levels varied from call to call, computed source levels were examined and observed to be ± 3 dB, about the same as observed from the same call to different sonobuoys. Variability in spectrum levels of components from the same call on multiple sonobuoys and from multiple calls on the same sonobuoy was observed to be much greater than 3 dB. Greater variability is to be expected in tonal propagation losses when compared to band level variability given the multi-component calls observed in this study. Fine-scale variability in propagation losses may be responsible for all the observed variability in source levels.

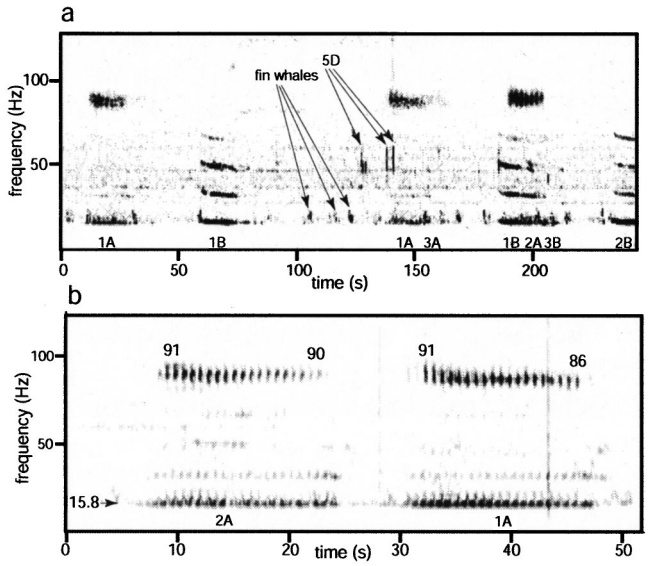


FIG. 1. (a) Acoustic data west of San Nicolas Island from 15 October at 1833 local time on the furthest west sonobuoy shown in Fig. 3, computed with a 1-s FFT, Hann window, and 75% overlap. Each call is labeled with the whale number shown in Fig. 3 and the call type designation. (b) Detail of type A calls from 15 October at 1759 local time, showing variability of 19 to 23 pulses between 92 and 85 Hz for the prominent overtone. Note the 90-Hz components are not harmonics of the 15.8-Hz fundamental. The FFT length is 0.5 s and overlap is 87.5% with a Hann window. The spectrograms shown have a frequency response emphasizing the higher frequencies by 4.5 or 6 dB per octave, depending whether a type 53 or 57 sonobuoy was used for the recording. This emphasis is built into the sonobuoy design and is intended to compensate for lower ambient noise levels at higher frequencies and thereby to maximize dynamic range.

III. RESULTS

A. Call characteristics

A total of 117 blue whales were seen in 78 sightings with 43 photographed and 33 identified. Acoustic recordings were obtained from an estimated 43 individual calling blue whales. These calls can be classified into two basic types, either patterned pairs, each of about 17-s duration, or irregular spaced calls of typically 2-s duration. Many of the estimated 43 individual blue whales acoustically recorded were never seen and recording was not always undertaken when blue whales were sighted.

A spectrogram of typical blue whale calls is shown in Fig. 1(a), as recorded on 15 October at 1833 local time. This spectrogram shows three blue whales producing two part broadcast calls, labeled as ''A'' and ''B'' in the spectrogram with each whale designated by a number. Each of these calling whales was individually tracked acoustically as described later. Another whale, designated five during this encounter, was producing downswept calls of about 1-s duration ranging from 60 to 45 Hz, these being labeled as type D calls. Fin whale calls are also present near 20 Hz in Fig. 1(a). In California coastal waters we find the type D call about as common as the type A-B call (A. Teranishi, unpublished data), though type D calls appear less common in mid-ocean re-

cordings. The character of the type D call can be quite variable, sometimes having characteristics intermediate with that of the type B call, and is not produced in such a regular pattern as the type A-B calls.

Figure 1(b) illustrates type A call variance between two animals. Most A calls have a weak tone near 16 Hz which precedes the pulsive part of the call by several seconds, and is followed by 19 to 23 pulses which are particularly apparent near 90 Hz. As previously noted (Thompson *et al.*, 1996), the component near 90 Hz is not a harmonic of the energy near 16 Hz. We encountered one animal producing A-B call pairs in which the type A calls had only five or six pulses. Beginning and ending peak frequencies are noted on the spectrogram of Fig. 1(b), the 90-Hz component typically shifting downward in frequency throughout the call. The average frequency and frequency shift for the 90-Hz component provides a measure of individual identification for each whale producing A-B calls in this encounter, though we do not suggest these differences will remain constant over longer time periods or that all individuals can be distinguished by such measures.

Both high- and low-frequency weak precursory components are commonly seen preceding the type B call when the signal-to-noise ratio is high [Figs. 2(a) and (b)]. In Fig. 2(a), a 10- to 12-Hz upsweep precedes the 16-Hz portion of the call. This precursor is relatively consistent in character (rate of change in frequency and amplitude), frequency, and duration when present, and has been reported previously (D'Spain *et al.*, 1995; Stafford *et al.*, 1999b). The precursory component near 400 Hz [Fig. 2(b)] ranges in different encounters, from about 300 to 500 Hz, from less than one to several seconds in duration, and shows significant variation in character. This component may be analogous to a 390-Hz component which preceded the last call segment in recordings of Chilean blue whales (Cummings and Thompson, 1971).

B. Acoustic activity

On three occasions during the October 1997 cruise sonobuoy arrays were deployed around animals which were biopsy sampled during the acoustic monitoring, and which were sufficiently separated from other whales to be visually tracked. On two of these occasions (12 and 16 October) the tracked animals did not produce any calls during the encounter. During the encounter of 15 October, one of the biopsied animals did produce calls.

On 15 October at 1652 local time, two blue whales (nos. 1 and 3 in Fig. 3) were sighted together in 800-m-deep water 31 km west of San Nicolas island. One of the RHIBs stayed with whale one for the next several hours collecting visual observations along with photo-identification and a genetic sample while the other RHIB obtained photo-identification on whale 3. Both of the visually observed animals were producing broadcast-type calls as was a third unseen animal (whale 2). Two more blue whales (nos. 4 and 5) could be heard in the area producing D calls. Five sonobuoys allowed acoustic tracking of the three animals producing the ''broadcast'' calls and general localization of the two animals producing D calls.

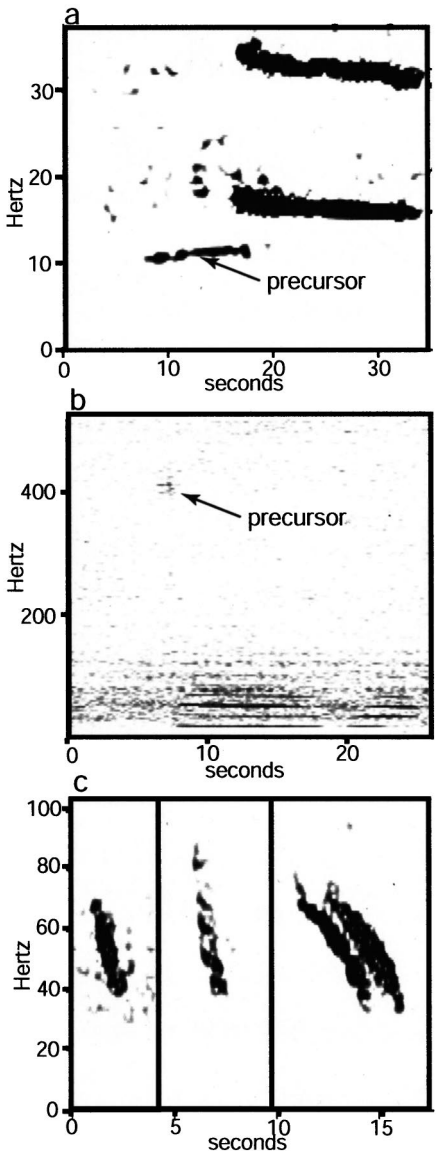


FIG. 2. Blue whale call spectrograms recorded with sonobuoys off California. (a) Onset of a type B call showing the common 10–12 Hz precursor. The FFT length is 2 s and overlap is 98% with a Hann window. (b) An example of the high-frequency precursor to the B call. The duration varies from less than one s to several seconds, the frequency varies by more than 100 Hz and the character is variable. This component is not always observed. The FFT length is 0.75 s and overlap is 75% with a Hann window. (c) A composite spectrogram of three separate recordings of type D calls illustrating the variability. The double call in the third example is considered to be two animals counter-calling rather than a propagation artifact. The FFT length is 0.74 s and overlap is 93.75% with a Hann window.

The tracking precision shown in Fig. 3 is highly variable, being as precise as 100 m for each call in the central portion of the track for animal 2 and as poor as providing bearing only on animal 1 when it was far from the array. Sonobuoy drift was less than 100 m during tracking, the largest position errors (~50 m) being GPS errors in the positions taken at buoy deployment time. The positions for each call from animal 1 are interpolated from GPS fixes in the final footprint of each surfacing. We are confident no surfacings were missed for this animal. Bearings from the DIFAR buoys matched the bearing to this animal, but the small array aperture in this direction prevents accurate range

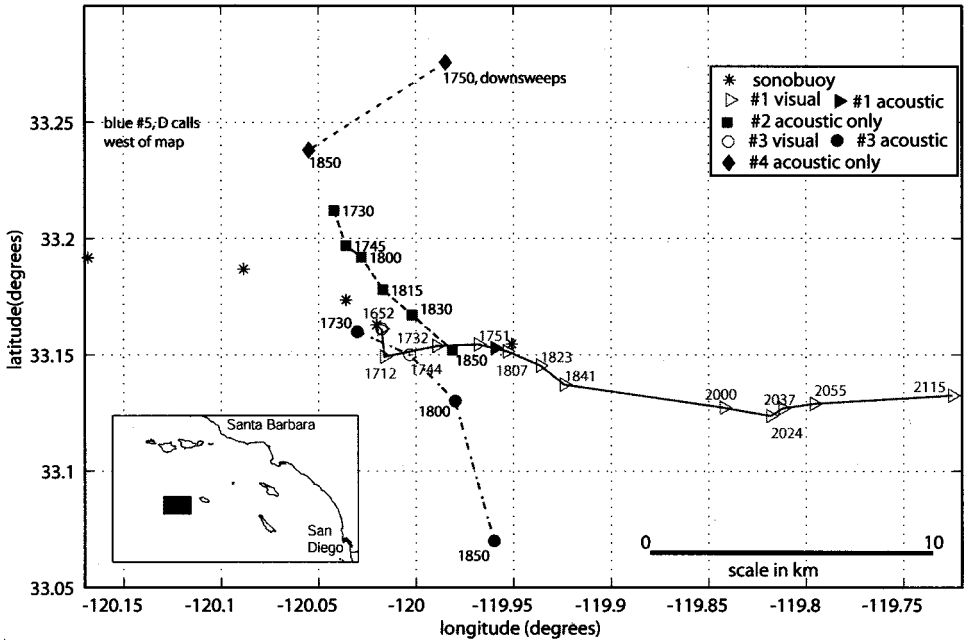


FIG. 3. Location map west of San Nicolas Island showing the positions and interpolated tracks for five blue whales from 1652 to 2115 local time 15 October 1997. Acoustic position errors are relatively small for whale 2 where the sonobuoy array aperture is excellent in the direction of that whale. When whale 1 is east of the sonobuoy array, the acoustic bearings are practically the same from all sonobuoys and acoustics alone cannot determine position.

estimation to this animal from acoustics alone. Sound pressure levels from animal one saturated the nearest buoy during one call near 1800 hours. The DIFAR bearings to each whale call, combined with the saturated signal during this one call, leaves no doubt as to the match between visual and acoustic data for this whale. Average swimming speeds were 5 km/h for animal 1, 7 km/hr for animal 2, and 9 km/hr for animal 3. Subtle but consistent differences in call character [illustrated in Fig. 1(b)] further identified the calls of each individual animal.

C. Call patterns

The calls from the five whales acoustically monitored on 15 October show no apparent coordination or interaction (Fig. 4). The visual observations of surface time coincide with gaps in the calling sequence for animal 1, with the possible exception of the last observed surfacing where there may be a small overlap with the onset of the type A call. A timing error of only 15 s in reading the wristwatch and recording our visual observation could account for this discrepancy. Such a match between calling gaps and visual surfacing intervals has been reported previously (Cummings and Thompson, 1971) although a possible exception has also been reported (Edds, 1982). A hypothetical model for sound production in the blue whale suggests the type B calls cannot be produced while at the surface (Aroyan *et al.*, 2000).

The most common call pattern for California blue whale broadcast calls is one type A call followed by one type B call, as produced by animals 1 and 3. The call pattern produced by animal 2, however, is one A call followed by a series of B calls. We have not seen type B calls which were not preceded by an A call, nor consecutive A calls. Each closely spaced call sequence starts with a type A call. From examination of the call timeline alone, we can see no direct evidence of communication between these whales as indicated by use of counter-calls or synchronized surfacing, though there can be little doubt these animals can hear each other. The calls from animals 4 and 5 are plotted as one timeline (Fig. 4) because it was not possible to distinguish which animal produced some of the calls though many of the calls could be attributed to one or the other of the two locations.

D. Acoustic source levels

The received sound pressure levels and implied source levels for the A-B paired broadcast calls from animal 1 are plotted as Figs. 5(a) and (b). The average source level is 178 dB *re*: 1 μPa at 1 m over the 10–110-Hz band for 82 ''A'' calls [Fig. 4(a)] and 186 dB *re*: 1 μPa at 1 m over the 10–110-Hz-band for 61 ''B'' call measurements [Fig. 4(b)], all from whale 1.

Because variability is apparent in the intensity ratios between the 16.5-Hz tone and the third harmonic or 50-Hz tone of the ''B'' calls, these ratios were examined to evaluate if the whale may be controlling the relative intensities of the harmonics. The spectrum level of the third harmonic is on average 10.3 dB lower than the fundamental over 55 calls, with a range of 16 to 0 dB lower. This ratio of fundamental and third harmonic for the same call observed on different sonobuoys shows nearly a 16-dB variability, indicating the variability is primarily due to propagation effects rather than changes in the sound production mechanism at the whale. The source and receiver depths are expected to have a profound effect on such frequency-dependent propagation effects (Urick, 1983). Surface reflection interference affects each frequency component differently and could potentially be used to estimate the depth of the calling whale.

As with the type B calls, the A calls also have most of the sound intensity in the lowest frequency portion of the call, though the 90-Hz portion often has a higher signal-to-noise ratio because of observed lower ocean ambient noise levels at 90 Hz. The spectrum level of the component near 17 Hz averages 2.5 dB lower than the 10–110-Hz band level while the spectrum level of the component near 90 Hz is 14

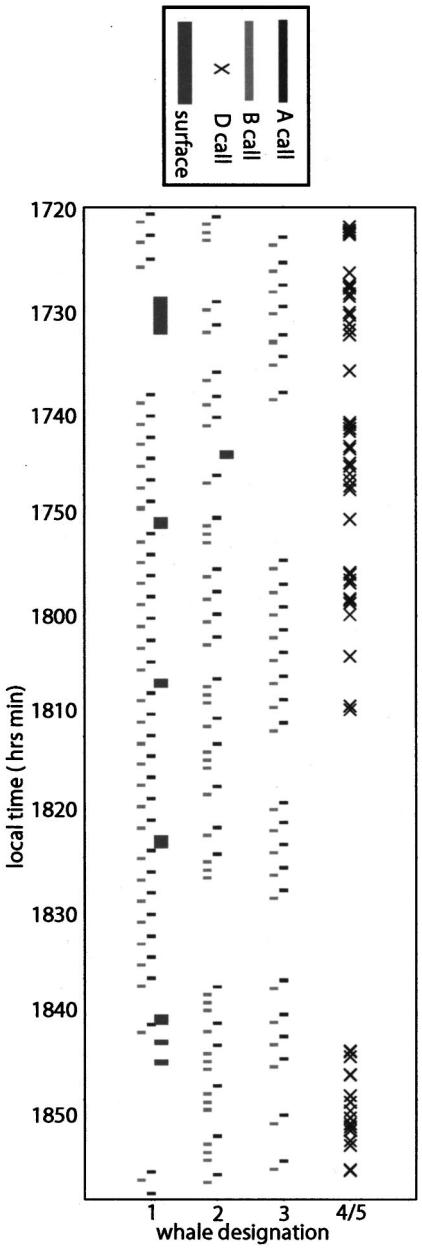


FIG. 4. The acoustic call patterns for five blue whales producing type A, B, and D calls, as well as visual observations of whales surface times. Whales 4 and 5, which produced only type D calls, could not always be separated and thus are plotted together. Note how whale 2 often follows a type A call with three type B calls. The visual observations of surface times correspond to gaps in calling.

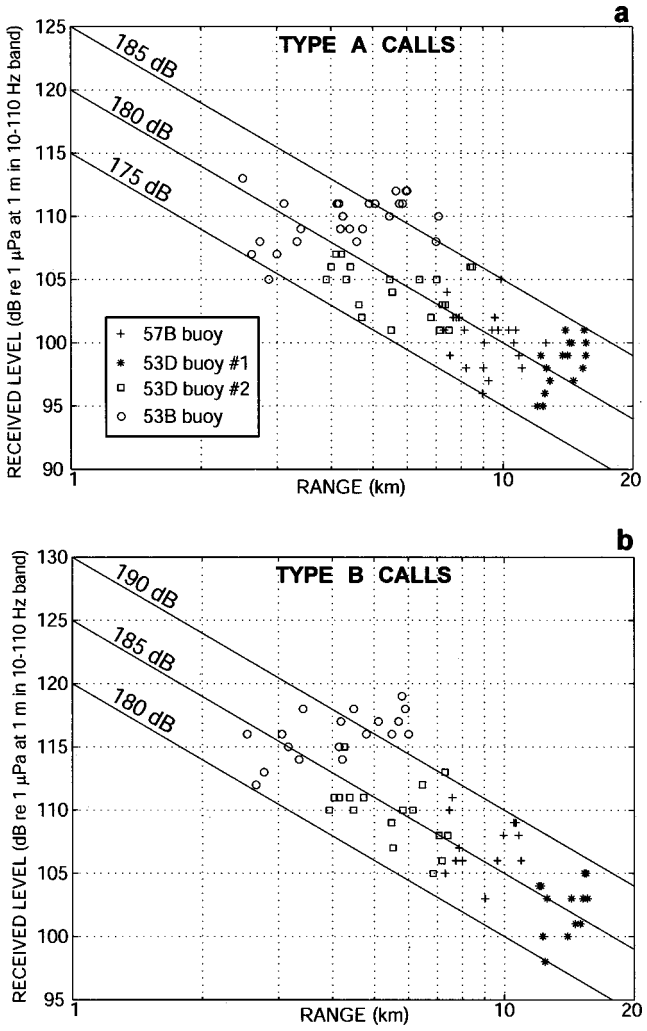


FIG. 5. The received sound pressure levels and inferred source levels for blue whale 1 on 15 October as measured over the 10–110-Hz band *re*: 1 μPa at 1 m. The levels were determined using a 10-s average, starting 1 to 2 s after the onset of each call, using a Hann window, 80% overlap and a 1 s frame length. Source level calculations use spherical spreading losses. The scatter in the data may be attributed to propagation variability, rather than actual differences in source level. The four symbols represent the four sonobuoys used.

dB lower than the band level. Examination of the clearest recordings of the pulsive type ''A'' calls show acoustic energy at five higher frequencies up to 110 Hz, none of which appear to be harmonics.

E. Gender, behavior, and history of acoustically monitored whales

A single female whale was acoustically monitored from 1308 to 1410 local time on 12 October (catalog ID #1323) at 33° 33′ N 119° 46′ W. She was traveling slowly and consistently in a NW direction. During the 47-min visual observation period prior to when the animal was struck with the biopsy dart the animal traveled a straight-line distance of 3.2 km. There was no detectable reaction to the biopsy hit. This animal had not been previously photo-identified. We are confident no calls were produced by this animal during the time in which it was acoustically monitored.

Whale 1 of the 15 October encounter was determined to be male. This large whale was initially spotted traveling in the lead of a second slightly smaller whale, although these two quickly separated. The whale traveled consistently in an easterly direction on a fairly direct course. Dive intervals and spacing were longer than typical for blue whales in this area. Most intervals ranged from 14 to 17 min and were interspersed with surface intervals where the animal would surface to breath repeatedly four to five times in close succession. The whale traveled a straight line distance of 34.2 km from first sighting at 1605 and last position at 2215. This animal (Cascadia Research catalog ID #673) had been seen seven times prior to the encounter described. These included six encounters between 27 June and 16 September 1992 with June and July sightings in the Santa Barbara Channel and a

September sighting in the Gulf of the Farallones. There was one sighting on 29 September 1993 in Monterey Bay. This whale was acoustically monitored for 5 h and 25 min and we are confident the broadcast type calls were produced by this animal.

A female whale was acoustically monitored for 20 min from 1205 to 1225 local time on 16 October (catalog ID #170) at 34° 52′ N 120° 54′ W. This was a large whale traveling very slowly consistently in a southerly direction. Dive intervals were fairly short (under 4 min) and the animal appeared to be remaining fairly shallow between dives. A biopsy sample was taken midway through the 39-min visual observation period and the animal made a quick dive apparently in response to the biopsy. This whale had been seen seven times prior to the encounter described. The earliest sighting was 1 September 1987 in the Gulf of the Farallones indicating this animal is >10 years old. Three sightings were made in June 1992 in the Santa Barbara Channel. This animal was also seen three times between 12 and 23 July 1997 in the Santa Barbara Channel. We are confident no calls were produced by this animal during the time it was acoustically monitored.

IV. DISCUSSION

A. Call activity and gender

The broadcast calls are often produced in a continuous pattern for many hours at a time, with pauses appropriate for surface breathing intervals as suggested for blue whales elsewhere (Cummings and Thompson, 1971; Edds, 1982) and by the combined visual and acoustic data presented here. The amplitude, duration, and repetitive nature of the broadcast call makes it well suited for long distance signaling. During observation periods of a few hours, an individual whale usually produces only one type of call, either the type D call (A. Teranishi, unpublished data) or the broadcast call (Stafford *et al.*, 1998). Only on a few occasions have blue whales been observed to mix these two call types (M. McDonald, unpublished data) (Thode *et al.*, 2000). Whales producing the broadcast call are often traveling at modest speeds (McDonald *et al.*, 1995; Stafford *et al.*, 1998; Tyack, 1998).

When multiple blue whales are producing the broadcast call in the same area, this study finds no evidence of coordination between the callers, as might be suggested by synchronized respiration intervals or patterns in call behavior among animals. The type D call, however, appears to be used as a counter-call among multiple whales. Type D calls sometimes occur in overlapping pairs, each produced by a different animal, and separated by relatively long intervals [Fig. 2(c)].

Call behavior of fin and humpback whales has been better studied than blue whales and provide a reference with regard to how call types may vary with gender. There may be analogies between the blue whale broadcast call and the fin whale doublet call, and between the blue whale D call and the fin whale irregular call. Acoustic evidence suggests widely separated calling fin whales synchronize respiration intervals and counter-call when using irregular call types, but do not show such coordination when producing the more regular doublet call type (McDonald *et al.*, 1995; Watkins *et al.*, 1987). It has been argued that only the male fin whale produces the doublet call as a breeding display (Watkins *et al.*, 2000), consistent with our finding of a male blue whale producing the broadcast call. The humpback song, which is produced only by males, may be analogous to the fin whale doublet call and the blue whale broadcast call.

B. Source levels

The average blue whale call source levels of 186 dB reported here for the type B call are measured in units comparable to the 188 dB levels previously reported for blue whales off Chile (Cummings and Thompson, 1971). The source level in our blue whale recordings is determined primarily by the fundamental tone, as it is sufficiently stronger than the harmonics and overtones to dominate the signal. The pressure spectrum level of the fundamental tone is therefore nearly equivalent to the band level and the band over which the level is measured is of relatively little consequence. One other calibrated recording of blue whales off California (Thode *et al.*, 2000) reports source levels in units of pressure spectral density over a 0.4-Hz bandwidth. The conversion of these units to either band or pressure spectrum levels for comparison results in an average level slightly greater than 180 dB for the type B call, notably lower than our results or those of Cummings and Thompson (1971). Knowledge of source levels is important in studies of the effects of man-made noise and in determining zones of masking which limit the communication potential of whales (Richardson *et al.*, 1995).

V. SUMMARY

We describe three blue whale encounters in which biopsies were obtained to determine gender and acoustic monitoring was in place to determine if the whales were calling. The observed travel and calling patterns are described for a group of blue whales including one known male producing broadcast calls. Two noncalling females were observed in separate comparatively brief encounters. We describe a previously unpublished characteristic of some blue whale broadcast calls, a 400-Hz precursor to the type B call. The source level of blue whale calls from one animal is found to average 186 dB for the type B, and 178 dB for the type A calls *re*: 1 μPa at 1 m over the 10–110-Hz band. Observed variability in intensity ratios of the third harmonic to the fundamental within the blue whale call is considered to be a propagation artifact. With further work we hope to determine if a gender bias is present in the production of the blue whale broadcast call and to quantify the blue whale acoustic detection function for purposes of abundance estimation. Such work could best be done by visually following acoustically monitored and biopsied animals. Tagging animals with recorders may also be useful, though visual contact may be necessary to determine if a second animal is nearby which could potentially produce calls indistinguishable from those of the tagged animal.

ACKNOWLEDGMENTS

Funding for components of this research was provided by National Marine Fisheries Service, Southwest Fisheries Science Center, the Oceanic Society, Olympic Coast National Marine Sanctuary, and by the Whale Adoption Project, a joint program of Cascadia Research. A number of individuals helped arrange support for this work including Jay Barlow, Ed Bowlby, and Todd Jacobs. The observers for this cruise were Lisa Schlender, Greg Falxa, Rene DeVito, Wade Gerdee, Annie Douglas, Darcy Bristow, Emily Walton, Sherwin Cotler, and Cherish Morrison-Price. The crew and officers of the NOAA ship *McArthur* assisted in getting identification photographs. Lisa Schlender, Kristin Rasmussen, Heather Medic, Hannah Smith, Annie Douglas, Nicole Stagmar, Wade Jerdee, and Shannon Wilhite printed and matched identification photographs and assisted in data compilation and analysis. We also appreciate the several benefactors who provided invaluable supplemental support.

Aroyan, J. L., McDonald, M. A., Webb, S. C., Hildebrand, J. A., Clark, D., Laitman, J. T., and Reidenberg, J. S. (**2000**). ''Acoustic Models of Sound Production and Propagation,'' in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. N. Fay (Springer, New York), pp. 409–469.

Baker, C. S., Lambertson, R. H., Weinrich, N. T., Calambokidis, J., Early, G., and O'Brien, S. J. (**1991**). ''Molecular genetic identification of the sex of humpback whales (*Megaptera novaeangliae*),'' in *Genetic Ecology of Cetaceans*, International Whaling Commission special issue 13, edited by R. Hoelzel, pp. 105–111.

Barlow, J. (**1995**). ''The abundance of cetaceans in California waters. Part 1: Ship surveys in summer and fall of 1991,'' Fish. Bull. **93**, 1–14 plus errata.

Calambokidis, J., Chandler, T., Rasmussen, K., Steiger, G. H., and Schlender, L. (**1999**). ''Humpback and blue whale photographic identification research off California, Oregon and Washington in 1998,'' Final report to Southwest Fisheries Science Center, available from Cascadia Research, 218 $\frac{1}{2}$ W. Fourth Ave., Olympia, WA 98501 (35 pp.).

Calambokidis, J., Steiger, G. H., Cubbage, J. C., Balcomb, K. C., Ewald, C., Kruse, S., Wells, R., and Sears, R. (**1990**). ''Sightings and movements of blue whales off Central California 1986–1988 from photo-identification of individuals,'' Report of the International Whaling Commission (special issue 12), pp. 343–348.

Chapman, N. R., and Cornish, J. W. (**1993**). ''Wind dependence of deep ocean ambient noise at low frequencies,'' J. Acoust. Soc. Am. **93**, 782–789.

Clark, C. W., and Charif, R. A. (**1998**). ''Acoustic Monitoring of large whales to the west of Britain and Ireland using bottom-mounted hydrophone arrays, October 1996–September 1997,'' Peterborough, Joint Nature Conservation Committee, Report #281, ISSN 0963-8091.

Cummings, W. C., and Thompson, P. O. (**1971**). ''Underwater sounds from the blue whale (*Baleanoptera musculus*),'' J. Acoust. Soc. Am. **50**, 1193–1198.

Curtis, K. R., Howe, B. M., and Mercer, J. A. (**1999**). ''Low-frequency ambient sound in the North Pacific: Long time series observations,'' J. Acoust. Soc. Am. **106**, 3189–3200.

D'Spain, G. L. (**1994**). ''Relationship of Underwater Acoustic Intensity Measurements to Beamforming,'' Can. Acoust. **22**, 157–158.

D'Spain, G. L., Hodgkiss, W. S., Edmunds, G. L., Nickles, J. C., Fisher, F. H., and Harris, R. A. (**1992**). ''Initial Analysis of the data from the Vertical DIFAR Array: Mastering the Oceans through Technology,'' I.E.E.E. OCEANS 92, Newport, Rhode Island, conference proceedings, pp. 346–351.

D'Spain, G. L., Kuperman, W. A., Hodgkiss, W. S., and Berger, L. P. (**1995**). ''3-D localization of a blue whale,'' MPL technical memorandum 447, San Diego, Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego.

Edds, P. L. (**1982**). ''Vocalizations of the blue whale, *Balaenoptera musculus*, in the St. Lawrence River,'' J. Mammal. **63**, 345–347.

Edds-Walton, P. L. (**1997**). ''Acoustic Communication Signals of Mysticete Whales,'' Bioacoustics **8**, 47–60.

Fiedler, P. C., Reilly, S., Hewitt, R. P., Demer, D., Philbrick, V. A., Smith, S., Armstrong, W., Croll, D. A., Tershy, B. R., and Mate, B. R. (**1998**). ''Blue Whale Habitat and Prey in the Channel Islands,'' Deep-Sea Res., Part II **45**, 1781–1801.

Heard, G. J., McDonald, M., Chapman, N. R., and Jaschke, L. (**1997**). ''Underwater Light Bulb Implosions: A Useful Acoustic Source,'' Oceans 97, Conference Proceedings, Honolulu, pp. 755–762.

Lambertsen, R. H. (**1987**). ''A biopsy system for large whales and its use for cytogenetics,'' J. Mammal. **68**, 443–445.

Mate, B. R., Lagerquist, B. A., and Calambokidis, J. (**1999**). ''Movements of North Pacific blue whales during the feeding season off southern California and southern fall migration,'' Marine Mammal Sci. **15**, 1246–1257.

McDonald, M. A., Hildebrand, J. A., and Webb, S. C. (**1995**). ''Blue and fin whales observed on a seafloor array in the Northeast Pacific,'' J. Acoust. Soc. Am. **98**, 712–721.

Richardson, W. J., Greene, Jr., C. R., Malme, C. I., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, San Diego), p. 576.

Stafford, K. M., Fox, C. G., and Clark, D. (**1998**). ''Long-range acoustic detection, localization of blue whale calls in the northeast Pacific Ocean,'' J. Acoust. Soc. Am. **104**, 3616–3625.

Stafford, K. M., Fox, C. G., and Nieukirk, S. L. (**1999a**). ''Low-frequency whale calls recorded on hydrophones moored in the eastern tropical Pacific,'' J. Acoust. Soc. Am. **106**, 3687–3698.

Stafford, K. M., Nieukirk, S. L., and Fox, C. G. (**1999b**). ''An acoustic link between blue whales in the Eastern Tropical Pacific and the Northeast Pacific,'' Marine Mammal Sci. **15**, 1258–1268.

Thode, A. M., D'Spain, G. L., and Kuperman, W. A. (**2000**). ''Matched field processing, geoacoustic inversion, and source signature recovery of blue whale vocalizations,'' J. Acoust. Soc. Am. **107**, 278–286.

Thompson, P. O., Findlay, L. T., and Vidal, O. (**1996**). ''Underwater Sounds of Blue Whales, *Balaenoptera musculus*, in the Gulf of California, Mexico,'' Marine Mammal Sci. **12**, 293–296.

Tyack, P. L. (**1998**). ''Acoustic communication under the sea,'' in *Animal Acoustic Communication*, edited by S. L. Hopp, M. J. Owren, and C. S. Evans (Springer, Berlin), pp. 163–220.

Tyack, P. L. (**2000**). ''Functional Aspects of Cetacean Communication,'' in *Cetacean Societies: Field Studies of Dolphins and Whales*, edited by J. Mann, R. C. Conner, P. L. Tyack, and H. Whitehead (Univ. of Chicago, Chicago), pp. 270–307.

Urick, R. J. (**1983**). *Principles of Underwater Sound* (McGraw-Hill, New York).

Watkins, W. A., Daher, M. A., Repucci, G. M., George, J. E., Martin, D. L., DiMarzio, N. A., and Gannon, D. P. (**2000**). ''Seasonality and Distribution of Whale Calls in the North Pacific,'' Oceanography **13**, 62–67.

Watkins, W. A., Tyack, P., Moore, K. E., and Bird, J. E. (**1987**). ''The 20-Hz signals of finback whales (*Balaenoptera physalus*),'' J. Acoust. Soc. Am. **82**, 1901–1912.

Wilcock, W. S. D., and Toomey, D. R. (**1991**). ''Estimating Hypocentral Uncertainties for Marine Microearthquake Surveys: A Comparison of the Generalized Inverse and Grid Search Methods,'' Mar. Geophys. Res. **13**, 161–171.

LETTERS TO THE EDITOR

This Letters section is for publishing (a) brief acoustical research or applied acoustical reports, (b) comments on articles or letters previously published in this Journal, and (c) a reply by the article author to criticism by the Letter author in (b). Extensive reports should be submitted as articles, not in a letter series. Letters are peer-reviewed on the same basis as articles, but usually require less review time before acceptance. Letters cannot exceed four printed pages (approximately 3000–4000 words) including figures, tables, references, and a required abstract of about 100 words.

Surface acoustic impedance and causality

Yves H. Berthelot[a)]
Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0405

(Received 1 April 2000; revised 15 December 2000; accepted 5 January 2001)

Several models that are commonly used to calculate the surface acoustic impedance of porous materials are examined in light of the causality requirement in its nonlocal and its approximate local forms. Strict (nonlocal) causality requires that the imaginary part of the surface acoustic impedance be the Hilbert transform of the real part. It is found that the local approximation is not a good test of causality for these models and that, even in its general form, the equation describing the surface acoustic impedance of these porous materials is not strictly causal at high frequencies. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1352089]

PACS numbers: 43.28.En [ANN]

I. INTRODUCTION

Causality is a powerful tool to analyze lossy, linear systems. The Kramers–Krönig relations, initially derived for causal optical systems,[1] have been used in acoustics to show that attenuation and dispersions of waves propagating in a lossy medium are in fact intrinsically connected.[2,3] In the context of surface acoustic impedance, they have been used by Miki[4] to modify the Delany–Bazley empirical power laws[5] to satisfy causality, and by Daigle and Stinson[6] to deduce the imaginary part of the ground reflection coefficient from phase gradient measurements of the magnitude at low frequencies. In this letter, the local and nonlocal version of the Kramers–Krönig relations are used to test the causality of several models of acoustic impedance commonly used to describe propagation in porous media.

II. BACKGROUND

The surface acoustic impedance of a material is defined as the ratio of the complex amplitudes of the pressure, $\hat{P}(\omega)$, to the normal velocity component into the surface, $\hat{V}_{IN}(\omega)$, both evaluated at the surface, S,

$$Z(\omega)=\left[\frac{\hat{P}(\omega)}{\hat{V}_{IN}(\omega)}\right]_S=R(\omega)+iX(\omega), \tag{1}$$

where the resistance, R, and reactance, X, are both positive quantities if the harmonic dependence $e^{-i\omega t}$ is chosen.[7] It follows that $\hat{V}_{IN}(\omega)=\hat{P}(\omega)Y(\omega)$, where $Y(\omega)$ is the surface admittance defined by

$$Y(\omega)=\int_{-\infty}^{+\infty} y(t)e^{i\omega t}\,dt, \tag{2}$$

where $y(t)$ is a real valued function representing the physical time response of the surface to an impulse excitation. If the incident pressure strikes the surface at time $t=0$, causality requires that the (passive) surface response $y(t)$ be zero for $t<0$, so that

$$\begin{aligned} Y(\omega)&=\int_0^{\infty} y(t)e^{i\omega t}\,dt \\ &=\int_0^{\infty} y(t)\cos(\omega t)dt+i\int_0^{\infty} y(t)\sin(\omega t)dt. \end{aligned} \tag{3}$$

Since

$$\sin(\omega t)=-HT_{\omega}[\cos(\omega t)]=-\frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{d\omega'}{\omega'-\omega}\cos(\omega t), \tag{4}$$

where HT_{ω} denotes the Hilbert transform with respect to frequency (the integral is taken to be the Cauchy principal value[8]), Eq. (3) indicates that the real and imaginary part of the surface admittance are related by a Hilbert transform. The same is true for the impedance.[9,10] This result is quoted, for instance, in Nussenzweig[11] as Titchmarsh's theorem[12] or as the first and second Plemelj formulas:

$$R(\omega)=\frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{X(\omega')}{\omega'-\omega}d\omega' \tag{5}$$

and

a)Electronic mail: yves.berthelot@me.gatech.edu.

$$X(\omega) = -\frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{R(\omega')}{\omega'-\omega}d\omega'.$$

In other words, the imaginary part of the surface impedance is the convolution[10] (understood as principal value) of its real part with the function $-1/\pi\omega$. Equation (5) can also be expressed over positive frequencies only, in which case it reduces to the standard form of Kramers–Krönig relations[3,13,14]

$$R(\omega) = \frac{2}{\pi}\int_0^{\infty}\frac{\omega' X(\omega')}{\omega'^2-\omega^2}d\omega'$$

and (6)

$$X(\omega) = -\frac{2}{\pi}\int_0^{\infty}\frac{\omega' X(\omega')}{\omega'^2-\omega^2}d\omega'.$$

O'Donnell *et al.*[3] have applied this formalism to find a relationship between ultrasonic attenuation and phase velocity. In the process, they have derived a very useful approximation to Eq. (6) for the case where the functions $R(\omega)$ and $X(\omega)$ are slowly varying functions of frequency. This so-called local approximation is

$$X(\omega) = -\frac{\pi}{2}\omega\frac{dR}{d\omega}. \tag{7}$$

In Ref. 3, Eq. (7) yields results that are indistinguishable for the nonlocal Kramers–Krönig relations, Eq. (5). In this letter we evaluate some common models of surface acoustic impedance of porous media in terms of local [Eq. (7)] or nonlocal [Eq. (5)] causality conditions.

III. MODELS OF POROUS MEDIA

The particle velocity of the gas inside a perfectly cylindrical circular pore can be described in terms of a complex density[15]

$$\rho(\omega) = \rho_0\left(\frac{1}{1-\Phi(\sqrt{i\omega\tau})}\right), \tag{8}$$

where ρ_0 is the density of the gas, and where $\Phi(x) = (2/x)\times[J_1(x)/J_0(x)]$, J_i being the cylindrical Bessel function of first kind and order i. The time constant τ is a^2/v, a being the pore radius (assumed to be much smaller than the wavelength), and $v = \mu/\rho_0$ is the kinematic viscosity of the gas. In Eq. (8), it is assumed that acoustic motion is in one direction, and that the effects of viscosity and heat conduction can be decoupled. Based upon this model, Attenborough has derived[16,17] an expression for the acoustic impedance of a porous material with arbitrary pore shapes. The result can be written as

$$Z_s = \rho_0 c_0\left(\frac{q}{\Omega}\right)[1-\Phi(\sqrt{i\omega\tau})]^{-1/2} \times[1+(\gamma-1)\Phi(\sqrt{i\omega N\tau})]^{-1/2}, \tag{9}$$

where the time scale is now $\tau = [(2\rho_0 q^2)/(\sigma_e\Omega^2)]$. The quantity q^2 is defined as the tortuosity, Ω is the porosity, and σ_e is the effective flow resistivity in mks-Rayls/m, defined as where sf is the pore shape factor ratio. [In Ref. 18, Miki uses a different definition of the effective flow resistivity, σ_e', and his equation for the acoustic impedance is identical with Eq. (9) above if one writes $\tau = [(8\rho_0)/\sigma_e']$, where $\sigma_e' = 4\sigma_e(\Omega/q)^2$.] In Eq. (9), N is the Prandtl number, and γ is the ratio of specific heats of the gas of characteristic impedance $\rho_0 c_0$. In the low frequency limit,[16] Eq. (9) reduces to what is often referred to as the Attenborough model,[17,19] and which is rewritten here as

$$Z_s(\omega) = \rho_0 c_0 B\left[\frac{1-\frac{1}{i\omega\tau_1}}{\sqrt{1-\frac{1}{i\omega\tau_2}}}\right], \tag{10}$$

with $B = A/\sqrt{\gamma a T}$, $A = 4q^2/3\Omega$, $a = \frac{4}{3}-[(\gamma-1)/\gamma]N$, where $\tau_1 = \tau/6$ and $\tau_2 = 3a\tau/4$, and τ has been defined previously. In the very low frequency limit, Eq. (10) reduces to Eq. (11) below and the surface acoustic impedance is described in terms of single parameter (the effective flow resistivity) according to

$$Z_s = \frac{\rho_0 c_0}{\sqrt{F}}(1+i), \tag{11}$$

where $\overline{F} = (\rho_0 f/\sigma_e)$ is a nondimensional frequency.

More recently, Allard and Champoux[20] proposed new equations for the surface impedance of porous media in terms of a dynamic density, $\rho(\omega)$, and a dynamic bulk modulus, $K(\omega)$. The dynamic density is given in terms of the function $F(\omega) = \sqrt{1-i\omega\tau'}$ introduced by Johnson *et al.*[21] in a heuristic manner to describe the dynamic tortuosity of the medium. The function F was chosen because it is simple and it satisfies the expected low and high frequency limits. Using this function F, Allard and Champoux have proposed a new model[20] for the acoustic impedance of a porous material, which is rewritten below as

$$Z_s(\omega) = \rho_0 c_0\left[\frac{1-\Psi(\omega)}{\gamma-\frac{\gamma-1}{\Psi(4N\omega)}}\right]^{1/2}, \tag{12}$$

where $\Psi(\omega) = (\sqrt{1-i\omega\tau_0})/(2i\omega\tau_0)$, with $\tau_0 = \rho_0/2\sigma$. Here, the flow resistivity σ is related to the effective flow resistivity σ_e by $\sigma = \sigma_e(4\Omega/s_f^2)$, where s_f is the pore shape factor ratio. In Eq. (3) of Ref. 20, it is implicitly assumed that the pore shape factor, the tortuosity, and the porosity are unity, in which case the dynamic density is $\rho_0[1-\Psi(\omega)]$. By comparing with Eq. (8), it is clear that the approximation consists in replacing $1/(1-\Phi)$ by $(1-\Psi)$. One also notes that, by using the low frequency limit of Eq. (12) with an isothermal incompressibility of air, one obtains Attenborough's very low frequency model, Eq. (11). In the next section, we examine the causality of these standard impedance models [Eqs. (9)–(12)] in view of the local and nonlocal causality conditions, Eqs. (7) and (5), respectively. Equations (9), (10), and (12) represent impedance models for rigid-porous absorbers in terms of characteristic time scales τ_i. As pointed out by Wilson,[22] these time scales correspond to re-

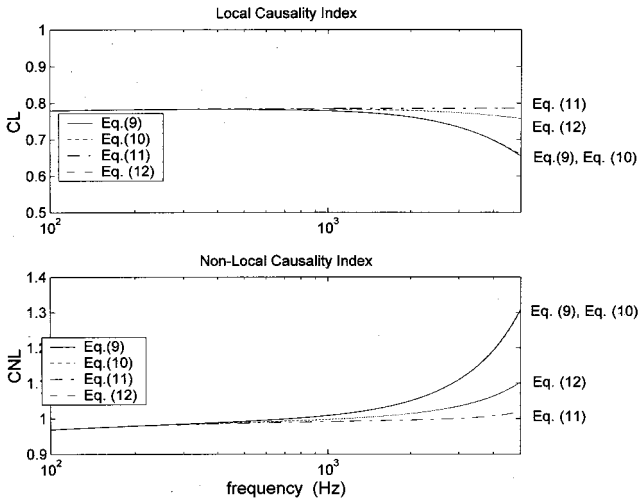


FIG. 1. (a) Local causality index defined by Eq. (13) as a function of frequency; (b) nonlocal causality index defined by Eq. (14), based on the Hilbert transform, as a function of frequency. The causality indices are evaluated for four different impedance formulations [Eqs. (9)–(12)], with $\sigma_e = 300\,000$ mks rayls/m, $\Omega = 0.3$, $T = 1.2$, $s_f = 1$.

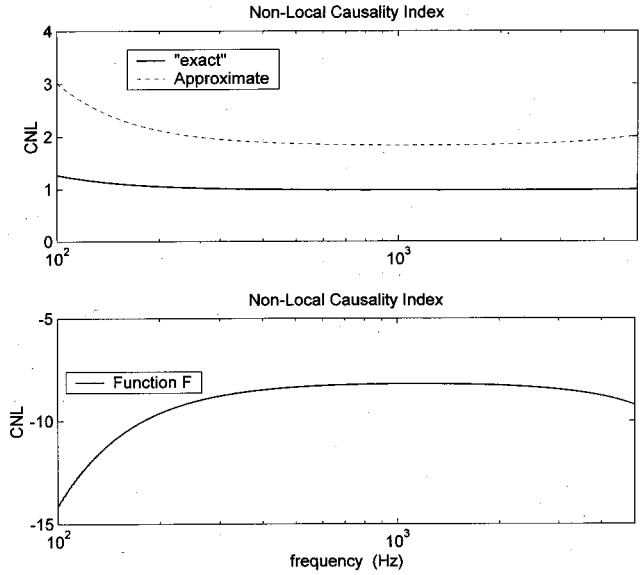


FIG. 2. (a) Nonlocal causality index for the response of the medium. Solid line: ''exact'' response $(1-\Phi)/i\omega$; dotted line: approximate solution based on $[i\omega(1-\Psi)]^{-1}$, with Φ and Ψ defined by Eqs. (8) and (12). (b) Nonlocal causality index of the function $F=\sqrt{1-i\omega\tau}$.

laxation times for the vorticity and entropy modes in the material.

IV. CAUSALITY

Based on Eqs. (5) and (7), one can define a causality index (for the local approximation) C_L such that

$$C_L = \frac{-\frac{\pi}{2}\omega\frac{d}{d\omega}[\mathrm{Re}(Z_s)]}{\mathrm{Im}(Z_s)} \tag{13}$$

and a causality index (in the nonlocal form), C_{NL}, such that

$$C_{NL} = \frac{HT[\mathrm{Re}(Z_s)]}{\mathrm{Im}(Z_s)}. \tag{14}$$

A causality index of unity indicates that the model for the impedance Z_s is causal. The causality indices C_L and C_{NL} are shown in Figs. 1(a) and (b), respectively, from 100 Hz to 5000 Hz. In each case, the solid line represents calculations based on the ''exact'' model [Eq. (9)]; the dotted line is based on Attenborough low frequency model [Eq. (10)]; the chain-dotted line represents the very low frequency approximation [Eq. (11)]; and the dashed line is obtained by using the Allard-Champoux model [Eq. (12)]. The material is assumed to have an effective flow resistivity $\sigma_e = 300\,000$ mks rayls/m, a porosity $\Omega = 0.3$, a tortuosity $T = 1.2$, and a pore shape factor ratio s_f equal to unity. With these parameters, the frequency at which $\omega\tau$ equals unity is about 1480 Hz.

Figure 1(a) reveals that the local causality index for the very low frequency approximation [Eq. (11)] is a constant equal to $\pi/4$, a result which can easily be derived analytically. Figure 1(b), on the other hand, indicates that the nonlocal causality index for this model [Eq. (11)] is unity (within the numerical error associated with the computation of the Hilbert transform), a result that could be expected on the basis that the Hilbert transform of $|f|^{-1/2}$ is also $|f|^{-1/2}$, i.e., that Eq. (11) is a low frequency causal model.[10] The departure from unity of the local causality index is a clear indication that the local version of the causality test [Eq. (7)] is not an appropriate test of causality, at least in this case. This supports Szabo's results[23] and the subsequent discussion by Waters *et al.*[24] for the case where the attenuation in the material does not follow a linear frequency dependence.

Figure 1 also indicates that a high frequencies equations [(9), (10), and (12)] do not satisfy causality, either locally or nonlocally. It is not surprising that Eq. (10) does not satisfy causality at high frequencies since it is a low frequency/high flow resistivity model. However, Eqs. (9) and (12) both show departure from causality at high frequencies, even though the response of the medium, which is proportional to $[i\omega\rho(\omega)]^{-1} = (1-\Phi)/(i\omega)$, with $\rho(\omega)$ defined in its ''exact'' form by Eq. (8), is causal (nonlocally) at all frequencies as shown in Fig. 2(a). Departure from causality in Eq. (9) is therefore due to the term that represents the complex adiabatic compressibility, i.e., the second bracket in Eq. (9). This term was derived from the analysis of the net heat flow per unit mass within a single circular cylindrical pore containing a heat-conducting, inviscid fluid with the assumptions of uniform pressure across the pore cross section and purely transverse heat flow. Figure 2(a) also shows the nonlocal causality index of the response of the medium for the approximate solution of Allard–Champoux. Here, the response is proportional to $[i\omega(1-\Psi)]^{-1}$ since the ''exact'' function $(1-\Phi)$ is now approximated by $1/(1-\Psi)$, where the function Ψ, defined with Eq. (12), is based on Johnson *et al.*'s heuristic function F. The dotted line in Fig. 2(a) indicates that the response of the medium for this approximate solution is not causal.

It is also interesting to note that, in the asymptotic limits, neither Johnson *et al.*'s function $F(\omega)$ nor the dynamic tortuosity $\alpha(\omega)$ satisfy local causality. For instance, in the high

frequency limit, $F(\omega) \rightarrow \sqrt{\omega\tau'/2}(1-i)$, and therefore the local causality index is $\pi/4$. In the low frequency limit, $\alpha(\omega)$ [and the dynamic density $\rho(\omega)$], are of the form $A[1-(i\omega\tau)^{-1}]$ where A is a real constant, a form which leads to a local causality index of zero. In fact, Fig. 2(b) indicates that Johnson *et al.*'s function $F(\omega)$ is not causal.

[1] R. Krönig and H. A. Kramers, ''Absorption and dispersion in X-ray spectra,'' Z. Phys. **48**, 174 (1928).
[2] M. O'Donnell, E. T. Jaynes, and J. G. Miller, ''General relationships between ultrasonic attenuation and dispersion,'' J. Acoust. Soc. Am. **63**, 1935 (1978).
[3] M. O'Donnell, E. T. Jaynes, and J. G. Miller, ''Kramers–Krönig between ultrasonic attenuation and phase velocity,'' J. Acoust. Soc. Am. **69**, 696–701 (1981).
[4] Y. Miki, ''Acoustical properties of porous materials—modifications of Delany–Bazley models,'' J. Acoust. Soc. Jpn. (E) **1**, 19–24 (1990).
[5] M. A. Delany and E. N. Bazley, ''Acoustic properties of fibrous absorbent materials,'' Appl. Acoust. **3**, 105–116 (1970).
[6] G. A. Daigle and M. R. Stinson, ''Impedance of grass-covered ground at low frequencies measured using a phase difference technique,'' J. Acoust. Soc. Am. **81**, 62–68 (1987).
[7] A. D. Pierce, *Acoustics: An Introduction to its Physical Principles and Applications* (McGraw-Hill, New York, 1981), p. 107.
[8] A. D. Pierce, *Acoustics: An Introduction to its Physical Principles and Applications* (McGraw-Hill, New York, 1981), see discussion page 136.
[9] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), pp. 372–373.
[10] R. N. Bracewell, *The Fourier Transform and its Applications*, 2nd ed. (McGraw-Hill, New York, 1986), pp. 359–365.
[11] H. m. Nussenzweig, *Causality and Dispersion Relations* (Academic, New York, 1972), pp. 21–28.
[12] E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals*, 2nd ed. (Oxford University Press, London, 1948).
[13] A. B. Bhatia, *Ultrasonic Absorption* (Dover, New York, 1967), pp. 90–91.
[14] A. I. Beltzer, *Acoustics of Solids* (Springer-Verlag, New York, 1988).
[15] C. Zwikker and C. W. Kosten, *Sound Absorbing Materials* (Elsevier, Amsterdam, 1949).
[16] K. Attenborough, ''Acoustical characteristics of rigid fibrous absorbent and granular materials,'' J. Acoust. Soc. Am. **73**, 785–799 (1983).
[17] K. Attenborough, ''Acoustical impedance models for outdoor ground surfaces,'' J. Sound Vib. **99**, 521–544 (1985).
[18] Y. Miki, ''Acoustical properties of porous materials—generalizations of empirical models,'' J. Acoust. Soc. Jpn. (E) **1**, 25–28 (1990).
[19] K. Attenborough, ''Ground parameter information for propagation modeling,'' J. Acoust. Soc. Am. **92**, 418–427 (1992).
[20] J.-F. Allard and Y. Champoux, ''New empirical equations for sound propagation in rigid frame fibrous materials,'' J. Acoust. Soc. Am. **91**, 3346–3353 (1992).
[21] 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).
[22] D. K. Wilson, ''Relaxation-matched modeling of propagation through porous media, including fractal pore structure,'' J. Acoust. Soc. Am. **94**, 1136–1145 (1993).
[23] T. L. Szabo, ''Causal theories and data for acoustic attenuation obeying a frequency power law,'' J. Acoust. Soc. Am. **97**, 14–34 (1995).
[24] K. R. Waters, M. S. Hughes, J. Mobley, G. H. Brandenburger, and J. G. Miller, ''On the applicability of Kramers-Krönig relations for ultrasonic attenuation obeying a frequency power law,'' J. Acoust. Soc. Am. **108**, 556–563 (2000).

An updated perspective on basin-scale tomography

John L. Spiesberger[a)]
Department of Earth and Environmental Science, 240 South 33rd Street, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6316

(Received 30 June 1999; revised 5 May 2000; accepted 15 December 2000)

There seems to be little doubt that pulse-like acoustic transmissions at 1000 to 5000 km can be used to measure small changes in average temperature in many oceanic regions, given the numerous results published by different groups. The early experimental results in this field which led to these conclusions have been confirmed by recent experiments. Our purpose in this Letter is to summarize the work in this field in the context of addressing a few reservations expressed about the early experiments. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1353591]

PACS numbers: 43.30.Pc, 43.30.Cq, 43.30.Qd [SAC-B]

I. INTRODUCTION

Because of the propensity of similar experimental results, there now appears to be little doubt that acoustic thermometers provide accurate estimates of temperature in many ocean regions at basin-scales. These experiments at 1000 to 5000 km in the Atlantic, Pacific, and Arctic oceans were used to reach findings about the interpretability of the acoustic signals, the short- and longer-term climatic temperature variability, and the algorithm for sound speed in sea water. Our purpose in this Letter is to briefly comment on a few reservations expressed about early experimental findings in this field.

II. EARLY RESULTS IN ACOUSTIC THERMOMETRY

The earliest basin-scale experiments in acoustic thermometry were conducted in the Atlantic and Pacific oceans at distances between 2000 and 4000 km.[1,2] Despite the long distances of propagation, many stable acoustic arrivals were observed with travel times that could be measured over experimental durations from months to years. These stable arrivals were successfully modeled by tracing rays through sound speeds derived from climatological records of temperature and salinity. This was important because the economic practicality of the technique depended on not having to sample the ocean along the sections. It was further found that changes in the path's travel times could be used to estimate temperature and its changes with great accuracy.

Comments have been made that these findings were inconclusive, that several open questions remained having to do with the accuracy of the way the data were interpreted, and that it was difficult to interpret the travel times of the paths.[3] Brief comments are made below to review these issues in the context of these experiments.

A. Kaneohe acoustic thermometry papers: Bottom-mounted source and 3709-km section

The Kaneohe source was used intermittently between 1983–1989 to transmit broadband signals near 133 Hz to many receiving stations at 2000–4000 km distance. The source was mounted on the steep northern slope of Oahu at 183-m depth. It was cabled to shore for power and timing. About ten Sound Surveillance Systems (SOSUS) picked up stable acoustic arrivals. Climatic changes in temperature due to El Niño and the Southern Oscillation were shown to be plausible causes for the half-second changes in travel time.[4] The data from one station at 3709-km distance has five stable arrivals, each of which consists of steep ray clusters that appear to be similar to rays in a bounded chaotic sea.[5] The stable arrivals reflected from the steep slope of Oahu one or more times before their long journey across the Pacific to the continental slope near northern California. The accuracy of this acoustic thermometer was estimated by quantifying nonthermal effects which changed the travel times. Comments have been made that nonthermal effects may not have been accounted for.[3]

The nonthermal effects that were discussed included tidal, mesoscale, and internal wave currents, sea floor spreading, pressure and salinity fluctuations, horizontal refraction, ray chaos, and bottom interactions. These effects are small, contributing up to 30 to 40 ms per stable arrival measured over a day. These aberrations limited the accuracy to 0.02 °C per day per stable arrival for the average temperature change in the upper kilometer.[2,5,6] The error of 0.02 °C is an upper bound because an averaging of the travel times from five arrivals would yield a more accurate temperature than obtained from a single arrival.

The nonthermal effects include the change in travel time due to the fact that the ray cluster geometries are affected by changes in the ocean. A comment has been made that because the depth of the bottom is not well known, it is difficult to estimate how much the ray's paths might change because it is not known to where their reflection points could move.[3]

An important component of this experiment was carried out by the Naval Oceanographic Office. They carefully measured the depth of the bottom at high resolution over a wide geographic region out to a distance of 130 km from the acoustic source using SEABEAM.[2,6,7] The locations where the stable paths interact with the bottom are well within this surveyed area. A modeling study was made to measure the amount that the travel times changed due to making changes in the measured depths of the bottom within experimental

a)Electronic mail: johnsr@sas.upenn.edu

errors. The conclusion reached was that the changes in travel times were small.[4] Indeed, these changes were much less than 30 ms. This finding is consistent with Fermat's principle which specifies that changes in travel time due to path variations are of second order because travel times are stationary to first order with regard to changes in path.

Experimental evidence supports the hypothesis that any changes in path geometries led to small changes in travel time. The travel times of the peaks of the five stable arrivals all changed by the same amount at the same time by up to one-half-second, within measurement error (Refs. 2 and 6). If these paths changed significantly, it is hard to see why their travel time changes would be the same since the points at which they reflect from the bottom are all different from one another, and it would be expected that their travel times would thus change in a discordant manner.

Other evidence based on detecting small tidal signals suggests that any path changes led to very small changes in travel time. The barotropic tides and internal tides generated by flat-topped seamounts several thousands of kilometers from the source were accurately estimated at five SOSUS stations, despite the fact that these signals amounted to only about 10 ms (Ref. 8). The ability to detect these small changes in travel time was due to the use of the phase and amplitude of the acoustic signals. Indeed, the measurements were made with an accuracy of 135 μs at two minute intervals over several months.[9]

In addition to the detection of tidal signals, nontidal signals amounting to 10 ms were measured among both the stable and unstable paths at the 3709-km section. In particular, the stable paths occupied the first quartile of the duration of the arriving energy. The changes in travel times of each quartile were measured separately, and all were within a few milliseconds of one another (Table II, Ref. 9). Therefore, the noise induced by any changes in the acoustic paths was small compared with 10 ms.

B. No bottom-interacting sound: 3000-km section

In 1987, broadband signals near 250 Hz propagated from a source near the depth of minimum speed in the waveguide to another SOSUS station at 3000-km distance in the eastern north Pacific.[2] The data exhibited sixteen clear, sharp, and beautifully stable arrivals throughout the duration of the four-month experiment. The arrivals were successfully identified using ray theory based on averaged historical measurements of temperature and salinity.[10] Estimates of temperature and the speed of sound were made using arrivals that did not interact with the sea bottom.

A comment has been made that these data were troublesome to analyze because the positions of the source and receiver were not precisely measured and because no measurements of sound-speed were used to initialize the procedure to estimate the temperature.[3]

The relative separation of the source and receiver were known to within 140 m in range and 10 m in depth.[11] Corrections to these positions were obtained using inverse techniques[12] that are used by most practitioners. The technique relies on the approximation that changes in acoustic travel time are linearly related to changes in the locations of the instruments for sufficiently small displacements. The approximation is valid for this experiment.

It was encouraging to find that the geophysical inversions for temperature could be made without any measurements of the sound-speed field between the source and receiver.[2] The standard inversion techniques were used and no difficulties were encountered in the procedure.

The strongest measure of the scientific validity of a result is its confirmation by others. The data from this 3000-km section led to the publication that the international algorithm for the speed of sound in sea water was too fast at high pressure.[2,13] Subsequently, other groups reached similar conclusions.[14–17] Thus, there are four scientific groups who have found that Del Grosso's algorithm for sound-speed is closer to the truth than the international standard. It seems unlikely that all these groups would reach similar conclusions about such a small error in the standard sound speed algorithm if the early results had significant problems in the analysis.

C. Identification of multipath and modes

The early Atlantic experiment at 2000 km (Ref. 1) and two Pacific experiments at 2000–4000 km (Refs. 2 and 18) were successful in identifying the stable acoustic arrivals with rays over the eleven sections analyzed to date. Comments are sometimes made that appear to imply that a quantitative identification of stable arrivals may not have been made.[3] However, since recent experiments are consistent with early analyses of temperature estimates and corrections to the sound speed algorithm, and because the early analyses of path identifications look like the data,[1,2,18] it appears that the early analyses of path identifications are correct. It is not yet known how far successful identifications can be made in different oceans, but the upper limits are important to measure and understand. Further successes at identifying stable arrivals have been made at distances exceeding 1000 km in the Pacific and Arctic oceans in more recent times.[3,19]

III. CONCLUSIONS

The early experiments in acoustic thermometry were successful in demonstrating the interpretability of the acoustic signals at distances between 2000 and 4000 km, from which accurate estimates of temperature and its change were estimated. It is wonderful that recent experimental results in the Pacific and Arctic are reaching similar conclusions. It is hoped that in the comments mentioned above we have addressed any remaining reservations about the early reports.

There now remains little doubt about the utility of the technique for estimating climate change in the oceans. The study of climate change was one of the motivating factors for the early experiments in the Pacific,[2] and the topic of climate change may become a more important topic in the next few centuries.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research Grant No. N00014-97-1-0484. I thank Michael Porter and other colleagues for their valuable comments.

[1] J. L. Spiesberger, T. G. Birdsall, K. Metzger, R. A. Knox, C. W. Spofford, and R. C. Spindel, ''Measurements of Gulf Stream meandering and evidence of seasonal thermocline development using long range acoustic transmissions,'' J. Phys. Oceanogr. **13**, 1836–1846 (1983).

[2] J. L. Spiesberger and K. Metzger, ''Basin scale ocean monitoring with acoustic thermometers,'' Oceanography **5**, 92–98 (1992).

[3] P. F. Worcester, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, B. M. Howe, J. A. Mercer, R. C. Spindel, J. A. Colosi, K. Metzger, T. G. Birdsall, and A. B. Baggeroer, ''A test of basin-scale acoustic thermometry using a large-aperture vertical array at 3250-km range in the eastern North Pacific Ocean,'' J. Acoust. Soc. Am. **105**, 3185–3201 (1999).

[4] J. L. Spiesberger, H. E. Hurlburt, M. Johnson, M. Keller, S. Meyers, and J. J. O'Brien, ''Acoustic thermometry data compared with two ocean models: The importance of Rossby waves and ENSO in modifying the ocean interior,'' Dyn. Atmos. Oceans **26**, 209–240 (1998).

[5] J. L. Spiesberger and F. Tappert, ''Kaneohe acoustic thermometer further validated with rays over 3700 km and the demise of the idea of axially trapped energy,'' J. Acoust. Soc. Am. **99**, 173–184 (1996).

[6] J. L. Spiesberger, K. Metzger, and J. Furgerson, ''Listening for climatic temperature change in the northeast Pacific: 1983–1989,'' J. Acoust. Soc. Am. **92**, 384–396 (1992).

[7] V. Renard and J. P. Allenon, ''Sea beam, multi-beam echo-sounding in 'Jean Charcot:' Description, evaluation, and first results,'' Int. Hydrogr. Rev. **56**, 35–67 (1979).

[8] R. H. Headrick, J. L. Spiesberger, and P. J. Bushong, ''Tidal signals in basin-scale acoustic transmissions,'' J. Acoust. Soc. Am. **93**, 790–802 (1993).

[9] J. L. Spiesberger, J. P. Bushong, K. Metzger, and T. G. Birdsall, ''Ocean acoustic tomography: Estimating the acoustic travel time with phase,'' IEEE J. Ocean Eng. **14**, 108–119 (1989).

[10] J. L. Spiesberger, E. Terray, and K. Prada, ''Successful ray modeling of acoustic multipaths over a 3000 km section in the Pacific,'' J. Acoust. Soc. Am. **95**, 3654–3657 (1994).

[11] J. L. Spiesberger and K. Metzger, ''A new algorithm for sound speed in seawater,'' J. Acoust. Soc. Am. **89**, 2677–2688 (1991).

[12] B. Cornuelle, ''Inverse methods and results from the 1981 ocean acoustic tomography experiment,'' Ph.D. dissertation, MIT and Woods Hole Oceanographic Institute, Woods Hole, MA, 1983.

[13] J. L. Spiesberger and K. Metzger, ''A basin-scale (3000 km) tomographic section of temperature and sound speed in the northeast Pacific,'' Trans., Am. Geophys. Union **71**, 157 (1990).

[14] P. F. Worcester, B. Dushaw, and B. M. Howe, ''Gyre-scale reciprocal acoustic transmissions,'' in *Ocean Variability and Acoustic Propagation*, edited by J. Potter and A. Warn-Varnas (Kluwer Academic, The Netherlands, 1991), pp. 119–134.

[15] B. D. Dushaw, P. F. Worcester, B. D. Cornuelle, and B. M. Howe, ''On equations for the speed of sound in seawater,'' J. Acoust. Soc. Am. **93**, 255–275 (1993).

[16] F. J. Millero and X. Lee, Comments on ''On equations for the speed of sound in seawater'' [J. Acoust. Soc. Am. **93**, 255–275 (1993)], J. Acoust. Soc. Am. **95**, 2757–2759 (1994).

[17] C. S. Meinen and D. R. Watts, ''Further evidence that the sound-speed algorithm of Del Grosso is more accurate than that of Chen and Millero,'' J. Acoust. Soc. Am. **102**, 2058–2062 (1997).

[18] D. E. Norris, J. L. Spiesberger, and D. W. Merdes, ''Comparison of basin-scale acoustic transmissions with rays and further evidence for a structured thermal field in the northeast Pacific,'' J. Acoust. Soc. Am. **103**, 182–194 (1998).

[19] P. N. Mikhalevsky, A. Gavrilov, A. B. Baggeroer, and M. Slavinsky, ''Experiment tests use of acoustics to monitor temperature and ice in Arctic ocean,'' EOS Trans. Am. Geophys. Union **76**, 265–269 (1995).

On the group velocity of guided waves in drill strings

Flavio Poletto[a] and José M. Carcione
Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (OGS), Borgo Grotta Gigante 42C, 34010 Sgonico, Trieste, Italy

(Received 5 September 2000; revised 8 January 2001; accepted 12 January 2001)

A key element to verify the drill-bit signal processing is the estimation of the time delay of the guided waves transmitted through the drill string from the bit to the surface. This delay can be measured from drill-bit vibrations or calculated using the mechanical properties of the drill string. The calculation of the complete solution and the exact dispersion relations requires expensive numerical modeling. In this work, a simple approach to compute the group velocity of extensional and torsional waves by averaging the drill-string properties in the low-frequency approximation is used. The result is a generalization of the group velocity for periodic systems in which stationary distribution of properties is assumed. Results obtained with real strings show that the group velocity in the bottom-hole assembly is greater than the velocity in a uniform rod and lower than the velocity in the drill pipes. © *2001 Acoustical Society of America.*
[DOI: 10.1121/1.1354196]

PACS numbers: 43.40.Cw, 43.40.At, 43.20.Mv [CBB]

I. INTRODUCTION

Drill-string guided waves contain information about the drilling conditions[1,2] and can be used to transmit data from the bit location to the surface. For instance, Rector and Marion[3] and Miranda *et al.*[4] investigated the potentials of these waves for while-drilling prediction ahead of the bit. The drill bit acts as a seismic source and, in addition, generates extensional and torsional vibrations in the drill string, whose wavelengths are much longer than the diameters of the drill string. These signals are used to correct the arrival time of the data acquired at the surface by a deployed seismic line. Because the correction is based on the group velocity of the guided waves, it is important to determine with accuracy this velocity, which depends on the drill-string acoustic properties and geometrical features.

Several authors investigated the behavior of guided waves in drill strings. Barnes and Kirkwood[5] analyzed the passband and stop band effects caused by the presence of tool joints in the drill pipes of an idealized (periodic) drill string. Drumheller[6,7] investigated the acoustic vibrations of the drill string for transmission of while-drilling information between the drill bit and the surface. Using the dispersion equation, Drumheller[6] calculated the group velocity of the extensional waves for different lengths and cross sections of the drill pipes and tool joints. He considered a periodic system. Using a full-wave modeling algorithm, Carcione and Poletto[8] solved the differential equations describing wave propagation through the drill string. They computed waveforms of the extensional, torsional, and flexural waves by modeling the geometrical features of the coupling joints, including piezoelectric sources and sensors.

In this letter, we propose a method to compute the group velocity of extensional and torsional waves in a drill string composed of several elements of arbitrary acoustic properties and geometrical characteristics. The method is based on averaging the forces, the mass, and the polar moment of inertia of the different components of the drill string. The condition of stationarity is required, that is, in a given length of drill string the proportion of each element (e.g., drill pipes, tool joints, etc.) remains constant. The resulting equations for the group velocity are not restricted to periodic systems.

II. ACOUSTIC PROPERTIES OF THE DRILL STRING

A drill string is composed of several sections and elements of different weights, diameters, and mechanical properties. It is divided into two main parts. The upper part, being the predominant one, is composed of several sections of drill pipes and coupling joints. The lower part is the bottom-hole assembly (BHA), which is heavier and of rather complex composition. The main parts of the BHA are the heavyweight drill pipes and the drill collars. The drill string is normally made of steel with uniform elastic properties, but parts of the BHA are made of aluminum and hard rubber.

Each element of drill string is characterized by a density ρ, a length d, an internal radius r_i, an external radius r_e, a Young modulus E, and a shear modulus μ. Because at seismic frequencies the wavelengths of the guided waves are long compared to the drill-string lateral dimensions, the drill string can be approximated as an effective rod.

Let z be the axial coordinate of the drill string, u the relative axial displacement, and θ the relative angular displacement. The axial and angular strains are

$$\epsilon = \frac{\partial u}{\partial z} \quad \text{and} \quad s = \frac{\partial \theta}{\partial z}, \tag{1}$$

respectively. The strains (1) vary in the drill string according to the properties of the different elements. We assume the low-frequency approximation, which holds for wavelengths that are much longer than the length of the drill-string sections. At the low-frequency approximation, these sections can be seen as homogeneous rods, with constant diameter and elastic properties, where constant axial and torsional

[a]Electronic mail: fpoletto@ogs.trieste.it

stresses generate constant axial and angular strains, respectively.

If k is the wave number and ω is the angular frequency, the exact equations for the phase and group velocities are given by

$$v_p=\frac{\omega}{k} \quad \text{and} \quad v_g=\frac{\partial\omega}{\partial k}=\left(\frac{\partial k}{\partial\omega}\right)^{-1}, \tag{2}$$

where $\omega=\omega(k)$ [or $k=k(\omega)$] is the dispersion relation. Instead of using the exact dispersion relation, which would require a lengthy and unnecessary numerical calculation, we invoke the long-wavelength approximation and substitute the nonuniform drill string by an effective rod of similar average properties. Then, the group velocity at seismic frequencies can be easily calculated from these average properties.

We denote m_{ext} (in kg/m) and m_{tor} (in kg m) the mass per unit length and the polar moment of inertia per unit length, respectively. They are given by

$$m_{\text{ext}}=\rho A, \quad \text{and} \quad m_{\text{tor}}=\rho I, \tag{3}$$

where A is the cross section of the effective rod and I is its polar moment per unit mass and unit length, given by $I=(\pi/2)(r_e^4-r_i^4)$. Because in a uniform rod, the group velocities of extensional and torsional waves equal the respective phase velocities, that is, $v_{g(\text{ext})}=(E/\rho)^{1/2}$, and $v_{g(\text{tor})}=(\mu/\rho)^{1/2}$,[9] the group velocities of the effective rod can be calculated as

$$V_{g(\text{ext})}=\sqrt{\frac{EA}{m_{\text{ext}}}}, \quad \text{and} \quad V_{g(\text{tor})}=\sqrt{\frac{\mu I}{m_{\text{tor}}}}, \tag{4}$$

where EA and μI are the average axial force per unit strain and the average torque per unit strain, respectively.

III. AVERAGE PROPERTIES

Let the drill string be composed of N elements, of axial lengths d_i and cross sections A_i, $i=1,...,N$. We calculate the average of a given property or field variable a by

$$\langle a\rangle=\left(\sum_{i=1}^{N} d_i\right)^{-1}\left[\sum_{i=1}^{N} d_i a_i\right]. \tag{5}$$

A. Average dynamical properties

The axial force (Newton's law) and the torque per unit length are,[9]

$$F_i=A_i\rho_i\frac{\partial^2 u_i}{\partial t^2}, \quad \text{and} \quad M_i=I_i\rho_i\frac{\partial^2\theta_i}{\partial t^2}, \quad i=1,...,N. \tag{6}$$

The average axial force and torque per unit lenght are then given by

$$F=\langle F_i\rangle, \quad \text{and} \quad M=\langle M_i\rangle, \tag{7}$$

respectively. We obtain the average masses $m_{\text{ext}}=A\rho$ and $m_{\text{tor}}=I\rho$ by assuming that the average axial force and torque are related to the average axial and angular displacements $\langle u_i\rangle$ and $\langle\theta_i\rangle$ by

$$F=A\rho\frac{\partial^2\langle u_i\rangle}{\partial t^2}, \quad \text{and} \quad M=I\rho\frac{\partial^2\langle\theta_i\rangle}{\partial t^2}. \tag{8}$$

Averaging Eq. (6) and using (7), (8), and $\partial^2\langle u_i\rangle/\partial t^2=\langle\partial^2 u_i/\partial t^2\rangle$ and $\partial^2\langle\theta_i\rangle/\partial t^2=\langle\partial^2\theta_i/\partial t^2\rangle$, we have

$$\left\langle A_i\rho_i\frac{\partial^2 u_i}{\partial t^2}\right\rangle=A\rho\left\langle\frac{\partial^2 u_i}{\partial t^2}\right\rangle,$$

and (9)

$$\left\langle I_i\rho_i\frac{\partial^2\theta_i}{\partial t^2}\right\rangle=I\rho\left\langle\frac{\partial^2\theta_i}{\partial t^2}\right\rangle.$$

Moreover, assuming statistical independence[10] between u_i and $A_i\rho_i$ and between θ_i and $I_i\rho_i$, we have

$$\left\langle A_i\rho_i\frac{\partial^2 u_i}{\partial t^2}\right\rangle=\langle A_i\rho_i\rangle\left\langle\frac{\partial^2 u_i}{\partial t^2}\right\rangle,$$

and (10)

$$\left\langle I_i\rho_i\frac{\partial^2\theta_i}{\partial t^2}\right\rangle=\langle I_i\rho_i\rangle\left\langle\frac{\partial^2\theta_i}{\partial t^2}\right\rangle$$

[for instance, since u_i, in the long-wavelength limit, varies linearly and very slowly from element to element, we may assume (omitting the time derivative) $\langle A_i\rho_i u_i\rangle=\langle A_i\rho_i\rangle\langle u_i\rangle$, even if $A_i\rho_i$ varies by large fractions from element to element].

Comparison of (9) and (10) yields

$$m_{\text{ext}}=A\rho=\langle A_i\rho_i\rangle, \quad \text{and} \quad m_{\text{tor}}=I\rho=\langle I_i\rho_i\rangle. \tag{11}$$

B. Average static properties

Let F be a constant axial force and M a constant torque acting on the drill string and let ϵ and s be the average axial and torsional strains, respectively. The strains are related to the axial force and torque by

$$\epsilon=\frac{F}{EA}, \quad \text{and} \quad s=\frac{M}{\mu I}, \tag{12}$$

where EA and μI are the averaged properties.

On the other hand, the average value of the axial and torsional strains are

$$\epsilon=\langle\epsilon_i\rangle, \quad \text{and} \quad s=\langle s_i\rangle, \tag{13}$$

respectively. Because

$$\epsilon_i=\frac{F}{E_iA_i}, \quad \text{and} \quad s_i=\frac{M}{\mu_i I_i}, \tag{14}$$

we obtain

$$\langle\epsilon_i\rangle=F\left\langle\frac{1}{E_iA_i}\right\rangle, \quad \text{and} \quad \langle s_i\rangle=M\left\langle\frac{1}{\mu_i I_i}\right\rangle. \tag{15}$$

Comparison of Eqs. (12) and (15) and use of (13) and (14) yields

$$EA=\left\langle\frac{1}{E_iA_i}\right\rangle^{-1}, \quad \text{and} \quad \mu I=\left\langle\frac{1}{\mu_i I_i}\right\rangle^{-1}, \tag{16}$$

respectively. Equations (16) are averages of the reciprocal of the acoustic properties. They are similar in form to the effective elasticity constant c_{33} of a thinly layered medium in the long-wavelength limit.[11]

TABLE I. Dimensions of the drill-collar elements.

System 1			System 2		
r_e (in.)	r_i (in.)	d (m)	r_e (in.)	r_i (in.)	d (m)
6.5	2.8125	0.47	8.0	2.8125	0.7
6.5	2.875	27.58	8.0	2.875	18.37
6.25	2.05	10.33	7.094	3.125	5.39
6.5	2.875	75.40	7.094	3.0	5.95
6.25	2.8125	1.89	8.0	2.875	55.44
6.5	2.875	18.65	8.0	2.75	2.0
6.75	2.8125	1.48	8.0	2.875	9.25
6.75	2.8125	1.48	8.0	2.8125	9.29
6.5	2.875	9.46	8.0	2.75	2.22
6.25	2.8125	0.66	9.25	2.875	0.3
8.5	0.1	0.25	9.625	2.0	9.64
			12.25	0.1	0.31

IV. GROUP VELOCITY

Substituting Eqs. (11) and (16) into (4) yields the group velocities for extensional and torsional waves in the long-wavelength (low-frequency) approximation

$$v_{g(\mathrm{ext})}=\left(\left\langle\frac{1}{E_iA_i}\right\rangle\langle A_i\rho_i\rangle\right)^{-1/2}$$
$$=\left(\sum_{i=1}^{N} d_i\right)\left[\left(\sum_{i=1}^{N} d_i\,\rho_iA_i\right)\left(\sum_{i=1}^{N}\frac{d_i}{E_iA_i}\right)\right]^{-1/2}, \quad (17)$$

and

$$v_{g(\mathrm{tor})}=\left(\left\langle\frac{1}{\mu_iI_i}\right\rangle\langle I_i\rho_i\rangle\right)^{-1/2}$$
$$=\left(\sum_{i=1}^{N} d_i\right)\left[\left(\sum_{i=1}^{N} d_i\,\rho_iI_i\right)\left(\sum_{i=1}^{N}\frac{d_i}{\mu_iI_i}\right)\right]^{-1/2}. \quad (18)$$

These equations are not restricted to periodic systems and provide a good approximation in the frequency range used for while-drilling investigations.

If the density and the Young modulus are constant, and the drill string is composed of a periodic system of pipes and coupling joints, Eq. (17) becomes the group velocity obtained by Drumheller.[7] Moreover, if the shear modulus is constant, Eq. (18) simplifies to

$$V_{g(\mathrm{tor})}=\sqrt{\frac{\mu}{\rho}}(d_1+d_2)\left[d_1^2+\left(\frac{I_1}{I_2}+\frac{I_2}{I_1}\right)d_1d_2+d_2^2\right]^{-1/2}, \quad (19)$$

where d_1 and d_2 and I_1 and I_2 are the lengths and polar moments of the pipes and coupling joints, respectively.

V. EXAMPLE

We consider wave propagation through the drill collars (part of the BHA) for two different cases. They are referred to as systems 1 and 2 in Table I, where d indicates the length of each element of the drill collar. In system 1, the radii are quite uniform, and in system 2 the radii show a larger variation. Each element has the same acoustic properties: $\rho=7840$ kg/m^3, $E=206$ GPa, and $\mu=78.5$ GPa. These values give rod velocities of 5126 and 3164.3 m/s for the extensional and torsional waves, respectively. The group velocities obtained from Eqs. (17) and (18) are 5123 and 3156.8 m/s and 5077 and 3044 m/s for the extensional and torsional waves corresponding to systems 1 and 2, respectively. As expected, due to uniform distribution of masses, the values for system 1 are close to those of the uniform rod.

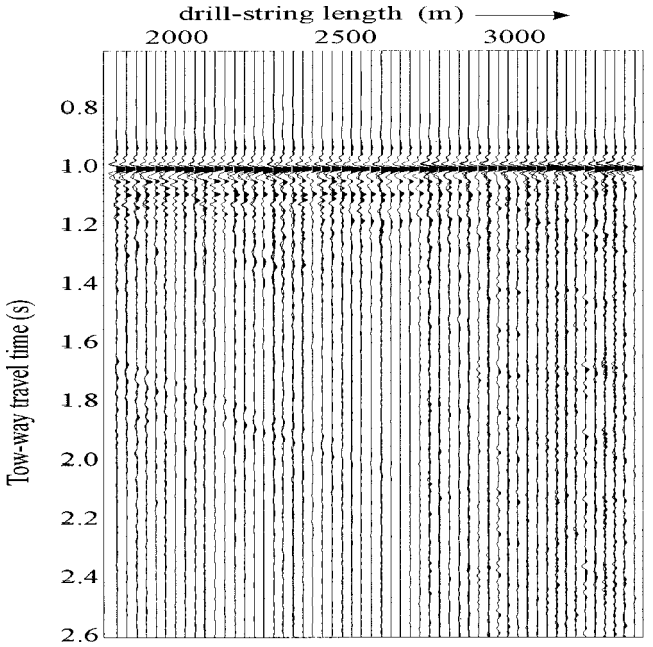


FIG. 1. Drill-bit axial pilot autocorrelation. Short- and long-period multiples have different group velocities in the BHA and in the drill pipes.

Let us consider now the drill pipe/coupling joint system above the BHA, with dimensions $r_e=5$ in., $r_i=4.275$ in., $d=9.2$ m (pipes), and $r_e=6.625$ in., $r_i=2.75$ in., $d=0.5$ m (tool joints). Assuming a periodic system, the extensional and torsional group velocities are 4727 and 2860 m/s, respectively. Note that the wave velocities in the drill collars for system 2 (with an uneven distribution of masses) are in between the velocities of a uniform rod and the velocities of the drill pipe/coupling joint system.

Figure 1 shows an example of real drill-bit signal autocorrelation corresponding to system 2 (see Table I). Each trace corresponds to approximately 30 min of listening time. These data are measured by an axial accelerometer on the top of the drill string, when drilling from 1900- to 3200-m depth. The direct arrival is aligned at 1 s. The dipping events are long period multiples of the drill string. Short-period reverberations are observable after the direct arrivals and after the long-period multiples. These events correspond to reverberations in the main parts of the BHA, i.e., drill collars (DC) and heavyweight pipes (HWDP). The analysis of the multiples in the DC gives a velocity of 5050 m/s, in good agreement with the calculated group velocity.

A similar analysis for the waves in the HWDP, assuming $E=206$ GPa, gives a velocity of 4740 m/s. Because of the different group velocities, the internal reverberations in HWDP and DC of equal length have different periods. The difference can be of 4 ms for a length of 150 m.

In summary, we obtained the group velocities for extensional and torsional waves in drill string at the long-wavelength (low-frequency) limit. For wavelengths longer

than a characteristic length (for which the system is stationary), the drill string behaves like a uniform, or nearly uniform rod, whose mass is the average mass and whose elastic coefficients are algebraic combinations of the elastic coefficients of the single elements. The results are a generalization of well-known expressions for periodic systems, which are not expressed as averages, and therefore cannot suggest a generalization to nonperiodic media or media in which the acoustic and geometrical parameters and density can take more than two values. The results can be used to estimate the delays of low-frequency acoustic signals in the different drill-string sections.

ACKNOWLEDGMENTS

We thank ENI-Agip S.p.A. for funding the research and permission to publish the real data.

[1] J. D. MacPherson, J. S. Mason, and J. E. E. Kingman, ''Surface measurement and analysis of drill string vibrations while drilling,'' SPE/IADC 25777 (1993).

[2] M. Hutchinson, V. Dubinsky, and H. Henneuse, ''An MWD Downhole Assistant Driller,'' Society of Petroleum Engineers, Paper No. 30523 (1995).

[3] J. W. Rector III and B. P. Marion, ''The use of drill-bit energy as a downhole seismic source,'' Geophysics **56**, 628–634 (1991).

[4] F. Miranda, L. Aleotti, F. Abramo, F. Poletto, A. Craglietto, S. Persoglia, and F. Rocca, ''Impact of seismic while drilling technique on exploration wells,'' First Break **14**, 55–68 (1996).

[5] T. G. Barnes and B. R. Kirkwood, ''Passbands for acoustic transmission in an idealized drill string,'' J. Acoust. Soc. Am. **51**, 1606–1608 (1972).

[6] D. S. Drumheller, ''Acoustic properties of drill strings,'' J. Acoust. Soc. Am. **85**, 1048–1064 (1989).

[7] D. S. Drumheller and S. D. Knudsen, ''The propagation of sound waves in drill strings,'' J. Acoust. Soc. Am. **97**, 2116–2125 (1995).

[8] J. M. Carcione and F. Poletto, ''Simulation of stress waves in attenuating drill strings, including piezoelectric sources and sensors,'' J. Acoust. Soc. Am. **108**, 53–64 (2000).

[9] H. Kolsky, *Stress Waves in Solids* (Clarendon, Oxford, 1953).

[10] M. K. Ochi, *Applied Probability and Stochastic Processes* (Wiley, New York, 1990).

[11] J. M. Carcione, D. Kosloff, and A. Behle, ''Long wave anisotropy in stratified media: A numerical test,'' Geophysics **56**, 245–254 (1991).

On the modeling of the vent path in hearing aid systems

B. Rafaely[a] and J. L. Hayes
Institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ, United Kingdom

(Received 3 October 2000; accepted for publication 23 January 2001)

The unamplified response of a vented hearing aid can be significant, in particular at the low frequencies. This letter discusses two different, previously published configurations, in which the vent response is included in the hearing aid model. Any confusion that can potentially arise from the difference in the models is clarified, and a preferred configuration is proposed, which is more consistent with measurable quantities. © *2001 Acoustical Society of America.*
[DOI: 10.1121/1.1356701]

PACS numbers: 43.66.Ts, 43.66.Yw [SPB]

I. INTRODUCTION

Modeling of hearing aid systems is essential in the analysis and design of hearing aids, and in particular when studying the effect of feedback. Typical components in a hearing aid model will include the microphone and receiver, the amplifier or compensation filter, the ventilation tube, or vent, the receiver tube, and the ear canal. Figure 1 shows a schematic diagram of an in-the-ear or in-the-canal hearing aid which includes these components.

The vent is an important component of the hearing aid system. The vent reduces the occlusion effect caused when the ear is blocked, but also allows sound to leak from the receiver back to the microphone, therefore causing feedback, which limits the hearing aid gain. Although the vent is a single acoustic component it is usually represented by two elements in a typical hearing aid system model. The forward part represents the response from a position outside the ear canal to a position near the eardrum, while the feedback part represents the response from a position near the eardrum to the outside.

Previous studies used different configurations for integrating the two parts of the vent in the hearing aid model, the main difference being the way in which the contribution of the forward vent path is combined with the receiver output. The aim of this paper is to compare the different configurations, and propose a preferred configuration that is more consistent with measurable quantities. It should be noted that slit leaks arising from improper seal of the hearing aid to the ear can also cause sound to leak in both directions, and their effect is similar to that of the vent. Although in this letter it is assumed that the leaks are caused by the vent, the same analysis can be used to include the effect of slit leaks by modeling the combined effect of the vent and the slit leaks.

II. CONFIGURATION 1

Figure 2 shows a simplified model of a hearing aid system. H includes the response of the microphone, the amplifier, or compensation filter and the receiver. B denotes the backward response of the vent and A its forward response. P_{ear} is the pressure near the eardrum, P_{ext} the pressure at the microphone position due to an external sound source, and P_{mic} is the total pressure at the microphone position due to both the external sound source and the receiver. A similar diagram was presented by Kates (1991, 1999) and Rafaely *et al.* (2000). The system response with the hearing aid in place and operating, referred to as the real-ear aided response (REAR) in Mueller *et al.* (1992) can be written from Fig. 2 as:

$$\frac{P_{\text{ear}}}{P_{\text{ext}}}=\frac{A+H}{1-HB}. \tag{1}$$

The system response with the hearing aid in place but switched off, referred to as the real-ear occluded response (REOR) in Mueller *et al.* (1992), can be calculated from (1) by setting $H=0$,

$$\left.\frac{P_{\text{ear}}}{P_{\text{ext}}}\right|_{H=0}=A. \tag{2}$$

But what is A representing? If A was to be measured in practice as the response from the pressure at the microphone position to the pressure near the eardrum with an external acoustic source, then the pressure at the microphone position will also be affected by B, the feedback path, as suggested in Fig. 2. The measured response in this case will be more accurately represented by the response from P_{mic} to P_{ear}, which can be written as:

$$\left.\frac{P_{\text{ear}}}{P_{\text{mic}}}\right|_{H=0}=\left.\frac{A+H}{1+AB}\right|_{H=0}=\frac{A}{1+AB}. \tag{3}$$

Equation (3) shows that if a measurement of the real-ear occluded response was to be performed between the hearing aid microphone and potentially another probe microphone placed near the eardrum, as detailed in Mueller *et al.* (1992), then the measured quantity will not be A but $A/(1+AB)$. Measuring A will require a measure of the contribution of the external source to the pressure at the microphone position, P_{ext}, therefore excluding the effect of the vent on the pressure at the microphone position. This measurement is more difficult to perform compared to a simple response between the two microphones. The response $A/(1+AB)$ is therefore

[a]Electronic mail: br@isvr.soton.ac.uk

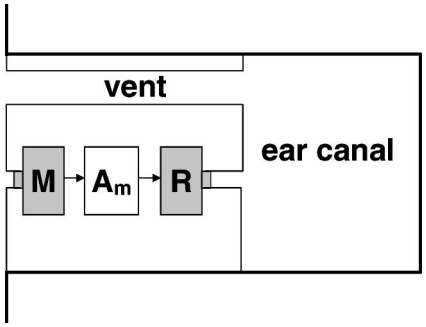


FIG. 1. A schematic diagram of an in-the-ear or in-the-canal hearing aid system, with M the microphone, A_m the amplifier and R the receiver.

easier to measure and provides a preferred choice for the definition of the vent ''forward'' path. Nevertheless, in practice, *A* usually has high magnitude at low frequencies, while *B* has high magnitude at high frequencies, so the product *AB*, particularly for narrow vents, tends to be smaller than unity [see, for example, Kates (1990) and Rafaely *et al.* (1999)], and $A/(1+AB)\approx A$. The next section describes a system configuration where the forward vent path *A* represents the measurable real-ear occluded response.

III. CONFIGURATION 2

Figure 3 shows an alternative system model for the incorporation of the forward vent response. The difference between it and Fig. 2 is that here the feedback signal includes the contribution from the receiver only, and not the forward vent path. This assumes that the feedback part of the sound through the vent which originated from the external source is already included in the response of *A*. A configuration similar to Fig. 3 was presented by Kates (1988a, b) and Rafaely and Roccasalva-Firenze (2000). The real-ear aided response can be written from Fig. 3 as:

$$\frac{P_{\text{ear}}}{P_{\text{ext}}}=A+\frac{H}{1-HB}. \tag{4}$$

The real-ear occluded response can be calculated from Eq. (4) by setting $H=0$, and is the same as in Eq. (2), i.e., equal to *A*. The real-ear occluded response when using the hearing aid microphone as the reference, which represents the forward vent response, can also be calculated as in Eq. (3), and is given by:

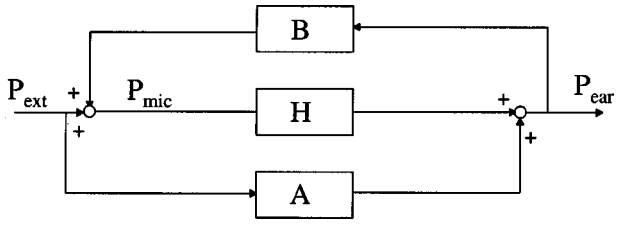


FIG. 2. A system block diagram of a vented hearing aid in configuration 1, with H including the response of the microphone, the receiver, and the amplifier, A the vent response from external pressure to ear canal pressure, and B the vent response from the receiver output pressure to the microphone input pressure.

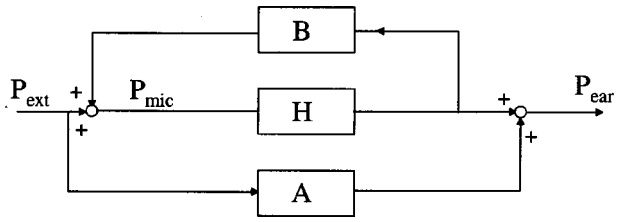


FIG. 3. A system block diagram of a vented hearing aid in configuration 2, with H including the response of the microphone, the receiver and the amplifier, A the vent response from external pressure to ear canal pressure, and B the vent response from the receiver output pressure to the microphone input pressure.

$$\left.\frac{P_{\text{ear}}}{P_{\text{mic}}}\right|_{H=0}=H+A(1-HB)|_{H=0}=A. \tag{5}$$

In this case the forward response of the vent, as measured from the microphone to the pressure near the ear is the same as the forward response *A* in the system diagram. The configuration in Fig. 3 and the corresponding equation (4) are therefore recommended as preferred system representations, since the forward path *A* represents a measurable quantity, which corresponds to the real-ear occluded response with the hearing aid microphone as a reference.

The analysis above leads to the interpretation that *A* can be considered as the complete vent response (forward and backwards) to an external sound source, while *B* is the complete vent response to an internal sound source—the receiver.

It should be noted that although the system configuration presented by Rafaely *et al.* (2000) is similar to Fig. 2, the system equations were similar to Eq. (4). The correct configuration for this study should therefore be as in Fig. 3.

It is also important to note that when measuring the real-ear aided response with the hearing aid microphone as the reference microphone and $|H|>0$, then the response as described in Eq. (5) rather than Eq. (4) will be measured. Only when the feedback signal is small, such that $|HB|\ll 1$, will both equal $H+A$, the combined amplified and acoustic responses.

IV. CONCLUSIONS

This letter presented two configurations for including the vent response in a hearing aid block diagram. It was shown that only for the configuration in which the vent forward path was separated from the backward vent path signal, was the forward vent path representing the measurable real-ear occluded response with the hearing aid microphone as a reference.

ACKNOWLEDGMENT

The authors would like to thank James Kates for the useful comments.

Kates, J. M. (**1988a**). ''Acoustic effects in In-The-Ear hearing aids response: results from a computer simulations,'' Ear Hear. **9**, 119–132.

Kates, J. M. (**1988b**). ''A computer simulation of hearing aid response and the effects of ear canal size,'' J. Acoust. Soc. Am. **83**, 1952–1963.

Kates, J. M. (**1990**). ''A time-domain digital simulation of hearing aid responses,'' J. Rehabil. Res. Dev. **27**, 279–294.
Kates, J. M. (**1991**). ''Feedback cancellation in Hearing Aids: Results from a computer simulation,'' IEEE Trans. Signal Process. **39**, 553–562.
Kates, J. M. (**1999**). ''Constrained adaptation for feedback cancellation in hearing aids,'' J. Acoust. Soc. Am. **106**, 1010–1019.
Mueller, H. G., Hawkins, D. B., and Northern, J. L. (**1992**). *Probe Microphone Measurements—Hearing Aid Selection and Assessment* (Singular Publishing Group, California).
Rafaely, B., and Roccasalva-Firenze, M. (**2000**). ''Control of feedback in hearing aids—a robust filter design approach,'' IEEE Trans. Speech Audio Process. **8**, 754–756.
Rafaely, B., Roccasalva-Firenze, M., and Payne, E. (**1999**). ''Control of feedback in hearing aids,'' Technical Memorandum 836, Institute of Sound and Vibration Research, University of Southampton, U.K.
Rafaely, B., Roccasalva-Firenze, M., and Payne, E. (**2000**). ''Feedback path variability modelling for robust hearing aids,'' J. Acoust. Soc. Am. **107**, 2665–2673.