CODEN: JASI

0001-4966

The Journal of the Acoustical Society of America

Vol. 108, No. 3, Pt. 1 September 2000

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 108, NO. 3, PT. 1, SEP 2000

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 108, NO. 3, PT. 1, SEP 2000

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 108, NO. 3, PT. 1, SEP 2000

CONTENTS—*Continued from preceding page*

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 asked 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.*

Results of the 2000 Acoustical Society election

The following candidates were elected Officers and Members of the Executive Council in the 2000 Acoustical Society election:

William M. Hartmann, *President-Elect*
Janet M. Weisenberger, *Vice President-Elect*
Donna L. Neff, *Member of the Executive Council*
Henrik Schmidt, *Member of the Executive Council*

CHARLES E. SCHMID
Executive Director

Preliminary Notice: Joint 141st ASA Meeting/ NOISE-CON 2000

The Joint Meeting of the Acoustical Society of America (ASA) and NOISE-CON 2000 will be held Sunday through Friday, 3–8 December 2000, at the Newport Beach Marriott Hotel in Newport Beach, California, USA. International standards meetings will be held both before and after the joint meeting. Meetings of Technical Committee 29, Electroacoustics, of the International Electrotechnical Commission [IEC/TC 29 (Electroacoustics)] will be held 27 November to 1 December. Meetings of Technical Committee 43 and Subcommittee 1, Noise, of the International Organization for Standardization, will be held 11–15 December.

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 Sunday through Friday, 3–8 December.

NOISE-CON 2000 will open on Sunday, 3 December, at the Newport Beach Marriott with parallel technical sessions, a plenary session, and a reception. Joint technical sessions on Monday, 4 December, and Tuesday, 5 December, will be held concurrently with ASA technical sessions. NOISE-CON 2000 will conclude on Tuesday evening with a second reception. ASA sessions will begin concurrently with NOISE-CON 2000 on 4 December, and will continue through 8 December. Meeting participants are invited to attend all ASA and NOISE-CON technical sessions.

List of Special Sessions

The special sessions described below will be organized by the ASA Technical Committees/Group.

Acoustical Oceanography (AO)

Acoustic measurements of sediment transport
Acoustical oceanography and satellite remote sensing
Bioacoustics

Animal Bioacoustics (AB)

Instrumentation for animal bioacoustics monitoring and measurements
Standards and animals, where are we?
Use of acoustics for wild animal survey
William Cummings session on the acoustics of whales and dolphins

Architectural Acoustics (AA)

Amphitheater acoustics and noise control
Building acoustics test standards in the U.S. and international arenas
Integration of synthesis techniques and ''acoustical'' music
Speech privacy and speech intelligibility in the built environment for work spaces and education spaces
Theme park acoustics

Biomedical Ultrasound/Bioresponse to Vibration (BB)

Detection and characterization of bubbles, acoustic cavitation, and associated physical effects (joint with Physical Acoustics and Acoustical Oceanography)

Education in Acoustics (ED)

Informal science education
Public education in noise (Joint with Noise)

Engineering Acoustics (EA)

International comparison of acoustical calibration and measurements
Ultrasonic sensors and motors

Musical Acoustics (MU)

Asian musical instruments and traditions
Historical brasses and related instruments
Modal analysis techniques in musical acoustics

Noise (NS)

Construction machinery noise limits and means of compliance (Joint with NOISE-CON)
Environmental noise focused on combined noise sources (Joint with NOISE-CON and Architectural Acoustics)
Hearing protection standards
International noise standards (Joint with NOISE-CON and Architectural Acoustics)
Noise standards—Challenges to quieter products (Joint with NOISE-CON)
Performance assessment of acoustical test rooms (Joint with NOISE-CON and Engineering Acoustics)
Power plants (Joint with NOISE-CON)
Soundscapes (Joint with NOISE-CON)

Physical Acoustics (PA)

Novel optical techniques for measuring surface vibration (Joint with Noise, S2, Signal Processing in Acoustics and Structural Acoustics and Vibration)

Signal Processing in Acoustics (SP)

Blind deconvolution and source separation in acoustics (Joint with Underwater Acoustics and Speech Communication)

Speech Communication (SC)

Alvin M. Liberman: Contributions to speech perception
Functional neuroanatomy and neurophysiology of spoken word recognition and speech perception
Kenneth Stevens: Contributions to speech production and perception

Structural Acoustics and Vibration (SA)

Acoustic microsensors
Diagnostics of vibration and noise in structures (Joint with Signal Processing in Acoustics and Physical Acoustics)

Underwater Acoustics (UW)

High-frequency sediment acoustics and associated sediment properties and processes (Joint with Acoustical Oceanography)

Topical Meeting on the Physics of Echo-Contrast Agents

A one-day colloquium and discussion on the topic ''The physics of echo-contrast agents'' will be held on Monday, 4 December. Six subtopic sessions will focus on modeling of contrast agents, imaging techniques using echo-contrast agents, *in vitro* use of contrast agents, bioeffects of contrast agents, novel contrast agents, and future directions. Each subtopic ses-

 0001-4966/2000/108(3)/855/20/$17.00

sion will begin with a 15-minute overview invited paper, followed by a moderated panel discussion.

Exhibit

The joint instrument and equipment exposition for the two meetings will open at the Newport Beach Marriott on Sunday evening, 3 December, and will close on Tuesday, 5 December. The exposition will include computer-based instrumentation, sound level meters, sound intensity systems, signal processing systems, devices for noise control and acoustical materials, active noise control systems and other exhibits on acoustics. Contact the Exhibit Manager, Richard Peppin, Tel.: (301) 495-7738; FAX: (301) 495-7739, rpeppin@aol.com for additional information.

Other Technical Events

A Distinguished Lecture on ''Ultrasonic Eyeglasses for the Blind'' will be presented by Leslie Kay of New Zealand. A Vern O. Knudsen lecture, sponsored by the Technical Committee on Architectural Acoustics, will be presented by Paul Scarbrough of Jaffee Holden Scarbrough Acoustics, Inc., Norwalk, CT.

A ''Hot Topics'' session sponsored by the Tutorials Committee is scheduled. The Committee on Archives and History will jointly sponsor the next in a series of lectures on the history of acoustics with the Technical Committees on Noise and Speech Communication.

Student Design Competition

The Technical Committees on Engineering Acoustics and Architectural Acoustics of the Acoustical Society of America, in association with the American Loudspeaker Association, JBL Professional, JBL Consumer, Infinity and Revel, are sponsoring a Student Design Competition, which will be professionally judged at the Newport Beach meeting. The 2000 competition involves the design of a loudspeaker system, one specimen of which is to be developed, that may be placed on a pedestal or surface mounted to a room surface. The submitted designs will be displayed at the Newport Beach meeting and shall be judged by a panel consisting of professional loudspeaker designers, design consultants, and university professors. An award of $1000 US will be made to the submitter(s) of the entry judged ''First Honors.'' Four awards of $500 US each will be made to submitters of entries judged ''Commendation.''

Gallery of Acoustics

The Interdisciplinary Technical Group on Signal Processing in Acoustics will sponsor its fourth Gallery of Acoustics at the Newport Beach meeting. The objective of the Gallery is to enhance ASA meetings by providing a compact and free-format setting for researchers to display their work to all meeting attendees in a forum emphasizing the diversity and interdisciplinary nature of acoustics. The Gallery of Acoustics provides a means by which we can all share and appreciate the natural beauty and aesthetic appeal of acoustical phenomena. The Gallery will consist of posters, videos, and audio clips of images and/or sounds generated by acoustic processes or resulting from signal processing of acoustic data. Images and videos can consist of actual visualizations of acoustic processes, or of aesthetically and technically interesting images resulting from various signal processing techniques. Audio clips and segments should also have both aesthetic and technical appeal. A panel of referees will judge entries on the basis of aesthetic/artistic appeal, ability to convey and exchange information, and originality. A cash prize of $350 will be awarded to the winning entry. The top three entries will, with the authors' permission and cooperation, be posted on the Gallery web site. Requests for information and all other communications regarding the Gallery should be directed to Randall W. Smith, Northrop Grumman Corporation, Oceanic and Naval Systems, P.O. Box 1488, Mail Stop 9105, Annapolis, MD 21404, Tel: 410-260-5732; Fax: 410-694-2188; e-mail: psw@bu.edu; e-mail: randall_wsmith@mail.northgrum.com.

Meeting Program

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

NOISE-CON 2000 Supplement

Meeting attendees who wish to receive the CD-ROM Proceedings containing the technical papers from the joint ASA/NOISE-CON sessions as well as the proceedings of NOISE-CON 96, NOISE-CON 97, NOISE-CON 98, and the INCE-sponsored 1998 Sound Quality Symposium and a drink ticket to the Monday evening Exhibit Reception may do so by paying the NOISE-CON supplement fee of $60 at the time they register.

Tutorial Lecture

A tutorial presentation on Virtual Musical Instruments will be given by Julius O. Smith III of the Center for Computer Research in Music and Acoustics, Department of Music, Stanford University, Stanford, CA, at 7:00 p.m. on Monday, 4 December. Notes containing the essential elements of the presentation and the musical illustrations will be available at the meeting. Those who register by 6 November 2000 are guaranteed receipt of 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 6 November 2000 and $25 thereafter including on-site registration at the meeting. The fee for students with current ID cards is $7.00 for registration received by 6 November and $12.00 thereafter, including on-site registration at the meeting. Use the registration form in the call for papers or online at ⟨http://asa.aip.org/newport/newport.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 December, 1:00 to 5:00 p.m. and Saturday Morning, 9 December, 9:00 a.m. to 1:00 p.m.

This 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 the various DSP techniques will be applied to an acoustical application. Participants 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, 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 course 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. Only those who have registered by 6 November will be guaranteed receipt of instructional materials. There will be a $50 discount for registration made prior to 6 November. Full refunds will be made for cancellations prior to 6 November. Any cancellation after 6 November will be charged a $50 processing fee. Use the registration form in the call for papers or online at ⟨http://asa.aip.org/newport/newport.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 October 2000 to 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 include your status as a student; whether you have submitted an abstract; whether you are a member of ASA; method of travel; if traveling by auto, whether you will travel alone or with other students; names of those traveling with you; and approximate cost of transportation.

Young Investigator Travel Grant

The Committee on Women in Acoustics is sponsoring a Young Inves-

tigator Travel Grant to help with travel costs associated with presenting a paper at the Newport Beach 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 Newport Beach 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; e-mail: elmich@m.njit.edu. Deadline for receipt of applications is 30 October 2000.

Students Meet Members for Lunch

The Education Committee has established a program for students to meet with members of the ASA over lunch. Students are strongly encouraged to contact Scott Sommerfeldt, N241 ESC, P.O. Box 24673, Brigham Young Univ., Provo, UT 84602-4673, USA; Tel: 801-378-2205; e-mail: s_sommerfeldt@byu.edu prior to 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 Scott Sommerfeldt. 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 at the Newport Beach Marriott Hotel. The Plenary session will be held on Wednesday afternoon where Society awards will be presented and recognition of Fellows will be announced. A Fellows' Luncheon will be held on Thursday.

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-$\frac{1}{2}\times$11 inch or A4 paper suitable for photocopy reproduction. Copies of available papers will be made for a nominal charge.

Technical Tours

A technical tour has been arranged to visit the Orange County Performing Arts Center, since a number of the members of the Acoustical Society were involved in the design of this fine performance space. Bus transportation will be provided at cost to and from the Newport Beach Marriott. A schedule and sign-up sheets will be available at the meeting.

Air and Ground Transportation

John Wayne Airport in Orange County is served by all major airlines and is a hub for Delta Airlines. The airport designation is SNA. Website: www.ocair.com

The Newport Beach Marriott Hotel is about 15 minutes north of the airport by automobile. Transportation from the airport to the hotel may be by car, taxi, or shuttle. The Newport Beach Marriott Hotel provides complimentary transportation to and from Orange County Airport. You can call the hotel from the hotel telephone board in the baggage claim area. The alternative to arriving at Orange County Airport is to arrive at Los Angeles International Airport (LAX). Scheduled air transportation is available from LAX to Orange County Airport and thence to the Newport Beach Marriott via their complimentary van or by taxi. Check with the airlines for departure times to determine if such service would be convenient for your arrival time.

Convenient transportation from LAX to the Newport Beach Marriott is provided by one of the shuttle van services. The cost of the van service is approximately $25 or $30US per person, including gratuity. After you have retrieved your baggage go outside to the designated van pickup island. Tell the van shuttle service supervisor that you want to go to the Newport Beach Marriott. The trip to the hotel will take 1$\frac{1}{2}$ to 2 hours after you have contacted the shuttle service. An alternative to using the shuttle van service from LAX is to rent a car and drive to the hotel. Detailed driving instructions to the Newport Beach Marriott are available from the car rental office at LAX. Because of the cost, travel by taxi from LAX to the Newport Beach Marriott is not recommended.

Hotel Accommodations

The meeting and all functions will be held at the Newport Beach Marriott Hotel. Please make your reservations directly with the hotel and ask for one of the rooms being held for the Acoustical Society of America. If attendees expect to stay in the conference hotel they are encouraged to make their reservations early. **The reservation cutoff date for the special discounted ASA rates is 12 November 2000; after this date the conference rate will not be available**.

The room rate is $130, single or double, $150 triple, and $170 quad. Room rates are subject to tax, currently 10% plus a $0.14 per day tourist tax. Reservation must be guaranteed by one night's advance deposit or credit card. Please be sure your reservation reaches the hotel by the cutoff date to ensure your accommodations. Send reservations to Newport Beach Marriott Hotel and Tennis Club, 900 Newport Center Drive, Newport Beach, CA 92660, USA, Telephone: +1 949 640 4000, Fax: +1 949 640 5055.

Room Sharing

ASA will compile 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 16 October to the Acoustical Society of America, preferably by e-mail: asa@aip.org or by postal mail to Attn.: Room Sharing, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502. The responsibility for completing any arrangements for room sharing rests solely with the participating individuals.

Weather

During December, the climate is generally mild with bright, sunny days and clear, cool nights. There is the possibility of showers. The maximum air temperature should be between 14 °C and 19 °C, the minimum between 8 °C and 12 °C. A sweater and jacket are recommended.

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.

Accompanying Persons Program

Spouses and other visitors are welcome at Newport Beach. A hospitality room, specifically designated for accompanying persons, will be open daily throughout the meeting at the Newport Beach Marriott. Information will be available about activities in Newport Beach and the surrounding area, such as restaurants, shopping, sports, museums, parks, nightlife, cultural events, and other sites.

The Fashion Island open-air shopping mall is directly across Newport Center Drive from the Newport Beach Marriott. Fashion Island has more than 200 stores and specialty shops for clothing, furnishings, gifts, and art. A movie theater is also located in the center. There are 40 restaurants, cafes, and two Food Courts in the center. Balboa Island in Newport Harbor is easily reached by a short car or taxi ride from the hotel. There are 70 shops, galleries, cafes, and restaurants. A boardwalk around the island affords beautiful views of the harbor. The ferry across Newport Harbor is a picturesque 15-minute ride from Balboa Island to the Balboa Peninsula. Located about 2 km from the Newport Beach Marriott is Rogers Gardens, a magnificent nursery with a wide variety of blooming flowers, hanging plants, shrubs, trees, topiaries, and antiques.

The Orange County Performing Arts Center is in nearby Costa Mesa ⟨http://www.ocpac.org⟩.

Preregistration and Registration

The registration desk at the meeting will open on Sunday, 3 December, at the Newport Beach Marriott. Please use the registration form in the call for papers or online at ⟨http://asa.aip.org/newport/newport.html⟩ to register for the meeting. **If your registration is not received at the ASA headquarters by 24 November, you must register on-site**.

Registration fees are as follows:

Category	Preregistration by 6 November	Registration after 6 November
Acoustical Society or INCE Members	$200	$275
Acoustical Society or INCE Members One-Day	$100	$140
Nonmembers	$245	$320
Nonmembers One-Day	$125	$160
Nonmember Invited Speakers (Note: The fee is waived for these speakers if they attend the meeting on the day of their presentation only)	$200	$275

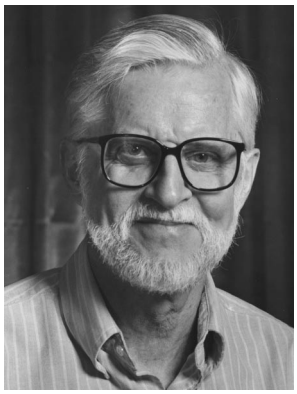
Students (with current ID cards)	Fee waived	Fee waived
Emeritus members of ASA (Emeritus status preapproved by ASA)	$35	$45
Accompanying Persons (For registrants who will not participate in the technical sessions)	$35	$45
NOISE-CON Supplement (Includes NOISE-CON CD-ROM proceedings and Exhibit Reception)	$60	$60

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 (2001) of membership. (Full price for dues: $100.)

Invited speakers who are members of the Acoustical Society of America or The Institute for Noise Control Engineering 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 TO THOSE WHO WISH TO CANCEL THEIR REGISTRATION AFTER 6 NOVEMBER.

Kenneth Stevens receives National Medal of Science



Kenneth N. Stevens

Kenneth N. Stevens, Clarence J. LeBel Professor of Electrical Engineering, is among this year's recipients of the National Medal of Science. Professor Stevens received the award at a White House ceremony on 14 March along with the 11 other recipients of the 1999 award.

''It was a surprise, partly because I'm in a field that normally doesn't receive that kind of recognition,'' Prof. Stevens said. ''It was gratifying for the field, in a sense, and meaningful in terms of my career and in terms of great colleagues and students I've had.

Honoring the discoveries and lifetime achievements of the nation's top scientists, the Medal of Science recipients are a diverse group that created new scientific fields such as conservation biology and speech sciences and led to discoveries that determined why the ozone ''hole'' exists, among other things. ''The contributions of these scientists are so profound, so connected to our everyday lives and so lasting that these medals go only a short way to express the gratitude the nation owes them,'' said Rita Colwell, director of the National Science Foundation (NSF).

Dr. Stevens, who received the Sc.D. in electrical engineering from Massachusetts Institute of Technology and bachelor's and master's degrees in engineering physics from the University of Toronto, has pioneered contributions to the theory, mathematical methods, and analysis of acoustics in speech production, leading to the contemporary foundations of speech science. He is a Fellow of the Acoustical Society of America who was awarded the ASA Silver Medal in Speech Communication in 1983 and the Gold Medal in 1995 ''for leadership and outstanding contributions to the acoustics of speech production and perception.'' Kenneth Stevens has served the Society as Member of the Executive Council, 1963–66, Vice President, 1971–72 and President, 1976–77. His award was announced at the Plenary Session of the Society held in Atlanta and included a video presentation of the award ceremony at the White House.

ASA presents special awards to young scientists at the 51st annual International Science and Engineering Fair (ISEF)

The Acoustical Society of America (ASA) presented four special awards at the 51st International Science and Engineering Fair (ISEF) held at the COBO Center in Detroit, Michigan from 7 to 13 May. The ISEF began in 1950 as a means of encouraging precollege students to conduct scientific research. Over the years, it has grown in size and prestige to become the crowning achievement in high school science competitions. This year's ISEF has attracted 1221 high school students from 40 countries around the world to compete for $2 million in prizes and scholarships for their scientific research projects. For this largest international science project competition by students in grades 9 to 12, ISEF 2000 brought together 1200 professionals, representing university faculty, industrial scientists and engineers, representatives of private and federal research centers and agencies, and medical researchers.

The ASA's First Prize Award of $500 went to Leanne E. Dodge, 18, senior from Bloomington High School South, Bloomington, Indiana, for her project titled, ''Toward A Framework For Quantifying Timbre: Phase III.'' Leanne developed a novel methodology to reveal and describe the timbre of violin by looking at the time rate of change of violin sounds, instead of merely studying the signals in the time- and frequency-domains.

In addition to the first prize, the ASA also presented three Honorable Mention Awards to George Hwang, 15, sophomore from Carman-Ainsworth High School, Flint, Michigan, for his project titled ''Can Acoustical Methods Be Used To Detect And Characterize Corrosion?'' Glen G. Cheng, 15, freshman from Ruidoso High School, Ruidoso, New Mexico, for his project titled ''Non-Intrusive Speed Determination By Doppler Shift;'' and Naveen N. Sinha, 15, freshman from Los Alamos High School, Los Alamos, New Mexico, for his project titled ''Sound Techniques For Studying Granular Material.''

All four winners also received a one-year student membership in ASA. These winners were selected from 30 projects determined by the ASA judges to be relevant to acoustics among 1000 projects presented in the ISEF 2000. The ASA judges included Dr. Jose Benki (Program in Linguistics, University of Michigan), Dr. Nickolas Vlahopoulos (Naval Architecture and Marine Engineering, University of Michigan), and Dr. Sean F. Wu (Chair of the ASA judges, Mechanical Engineering Department, Wayne State University).

SEAN F. WU

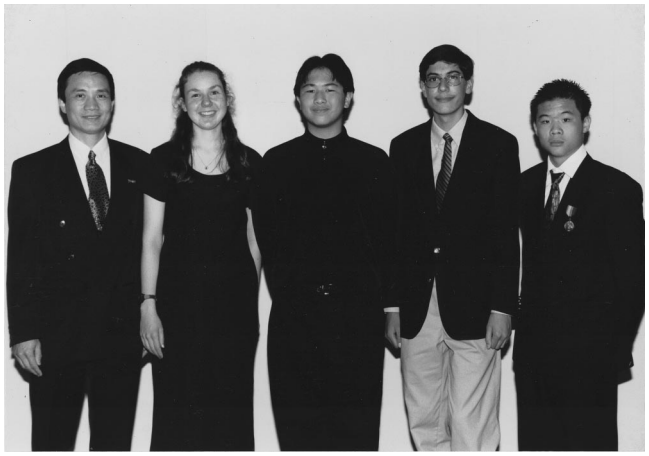


FIG. 1. Sean Wu, Chair of ASA judges, with ASA First Prize Award winner Leanne Dodge and Honorable Mention winners Glen Cheng, Naveen N. Sinha, and George Hwang.

Reports of Technical Committees

Acoustical Oceanography

The past year has been a good one for Acoustical Oceanography. Both the Columbus and Atlanta meetings were well attended, had exciting sessions, and showed the progress and growth of our committee. Rather than dwell on the sessions, which we can all recall simply by looking at our meeting programs, let me talk briefly about some of the new and exciting developments in AO.

One of the nicest happenings of the past year was the endowment of a prize in Acoustical Oceanography by Hank Medwin, AO's founder. The ''Medwin Prize for Acoustical Oceanography'' will be given every year to an accomplished young scientist doing Acoustical Oceanography (not necessarily from the AO committee), with a cash prize of $2000 attached. It has taken about a year for the administrative structure of this award to be put into place, and to get ASA approval, but the process is drawing to a close, and we hope to have the public announcement out in July, 2000, so that the first Award may be given in Newport Beach, still in time for AO's tenth anniversary.

Since I have already mentioned AO's tenth anniversary, let me continue to pursue that topic. At Columbus, as part of the AO Technical Committee meeting, we also had a special celebration of the founding of AO by Hank Medwin and his ''three wise men,'' David Farmer, Van Holliday, and Michael Buckingham. Some good speeches as well as some serious reflection on our future as a committee highlighted the celebration. For those interested in a lighter-hearted view of that celebration, I might also direct you to a special web page on it under the AO Technical Committee's ASA web page.

Another new, and seemingly successful addition to AO's technical agenda this year was the ''Mini-Tutorial in Oceanography'' series, in which we invite various noteworthy speakers from various disciplines in ''classical'' oceanography to give one-hour tutorial lectures and also interact with the Acoustical Oceanography community. The Columbus meeting featured two great lectures on Physical Oceanography (specifically internal waves) by Lev Ostrovsky and Rob Pinkel, whereas Atlanta focused on Marine Geology (specifically sediment stratigraphy), with two memorable lectures by Lou Bartek and Peter Traykovski. Our next topic (in Newport Beach) will be Marine Biology. I have no doubt that these lectures will continue to be as good and as well received as our previous lectures, and that our Mini-Tutorial series will continue to thrive.

Let me next tout our Best Student Paper Award. At each meeting, we give three awards (of $250, $150, and $100) to the three papers that our judging panel deems best. This years winners were: Tracianne Neilson, Reza Dezaji, and Yanqing Zeng in Columbus and Luis Souza, Kyle Becker, and Gopu Potty in Atlanta. Our congratulations to all of them.

Finally, for those reading this who are in AO, let me remind you that the AO Chairman's Memos will routinely augment this information. For those not in AO, but who might want to know more about our activities, please feel free to contact the AO Chairman (Jim Lynch) at jlynch@whoi.edu. I'd be happy to include your name on our distribution list.

James Lynch
Chair 1998–2001

Animal Bioacoustics

Animal Bioacoustics (AB) continues to make good progress in terms of growth and interests, particularly from individuals who do not normally attend meetings of the Acoustical Society of America (ASA). The field is gaining in momentum with increasing numbers of participants at meetings and with a steady increase in the number and quality of manuscripts being published in the *Journal of the Acoustical Society of America* (JASA).

AB conducted six half-day sessions at the Columbus meeting. One of the sessions was jointly sponsored by AB, Speech Communication, and Signal Processing in Acoustics (SP), and two were joint sessions with Underwater Acoustics and Acoustical Oceanography. There were a total of 76 papers presented. The first session included two tutorial talks, one by Leon Sibul on wavelet analysis and the other on time-frequency analysis by Patrick Loughlin. Both of these tutorial talks were arranged through Jim Candy, chair of the Signal Processing interdisciplinary technical group. Two of the sessions entitled ''The effect of man-made sounds on marine mammals'' were conducted jointly with Underwater Acoustics and Acoustical Oceanography, although most the speakers were from AB. This session was followed by a panel discussion. The topic was very timely and the overall consensus was that the session was very good. The panel discussion was also good and perhaps a little tame relative to the heated arguments that often accompany this topic in other forums. Chris Clark was inducted into the College of Fellows. AB also participated in the ''Hot Topics'' session with Sam Ridgway representing AB.

There were two student paper awards at the Columbus meeting. The first prize went to Carolyn Schlundt of San Diego State Unversity, and the second prize went to Micheal Dent of the University of Maryland. Honorable mention: R. S. de Sousa Lima at the Zoologia/Univ. in Brazil.

AB conducted four half-day sessions at the Atlanta meeting, with 35 total contributions. One of the sessions was a jointly organized by AB and SP. The AB sessions were probably the most diverse ones in many years with an interdisciplinary component and a wide variety of different types of animals. Organizers of the different sessions included Whit Au, Bob Dooling, Cindy Moss, and Larry Pater. At the end of the Atlanta meeting, Mardi Hastings and Whit Au swapped positions with Mardi becoming the chair of the Animal Bioacoustics Technical Committee from Whit representing AB on the Medals and Awards Committee. Paul Thompson and Peter Tyack were inducted into the College of Fellows. Elizabeth Brittan-Powell from the University of Maryland won the best student paper award. Approval was granted by the Executive Council to have a named session in honor of William Watkins at the Chicago meeting.

Some of the major issues brought out and discussed at the AB Technical Committee meeting include the following:

(1) The frequency of meetings—too often
(2) Animal Acoustics communication symposium—should be held in spring of 2002
(3) Jim Simmons experienced some problems with putting mixed media material on ARLO.

WHITLOW W. L. AU
Chair 1997–2000

Architectural Acoustics

The Technical Committee on Architectural Acoustics (TCAA) had a very diversified year in presentations from our meetings in Berlin [jointly with the European Acoustics Association (EAA) and DEGA, the German Acoustical Society] and Columbus to the activities supported by Technical Initiatives from the Technical Council.

In summary the major events for the year included the ASA meetings in Berlin and Columbus, the Concert Hall Acoustics Summer Institute held at Tanglewood Music Center, Classroom Acoustics Workshop in New York City, Workshop on Acoustics of Themed Entertainment, preparation of presentation materials in support of Classroom Acoustics Workshops, continued work in support of the standards group on Classroom Acoustics, Student Design Competition, and development started on the TCAA web page.

The Berlin meeting opened at the Berlin Philharmonic Hall with a special concert hosted by Jürgen Meyer explaining the impact that special acoustical properties of the hall had on the music being performed. Sessions presented at the Berlin meeting included Concert & Opera Halls; Case Studies of New Halls; Opera Houses, Modeling of Halls; Halls with Special Features; Modeling of Noise in Work Rooms, Worship and Theater Spaces; Perceptual Aspects in Rooms; Verification of Auralization and Modeling Programs; Sound Insulation and Subjective Annoyance; Development and Verification of Prediction Methods in Building Acoustics; Airborne and Impact Sound Transmission and Classroom Acoustics, Reverberation, Absorption, and Scattering, Acoustical Measurements. There were a large number of posters presented from the Architectural Acoustics technical area. Technical Committee organizers for Architectural Acoustics were Angelo Campanella and Richard Campbell. Michael Vorländer was our liaison technical representative at EAA/DEGA. The session chairs and cochairs for the Berlin meeting included Leo L. Beranek, Jürgen Meyer, Murray Hodgson, Wolfgang Probst, Emmanuel Tzekakis, Daniel Commins, Hideki Tachibana, Roberto Pompoli, Michael Vorländer, Robert Coffeen, Jens Holger Rindel, Richard Campbell, U. Peter Svensson, Trevor Nightingale, Eddy Gerretsen, Sten Ljunggren, Alfred Warnock, K. Heinrich Kuttruff, and Angelo Campanella. TCAA under the direction of Bob Coffeen also held a Student Design competition for the design of a university music rehearsal room and music practice rooms. First Honors ($1000) went to Robin Glosemeyer and Ben Seep, Architectural Engineering, University of Kansas. Commendation

($500) went to Meghan Howard and Amanda Mundis, Architectural Engineering, Penn State; Emily Hulce and Matt Linn, Architectural Engineering, University of Kansas; Jason Jennings, Architecture, Oklahoma State University; and Brent Swain, Architecture, Texas A&M University.

The Columbus meeting was an exciting one gaining national press coverage of some of the sessions that were held. Sessions included Integration of Synthesized Musical Instruments with Acoustics Environments, Measurements, Archeological Acoustics, Tribute Session to Buzz Towne—Classroom Acoustics, and Architectural Renovation and Restorations: Acoustics Issues. There was also a Distinguished Lecture on Acoustics at the End of the 20th Century—An Overview of the State of the Art by Malcolm J. Crocker, a Vern O. Knudsen Distinguished Lecture: ''The complete concert hall: A century of acoustics development and its lessons for the design of concert halls'' by Leo L. Beranek, and a joint session with Musical Acoustics: Daniel W. Martin Tribute Session. The Technical Committee organizer for the Columbus meeting was Angelo Campanella. The session chairs and cochairs at the Columbus meeting were Dan Clayton, Richard Campbell, Anthony Hoover, Richard Godfrey, David Lubman, Bill Cavanaugh, and Scott Pfeiffer. Of special interest at the Columbus meeting was the Dan Martin Organ Tribute Concert and tours of Owens Corning Testing Systems and the Southern Theater. TCAA also sponsored a Student Reception jointly with Noise. Dorie Najolia was selected as the TCAA student representative to the student council meeting in Atlanta.

Technical Initiatives supporting TCAA were used for the following projects in 1999:

- Classroom Acoustics Workshop held at the City University of New York Graduate Center on 26 and 27 February 1999 chaired by John Erdreich. There were 50 attendees.
- Workshop on Acoustics of Themed Entertainment held on 20 and 21 April 1999 in Murro Bay, CA, chaired by Dana Hougland and Tobin Cooley. There were approximately 50 participants.
- Concert Hall Acoustics Summer Institute held at Tanglewood Music Center, Lenox, MA from 29 August to 1 September 1999 chaired by Bill Cavanaugh and Tim Foulkes. Sponsorship came from ASA/TCAA, National Council of Acoustical Consutlants, Concert Hall Research Group, and the Robert Newman Student Award Fund. There were 31 participants in the institute.
- Classroom Acoustics Workshop presentation materials were developed by Bob Coffeen and his students. Publication of this material is scheduled for 2000.
- TCAA provided continued support of data reduction from concert hall measurements. Dick Campbell is overseeing this project. This project is co-sponsored by the Concert Hall Research Group.
- Development of the TCAA website started at the Columbus meeting. Les Blomberg was responsible for the completion of the website in 2000.
- TCAA also maintains its participation in Continuing Education for architects by providing American Institute of Architects credits for presentations such as the Classroom Acoustics Workshop.

Many of the members from TCAA participated in the Standards Working Group in the development of guidelines for acoustics properties of properly designed classrooms. The group received funding from the Access Board with the goal of completing the standard by 2001.

1999 was an exciting year with our international meeting and the various topics covered at the ASA meetings as well as the workshops and events that happened outside the normal meetings. Participation remains strong in the technical committee with an increase in the number of students recently attending.

RONALD R. FREIHEIT
Chair 1998–2001

Biomedical Ultrasound/Bioresponse to Vibration

Thirty-six members attended the Biomedical Ultrasound/Bioresponse to Vibration technical committee (B&B TC) meeting in November 1999, in Columbus, OH. Special sessions in Columbus were organized by Shira Broschat, Robin Cleveland, and Inder Makin, with other sessions chaired by Diane Dalecki and Brian Fowlkes. Bioresponse to Vibration member Ron Verrillo was awarded an ASA Silver Medal. Robin Cleveland organized the student paper contest at Columbus. Inder Makin was the Technical Program Organizing Meeting (TPOM) representative and arranged the papers into sessions.

Technical Initiatives from B&B for the year 2000 include webpage development (please send links to your website to Shira Broschat for inclusion on the B&B homepage), a student paper contest in Atlanta, travel expenses of some invited speakers, and funds for inviting a subset of American Institute of Ultrasound in Medicine (AIUM) members to join ASA B&B (letter mailed July 2000). Pat Kuhl's Societal Growth Initiative on targeting youth identified a need for a council of students to advise the president in the way the technical committee chairs do. Sandy Poliachik was picked to be our student liaison to this student council with a term through 2000.

The attendees of the Berlin B&B TC meeting had voted to emphasize only one ASA meeting each year, in particular the Spring meeting, to avoid conflicts with AIUM or the Institute of Electrical and Electronics Engineers (IEEE). At the Columbus meeting, it was agreed that the Fall meetings might have a different character, something like a ''topical meeting.'' Thus for the Newport Beach ASA meeting, a one-day (Monday) colloquim and discussion on the topic ''Physics of Echo-Contrast Agents'' will be put together by a subcommitee chaired by Carr Everbach. The structure of this meeting consists of six short subtopics: modeling of echo-contrast agents, imaging techniques using echo-contrast agents, *in vitro* use of echo-contrast agents, bioeffects of echo-contrast agents, novel uses of echo-contrast agents, and an overview. There is planned only one (short inspirational) talk for each subtopic by one invited speaker. Following this kick-off talk, a three-person panel (with moderator) will initiate open discussion for about an hour on each subtopic.

Three such subtopics in the morning and three in the afternoon would comprise the day-long meeting, and no contributed papers would be presented. No abstracts other than the six invited kick-off talks would appear in the program for this day-long topical meeting, but a contributed paper session later in the week would allow presentations of traditional 15-minute talks by participants. For the Newport Beach meeting, there will be only one special session, ''Detection and Characterization of Bubbles, Acoustic Cavitation, and Associated Physical Effects'' organized by Ron Roy and Tim Leighton. The special session will be cosponsored by the Technical Committees on Physical Acoustics and Acoustical Oceanography derive largely from the work of SWG-22, the Standards Working Group that Ron Roy and Wes Nyborg have been working on, focused on cavitation detection and monitoring. Larry Crum will deliver the B&B lecture in ''Hot Topics'' at Newport Beach.

The Atlanta TC meeting in June 2000 was attended by 16 people. Special sessions were organized by Ibrahim Hallaj and Thomas Szabo on ''New techniques in biomedical imaging,'' by Christy Holland, Pierre Mourad, and Suk Wang Yoon for ''Ultrasound for disease treatment and diagnosis,'' and Robin Cleveland for ''Lithotripsy.'' Robin Cleveland was awarded the R. Bruce Lindsay Award and Larry Crum the Helmholtz-Rayleigh Interdisciplinary Silver Medal, thereby making the awards ceremony at the plenary session seem very B&Bish. Bob Apfel, Floyd Dunn, Wes Nyborg, David Blackstock, Mark Hamilton, and Ron Roy contributed to the encomia for Larry and Robin's awards. Ibrahim Hallaj was the TPOM representative for Atlanta and arranged the papers into sessions.

The student paper contest in Atlanta was organized by Carr Everbach and was very successful, with 14 entries and 17 judges. First prize went to Constantin C. Coussios for ''Ultrasonic scattering from blood as a means of measuring hemolysis;'' second prize went to Xufeng Xi for ''Dynamic photoelastic study of the transient stress fields in solids during shock wave lithotripsy;'' and third prize went to Dahilia Sokolov for ''Bubble translation due to radiation force in SWL.'' In Columbus, it was suggested that there be separate prizes for categories of undergraduate, master's, and doctoral students, but there were insufficient numbers in each category to make this feasible.

At the Technical Council meeting following the Atlanta sessions, B&B put forth and was granted a technical initiative to buy a portable computer projector, with the proviso that when we are not using it, other Technical Committees may do so. It will be the responsibility of the B&B chair to ensure that the computer projector makes it to each meeting. Plans for the Chicago meeting in 2001 include sessions on beamforming, novel imaging techniques (tissue harmonic imaging, sonoelasticity, etc.), image-guided ultrasound therapy, and ultrasound bioeffects. We hope to see you at Newport Beach, Chicago, or another upcoming meeting.

E. CARR EVERBACH
Chair 1999–2002

Engineering Acoustics

The Engineering Acoustics Technical Committee (EATC) met on 2 November 1999 at the Columbus ASA Meeting and again on 31 May 2000 at the Atlanta ASA Meeting.

The EATC paper sorting for the Columbus Meeting was handled by Ahmet Selamet and included 70 papers in eight EATC sessions with four special sessions. The first special session, titled ''Transducer Loss Mechanisms,'' was organized by Art Ballato and Harold Robinson and included four invited and one contributed papers. The second special session, ''Compatibility of Hearing Aids and Cellular Telephones: Standards Progress,'' was organized by Stephen Thompson and included six invited presentations and one contributed paper. ''Acoustic Nondestructive Evaluations: New Directions and Techniques,'' was the third EATC special session. This session was organized by P. K. Raju and included two invited and seven contributed papers. Our fourth special session was ''Acoustics of Cellular Telephones,'' which was organized by Marty Alexander and included five invited and one contributed paper. A student paper contest was organized and conducted by Hal Robinson, the winner was A. Arif Ergin of the University of Illinois at Urbana Champaign, for the paper ''Exact global boundary condition with reduced computational complexity for grid truncation in acoustic finite element time domain simulations.''

The EATC paper sorting for the Atlanta Meeting was handled by Kim Benjamin and Tom Howarth. This meeting included 44 papers in six EATC sessions including one special session, one historical overlook, and one current topics paper. The special session was organized by Steven Garrett and was titled ''Combustion Acoustics.'' EATC also cosponsored a 1 hour presentation by Stan Ehrlich on ''The History of Engineering Acoustics'' with the Committee on Archives and History. Thomas Howarth presented the ''Hot Topics in Engineering Acoustics'' paper sponsored by the Tutorials Committee. The student paper contest was organized and conducted by Hal Robinson. The winner at the Atlanta meeting was Francois Guillot of Georgia Institute of Technology, for the paper ''An energy-based method for the experimental determination of electrostrictive coefficients on polyurethane films.''

This meeting also marked the final meeting with Thomas Howarth as the EATC Chair. Stephen Thompson has been elected as the Chair for 2000 to 2003.

THOMAS R. HOWARTH
Chair 1997–2000

Musical Acoustics

1999–2000 was an active year for the Technical Committee on Musical Acoustics (TCMU). The ASA meeting in Columbus in November 1999 featured a large number of sessions sponsored by the committee. Jim Pyne and Daniel Avorgbedar organized a session on African musical instruments and traditions, which included performances as well as lectures. Performance was also included as part of a session on free reed musical instruments organized by Jim Cottingham. ''Music, Rhythm, and Development'' was the title of a session organized by Caroline Palmer and Mari Riess Jones. The TCMU and the Committee on Archives and History sponsored a lecture on the history of musical acoustics presented by Gabriel Weinreich. A performance technical session organized by Jim Pyne featured a concert by the Ohio State University Marching Band.

Three sessions at Columbus were co-sponsored with Architectural Acoustics. Doug Keefe, Dan Clayton, and Bill Cavanaugh organized the Dan Martin Tribute Session. The Dan Martin Tribute Concert, an organ demonstration and recital at the First Congregational Church on Tuesday evening, served as a prelude to the technical session on Wednesday morning. A session on archeological acoustics, organized by David Lubman, included a number of musical acoustics papers. Richard Campbell and Anthony Hoover organized a session on integration of synthesized musical instruments with acoustic environments.

Two contributed paper sessions at Columbus chaired by Tom Rossing and Peter Hoekje included papers on a wide range of topics in musical acoustics. Among these was a special invited paper by Andres Peekna on the acoustics of the Baltic psaltery, including demonstrations on several instruments. Musical acoustics was the focus of the Monday evening tutorial lecture, ''An Evening with the Art and Science of Music,'' presented by Uwe Hansen and Jim Pyne, assisted by a number of musicians from The Ohio State University.

The June 2000 meeting in Atlanta included four special sessions. Antoine Chaigne and Stephen McAdams organized a session on modeling and perception of musical sound sources that brought together presenters working in these different areas of musical acoustics. A session on visualization methods in musical acoustics was organized by Peter Hoekje and chaired by Uwe Hansen. Peter Hoekje organized and chaired a session on choral and solo vocal performance. A session on audio signal data compression for musical acoustics was organized by Jim Beauchamp.

Efforts to increase student involvement in musical acoustics continue through the student paper competition and joint sponsorship of student socials. For the Atlanta meeting the technical committee appointed Rachel Romond of Occidental College as its representative on the ASA Student Council. There were ten entrants in the student paper competition in Columbus and six entrants in Atlanta. The winner of the Best Student Paper Award in Musical Acoustics at Columbus was Teresa Marrin Nakra, of the MIT Media Lab, who presented ''Incorporating gestures into the musical control stream with the conductor's jacket.'' The award winner at the Atlanta meeting was ''Lip and throat configuration in flute playing,'' presented by Joon-Hee Beth Hwang of the University of Sydney, Australia.

Education outreach this year included a continuing technical initiative in support of educational workshops on acoustics for teachers conducted by Uwe Hansen. A new technical initiative for 2000 being conducted by Peter Hoekje is a project to prepare a set of demonstrations in musical acoustics to be available on CD-ROM or on the web.

The representatives to the Technical Program Organizing Meetings were Jim Cottingham for the Columbus meeting and Jim Beauchamp for the Atlanta meeting. Representation from the TCMU on other ASA committees continues with Uwe Hansen on the Membership Committee, Roger Hanson on Medals and Awards Committee, and Ian Lindevald on the ASA Committee on Standards (ASACOS). *Ex officio* members on the TCMU include JASA Associate Editors Dean Ayers and Stan Ehrlich. Members appointed or reappointed to serve on the TCMU for terms expiring in 2003 include Anders G. Askenfelt, James W. Beauchamp, Xavier Boutillon, Matthew Brooke, D. Murray Campbell, René Caussé, Nicholas J. Giordano, J. M. Harrison, William M. Hartmann, Norman C. Pickering, James M. Pyne, Daniel A. Russell, Punita Singh, Paul A. Wheeler, and Shigeru Yoshikawa.

The web site for the TCMU maintains current information on musical acoustics at http://www.public.coe.edu/~jcotting/tcmu.

JAMES P. COTTINGHAM
Chair 1999–2002

Noise

The TC Noise met two times since last summer: in the fall in Columbus, OH, and in the spring in Atlanta, GA. Both meetings were well attended. The committee had several interesting sessions at both meetings, all well attended.

Again, the most visible item to discuss is the efforts several members made on the classroom acoustics task force. Funding, measurement standards, and criteria are all developing from this effort. We expect a standard to develop from this effort in about one year.

One survey of membership, taken at the request of the Technical Council, asked for opinions on the frequency of the ASA meetings. By far the majority opinion was that fewer than two meetings per year would be preferable. That is, most members who responded to the survey preferred one meeting per year or one meeting every 18 months. However, the Technical Committee opinion did not prevail as ASA elected to remain with the present number of meetings per year (two).

Miscellaneous issues:

(i) We distribute all minutes by e-mail only. This seems efficient and definitely a cost savings.

(ii) The qualifications for Committee ''membership,'' if any, remain vague. As a result, the membership list grows because of the reluctance to drop long-time, active members. It appears that people participate on the Committee independent of their membership status. That is, the conferring of ''member'' status does not assure attendance at the TC Noise meetings.

We have had help from many people, from session organizers, to program planners, to invited speakers, to people bringing refreshments, to workshop organizers, to contributors to the lively discussions at the meeting.

We continue to welcome new people to our open meetings and solicit their participation.

We had a lively discussion about the Institute of Noise Control Engineering's (INCE) influence on the Technical Committee's operations and goals. Only a small minority had felt the influence was negative. In fact, the TC enthusiastically supported a joint INCE and ASA meeting in December 2000.

Finally, there was an election to replace me as Chairperson. Bennett Brooks was elected and hence this is my last report for TC Noise. I am happy for the opportunity to serve as Chairperson. It was a very nice role in ASA: not a lot of work, and it gave me a chance to more formally work with a great group of enthusiastic and talented people. I appreciate the opportunity to help the Society. It was a nice part of my professional career.

I thank all members, and interested persons. The TC prospered thanks to your efforts.

RICHARD J. PEPPIN
Chair 1997–2000

Physical Acoustics

Physical Acoustics (PATC) continues in its tradition of having lively, amiable Technical Committee meetings. Seventy-four people attended the Columbus and seventy-nine attended the Atlanta PATC meetings. James Sabatier's stewardship over the last three years has fostered a healthy group of characters, and this new chairman hopes to do the same.

After arranging for special sessions for upcoming meetings, the PATC mulled some interesting issues. At the Columbus meeting, with a vote of 32 for and 17 against, the committee decided to start offering student paper awards. The issue of our group sponsoring poster sessions came up again but was voted down by a 27 to 7 margin. We favored the creation of the ''affiliate member'' category for the ASA as proposed by Larry Crum by a 40 to 6 vote. We also agreed to start an e-mail mailing list. Our thanks go to Ralph Muehleisen and Elaine Moran for setting it up. At the Atlanta meeting, overwhelming support was given for the continued publication of References to Contemporary Papers in Acoustics. About two-thirds of those present thought they would purchase Samuel Temkin's ''Elements of Acoustics'' if republished by the Society. The membership was pleased with the announcement that the 1929–1996 issues of JASA would soon be available on CD-ROM. When asked how much they would pay for it out of their own pockets, 12 said they would pay $300, 27 would buy it at $200, and 48 of the 79 present said they would purchase the CD set if it were priced at $100.

Special thanks go to Henry Bass, David Chambers, Joseph Dickey, Steven Garrett, Ibrahim Hallaj, Kent Lewis (twice), Philip Marston, Julian Maynard, Lev Ostrovsky, Andrea Prosperetti, P. K. Raju, and Thomas Szabo for their work in organizing or chairing Physical Acoustics special sessions for the Columbus and Atlanta meetings. Our thanks also go to Yves Berthelot, James Chambers, Richard Raspet, James Sabatier, and myself for attending the Technical Program Organizing Meetings. James Sabatier will be our new representative to the Medals and Awards committee. Thank you Mark Hamilton for your past service on that committee. The work of John Stroud and Christy Holland in keeping up our web page is very much appreciated. It will now be taken care of by Elaine Moran.

For technical initiatives, we held student receptions both at the Columbus and Atlanta meetings, organized by James Chambers, James Sabatier and myself. These have been very well attended, but we are struggling with keeping costs under control. We also supported the travel of two young invited speakers to the Atlanta meeting. For the upcoming year, our committee has asked the ASA to cosponsor with Eindhoven University of Technology, The Netherlands, a First International Workshop on Thermoacoustics, to be held 23–25 April 2001, at s'Hertogenbosch, The Netherlands.

It is a pleasure to note that the winners of awards and medals at the Atlanta meeting are all associated with our technical committee: Robin O. Cleveland received the R. Bruce Lindsay Award, Lawrence A. Crum received the Helmholtz-Rayleigh Interdisciplinary Silver Medal, and Murray Strasberg was the Gold Medal recipient.

Any questions, complaints, advice or whatever can be directed to me at bonzo@sabine.acs.psu.edu.

ROBERT M. KEOLIAN
Chair 1999–2002

Psychological and Physiological Acoustics

The ASA meeting held in Columbus was surprisingly well attended by P&P members considering the committee's long-standing policy of deemphasizing the fall meeting. An attraction, perhaps the highlight of the meeting, was the special session honoring Robert Bilger. The session was lively and diverse and was an appropriate acknowledgement of Bob's many contributions to psychoacoustics. Thanks to Larry Feth and Walt Jesteadt for organizing this successful session. Thanks also to Larry for serving as Technical Program Chair and to Lisa Stover for organizing the P&P sessions. The evening P&P meeting was mercifully brief and without serious controversy. Donna Neff was thanked for her excellent service as outgoing P&P chair. There was discussion of the possibility of a new ASA membership category, Affiliate. This was received with little enthusiasm, perhaps because of uncertainty regarding its effects on the Society. There was brief discussion of the special sessions planned for Atlanta and members were encouraged to begin thinking about special sessions for the Chicago meeting in spring 2000. No special sessions are planned for Newport Beach. The results of the ASA-wide membership survey were briefly discussed: P&P members who responded continue to overwhelmingly support one meeting per year. It was announced that Sid Bacon will present a Hot Topic for P&P at Atlanta and that Nandini Iyer will be the P&P representative on the new Student Council. The death of Edward Carterette was sadly noted and plans for an ASA obituary were discussed. It was announced that Laurel Carney replaced Bob Frisina as Associate Editor. Bob was warmly thanked for his excellent and extended editorial work. Tina Rankovic presented a progress report on the CD-ROM of the Bell Labs lab notes and Peggy Nelson provided an update of her work on the committee on classroom acoustics.

The Atlanta meeting was better attended by P&P members than the Columbus meeting but there were somewhat fewer papers and posters than at a typical spring meeting. The excellent special session ''Characterization and possible neural bases of age-related hearing loss'' was organized by Bob Frisina. Also well received was the special session organized by Bob Dooling on ''Cognitive aspects of complex sound perception in animals.'' This session was cosponsored by Animal Bioacoustics, P&P, and Speech. Sid Bacon represented P&P very well in his Hot Topic on compression. Thanks to Marjorie Leek who organized the P&P papers for the meeting. The evening P&P meeting was not mercifully brief, primarily because of numerous information items that were potentially of concern to P&P. One item that was not greeted with enthusiasm was the decision by the Executive Council to continue with two meetings per year. P&P reaffirmed its policy of no special sessions during the fall meeting. We also agreed unanimously that the proposed membership category of ''e-associate'' should continue to be examined. During the P&P meeting Jont Allen demonstrated the JASA on CD-ROM. This was well received and several future enhancements were suggested for this useful and appreciated project. Several books were proposed for reprinting by ASA, including the Swets readings, Fechner's *Elements of Psychophysics*, Hirsh's *The Measurement of Hearing*, the collection of Jeffress' papers, and Titchener's *Experimental Psychology, Volume 2*. Additional suggestions are welcome. Several public policy issues of relevance to P&P were suggested and will be forwarded to the *ad hoc* committee on public policy. There was some heated discussion of Acoustics Research Letters Online (ARLO) and the possibility that it will compete with JASA (while using JASA editorial resources) once it becomes independent of JASA on July 1 2000. As a preface to the JASA portion of the evening meeting, the chair briefly discussed the great importance of the Journal in the field of hearing and emphasized the crucial role of our Associate Editors and reviewers. The appointments of Les Bernstein and Marjorie Leek as Associate Editors were announced. They replace Wes Grantham and Bob Shannon whose excellent service was appropriately acknowledged. Based on the reports of the Associate Editors there may be a slight increase in the rejection rate and also a reduction in the number of submissions.

Six new members of the Technical Committee were elected during the meeting: Bruce Berg, Joan Besing, Emily Buus, Monita Chatterjee, Bob Frisina, and Andrew Oxenham. Their terms begin after the meeting in spring 2000. My thanks to the 11 people who agreed to be on the ballot and to the outgoing members, who prepared the ballot: Mary Florentine (Election Chair), David Eddins, John Grose, Marjorie Leek, Beth Strickland, and Fan-Gang Zeng. There was then a discussion of special sessions for Chicago. Four were proposed, including one joint session with Speech. The meeting concluded with a demonstration by Jont Allen of the Bell Labs CD-ROM.

P&P representation in ASA continues to improve. In the latest elec-

tions Bill Hartmann became president elect, Jan Weisenberger became vice-president elect, and Donna Neff was elected to the Executive Council. Congratulations to all three. Finally, P&P continues with its standard Technical Initiatives which include travel support for invited speakers, a student reception, and homepage maintenance. Suggestions for innovative uses for these funds, such as workshops, satellite meetings, etc., are welcome (nfv@tc.umn.edu).

NEAL VIEMEISTER
Chair 1999–2002

Signal Processing in Acoustics

Signal and image processing techniques are the technological thread that is entwined within each of the individual acoustical application areas represented by the Technical Committees (TCs). Over this year the Interdisciplinary Technical Group on Signal Processing in Acoustics (ITG-SP) has sponsored many technical sessions jointly with the TCs and individually on Processing Tutorials discussing critical technologies (e.g., wavelets, time-frequency, and model-based processing, etc.) in signal processing that directly impact many of the TC areas of high interest. The Gallery of Acoustics sponsored and administered annually by the ITG-SP has become a staple at the ASA meetings. Our new innovation of Technical Liaisons assures that we are continuously interacting at all levels with the TCs. The new Signal Processing Challenge Problem (under development) on the ASA website along with our own ITG-SP website provides a venue for signal processing students to interact with the ASA.

Signal and image processing are the underlying technologies that encompass many of the efforts in acoustics through the necessity to model, acquire, analyze and extract the required information from noisy measurement data. It is for these reasons that the Interdisciplinary Technical Group on Signal Processing (ITG-SP) has and continues to foster a growing interaction with each of the Technical Committees through the sponsoring of joint sessions and tutorials. This effort was made by the various signal-processing-related sessions throughout the Columbus and Atlanta meetings.

At the 138th ASA meeting in Columbus the ITG-SP sponsored or cosponsored nine sessions: two contributed sessions and seven special sessions. These sessions were entitled:

- ''Signal Processing Applications for Low Diffraction Beams'' (SP/Physical Acoustics), Chair: David H. Chambers, Co-Chair: D. Kent Lewis
- ''Signal Processing Techniques for Animal Bioacoustics'' (Animal Bioacoustics/Speech Communication/SP), Chair: Larry L. Pater
- ''Time-Frequency Applications in Acoustics I & II,'' (SP), Chair: Patrick J. Loughlin
- ''Bionic Ears: Influence on Perception and Production'' (Speech Communication/SP), Chair: Rosalie M. Uchanski
- ''Identification of Structural Systems for Model Validation and Health Monitoring'' (SP/Structural Acoustics and Vibration), Chair: David McCallen
- ''Wavelets in Acoustics'' (SP), Chair: Leon H. Sibul
- ''Multi-Channel Techniques in Signal Processing'' (SP), Chair: John Impagliazzo
- ''Signal Processing Applications in Acoustics'' (SP), Chair: Charles F. Gaumond.

James V. Candy also presented a ''Hot Topics'' lecture entitled, ''New Approaches in Acoustical Signal Processing'' which highlighted novel acoustic applications in time-frequency estimation, matched-field processing, time-reversal processing, and model-based signal processing.

We would also like to recognize our ''time-frequency and wavelet folks'' for organizing a full suite of special sessions during this year to help bring these exciting signal processing constructs to the ASA. We especially would like to mention L. Cohen, P. Loughlin, C. Gaumond, J. Impagliazzo, and L. Sibul and all of their efforts in stimulating the interest in this area.

The ''Gallery of Acoustics'' sponsored by the ITG-SP and compiled, set up, and displayed by Randall W. Smith and Preston S. Wilson was one of the highlights of the meeting. There was a gallery set up at the Columbus meeting and there will be another at the Newport Beach meeting. The Gallery provides ''creative'' acousticians the opportunity to display the results of their efforts in a more creative light. The 1999 entries in the Columbus Gallery of Acoustics were as follows.

Still Entries

- Those Twisted Beats (winner of judging and $350.00 cash prize), Pantelis Vassilakis, UCLA
- Single-Image Stereogram Visualizations of Acoustic Fields, T. Douglas Mast, Pennsylvania State University
- Side Look Sonar Images of a Bridge and a Surface Effect Vehicle, Thomas Hartka, Northrop Grumman Corporation

Audio Entry

- Reverse Time Stretching for Underwater Communication (winner of People's Choice Award), Fred G. Geil, Northrop Grumman Corporation

Video Entries

- Complex Motions of a Point on a Vibrating Wire, Roger J. Hanson, University of Northern Iowa
- Acoustic Landmine Detection Using Laser Scanning Vibrometer, James Sabatier and Ning Xiang, National Center for Physical Acoustics

The 139th ASA meeting in Atlanta, GA featured more joint ITG-SP sponsored/co-sponsored sessions with ten special sessions. These sessions were

- ''Signal Processing for Low Diffraction Beams'' (Physical Acoustics/SP), Chair: D. Kent Lewis
- ''Acoustical Imaging and Tomography'' (SP/Biomedical Ultrasound/Bioresponse to Vibration/Underwater Acoustics/Acoustical Oceanography), Chair: David H. Chambers
- ''Audio Engineering'' (Engineering Acoustics/SP), Chair: P. K. Raju, Co-Chair: David I. Havelock
- ''Auditory Temporal Processing in Animals'' (Animal Bioacoustics/SP), Co-Chairs: Cynthia F. Moss and Whitlow L. Au
- ''Time-Frequency and Wavelet Processing in Acoustics'' (SP/Noise), Co-Chairs: Leon H. Sibul and Charles F. Gaumond
- ''Underwater Signal Processing: Algorithms and Performance I,'' Chair: Charles F. Gaumond
- ''Underwater Signal Processing: Algorithms and Performance II,'' Chair: James C. Presig
- ''Audio Signal Data Compression for Musical Applications,'' Chair: James W. Beauchamp
- ''Model-Based Processing of Sources in Motion I & II'' (Underwater Acoustics/SP/Acoustical Oceanography), Co-Chairs: Peter G. Cable and William M. Carey.

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

Acoustical Oceanography	James Miller
Animal Bioacoustics	Brian Ferguson
Architectural Acoustics	Open
Biomedical Ultrasound/Bioresponse to Vibration	Shira Broschat/ Stergios Stergiopoulos
Engineering Acoustics	Stan Ehrlich
Musical Acoustics	John Impagliazzo
Noise	Dave Evans/Joe Pope
Physical Acoustics	Dave Havelock/Kent Lewis
Psychological & Physiological Acoustics	Open
Speech Communication	Jose Diaz
Structural Acoustics and Vibration	Stuart Bolton
Underwater Acoustics	Ed Sullivan/Geoff Edelson.

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

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

The efforts of David Evans on the Standards Committee are also important to the ITG-SP. Here is the report of his effort for this year. The International Organization for Standardization (ISO) Technical Committee on Mechanical Vibration and Shock (TC 108) circulated two New Work Item Proposals (NWIPs) for ballot in November 1999 on Signal Processing Methods for the Analysis of Stationary Mechanical Vibration, and Nonstationary Mechanical Vibration and Shock. Except for the few participating members of ISO/TC 108 that did not reply to this ballot, all participating members voted to approve these NWIPs with approximately 80% indicating that they were prepared to participate in the development of the projects. The working groups tasked with these new items, ISO/TC 108/WG 26 and ISO/TC 108/WG 27, are tentatively planning to meet jointly in the autumn of 2000.

We would also like to recognize Martin Barlett for creating, updating, and maintaining our website, another task that requires much effort. Surf our site and get all of the current tidbits of information and links to other processing websites. Our web address is http://www.arlut.utexas.edu/asaspweb/

JAMES V. CANDY
Chair, 1997–2003

Speech Communication

Productivity for the Speech Technical Committee has been exceptional this year with a number of individuals providing substantial contributions. In the following paragraphs, a brief account will summarize the activities of the Committee and describe key issues which are being considered by the Committee. Feedback on any issue is welcomed and may be addressed to any member of the Committee. Current members of the Committee are Mariko Aoki, Fredericka Bell-Berti, Shari Baum, Suzanne Boyce, Timothy Bunnell, Christopher Campbell, Randy Diehl, David Dorado, Carol Espy-Wilson, Marios Fourakis, Bruce Gerratt, John Hansen, Jean-Paul Haton, Sarah Hawkins, John Hawks, Mark Hedrick, Sun-Ah Jun, William Katz, Patricia Keating, Klaus Kohler, Qiquang Lin, Paul Luce, Nancy McGarr, Deborah Rekart, Ian Rogers, Robert Ruiz, Astrid Schmidt-Nielsen, Juergen Schroeter, Caroline Smith, Janet Stack, Winifred Strange, Mario Svirsky, Greg Turner, Rosalie Uchanski, Nagalapura Viswanath, Gary Weismer, and Doug Whalen. The *ex officio* members of the Committee are Keith Kluender, Anders Lofqvist, Douglas O'Shaughnessy, Winifred Strange, Maureen Stone, Christopher Turner, and Diane Kewley-Port. If you are interested in becoming a member of the Committee, please contact Emily Tobey.

Web Page Improvements

Bob Port has done a tremendous job developing the web page for the Speech Communication Committee ⟨asa.aip.org/committees.html⟩. The page describes the mission and scope of our division within the Society, lists interesting web sites related to speech communication, provides a listing of upcoming meetings related to speech communication, and describes student activities sponsored by our Committee. Please contact Bob Port if you have ideas regarding new additions to the page.

Student Activities

The Committee sponsors two activities, an evening reception and competition with a cash award for best oral or poster presentation, geared especially for students in Speech Communication at each ASA meeting. Attendance at the Columbus meeting student reception was excellent with approximately 45 students participating in the event. The reception for students at the Atlanta meeting was ''standing-room'' only with over 50 attendees from Speech Communication, Psychological and Physiological Acoustics, Animal Bioacoustics, and Musical Acoustics. Through the efforts of Richard McGowan, Sensimetrics Corporation donated $200 to our student reception in Atlanta. This donation is gratefully acknowledged by all who attended the reception.

John Hawks has headed a subcommittee whose responsibility is to judge student presentations at the meetings. Members of the Speech Communication community serve as judges to independently rate the presentations over the course of the meeting. The outstanding student presentation in Columbus went to J. Sean Allen and the outstanding student presentation in Atlanta went to Jing-Yi Jeng for her paper on ''Speech intelligibility and acoustic characteristics of Mandarin speakers with cerebral palsy.''

Melissa Epstein from the University of California Los Angeles served as our representative for the inaugral meeting of the ''ASA Student Council,'' an initiative started by ASA President, Pat Kuhl. She reported favorably on the meeting, and the items discussed by the students will form an exciting agenda throughout the upcoming years.

Editors

Current editors for our divisions within the Journal of the Acoustical Society are Keith Kluender, Anders Lofqvist, Douglas O'Shaughnessy, and Christopher Turner. They encourage all to submit manuscripts for consideration to the Journal. In addition, Jim Hillenbrand is editing sections of Acoustics Research Letters Online (ARLO).

Raymond H. Stetson Scholarship in Phonetics and Speech Production

Through donations made by Arthur Benton and Mac Pickett, a scholarship has been established in honor of Raymond Stetson. Members who are interested in contributing to this fund should send their donations to the Acoustical Society Foundation and specify the Stetson Scholarship. Students who are interested in motor speech production (motor coordination of speech organs, kinematics of speech articulation, aerodynamics of speech, or combined acoustic/kinematic modeling of speech), phonetics (linguistic issues related to speech, articulatory phonology, or experimental phonology), or bio-developmental factors in speech (neural function or infant/child development of speech communication) are invited to apply. Patricia Keating chaired a subcommittee composed of Abeer Alwan, Dani Byrd, and Peter Ladefoged to review the selection criteria and mechanics of selection. The subcommittee has revised the application form and streamlined the selection procedures. The person selected to receive the 2000 Stetson Scholarship was Elizabeth Johnson, Department of Psychology, Johns Hopkins University.

Special Initiatives

At the Seattle meeting, our committee sponsored a special session entitled ''Speech Communication: A Half Century of Speech Research.'' The session hosted invited papers by Gunnar Fant, Ken Stevens, Jim Flanagan, Al Liberman, L. Chistovich, Mac Pickett, Bob Porter, Katherine Harris, Peter Ladefoged, Vicki Fromkin, and Hiroya Fujisaki. Video-tapes were made of the presentations. This year, through a special ASA technical initiative, the University of Washington has made a composite tape available for members. Any member who has an idea for a special ASA technical initiative in Speech Communication should contact one of the Committee members.

Medals, Awards, and Fellows

Winifred Strange is the representative for Speech Communication to the Society's Medals and Awards Committee and Maureen Stone is our representative to the Membership Committee which makes decisions regarding Fellows of the Society. Please contact them if you wish to nominate an individual for either honor.

EMILY A. TOBEY
Chair 1998–2001

Structural Acoustics and Vibration

The primary business of the TC was conducted at its two regular meetings held at the Fall 1999 in Columbus, OH, and the Spring 2000 ASA meeting in Atlanta, GA. A full picture of our activites are reflected by the minutes of those meetings, which are available for viewing at http://fubini.swarthmore.edu/~bbard/savic.htm. Certain activities occur on a regular basis. These include discussing the editorial activity for JASA as a way of monitoring the level of activity in the research area, discussing and approving topics for special sessions at future ASA meetings, supervising and approving the Best Student Paper Award for submissions at the ASA meetings, and responding to queries from the Technical Council regarding matters affecting the membership. Our committee is also responsible for organizing the technical sessions in structural acoustics and vibration, which is achieved by volunteeer participation at the Technical Program Organizing Meetings.

One of the notable committee accomplishments was bringing a web

page on-line. (The address is cited above.) Our webmaster is Ben Bard of BBN, and Carr Everbach of Swarthmore College graciously provides the host computer. At the Fall 1999 meeting, the committee sponsored a Distinguished Lecture on MEMS devices given by Professor B. T. Khuri-Yakub of Stanford University. Special sessions sponsored by the committee at that meeting were ''Mid-frequency modeling in structural acoustics I & II,'' organized by Sean Wu of Wayne State University, and ''Acoustic nondestructive evaluation: New directions and techniques I & II,'' organized by Joseph Dickey of Johns Hopkins University. Special sessions at the Spring 2000 meeting were ''Aeroelasticity,'' co-organized by Ken Frampton of Duke University and Sean Wu, ''Combining active and passive control of vibration and noise,'' organized by Alain Berry of the University of Sherbrooke, and ''Tire vibration,'' organized by Stuart Bolton of Purdue University.

An election was held for Chair of the SAV Technical Committee in the winter of 2000. Scott Sommerfeldt of Brigham Young University was the winner, and took over the position at the Spring 2000 meeting.

JERRY H. GINSBERG
Chair 1997–2000

Underwater Acoustics

The Columbus meeting featured special sessions on ''Multiple Volumetric Scattering,'' chaired by David R. Palmer, and ''The Effect of Man-Made Sound on Marine Mammals,'' chaired by James F. Lynch and Peter L. Tyack and jointly sponsored by Acoustical Oceanography and Animal Bioacoustics. In addition, a memorial session honoring Robert J. Urick was jointly sponsored by Engineering Acoustics and chaired by David L. Bradley. Each invited talk in this session focused on Urick's contributions to our understanding of a particular component of the sonar equation and illustrated his broad influence on the field of underwater acoustics. Student activities included a graduate student reception, held jointly with Acoustical Oceanography, and a Best Student Paper Award competition. The winners were Michael R. Dungan (University of Michigan)—first prize; Michael J. Daley (Pennsylvania State University)—second prize; and George Kapodistrias (University of Washington)—third prize.

At the Atlanta meeting, Underwater Acoustics, Signal Processing in Acoustics, and Acoustical Oceanography sponsored a special session on ''Model-Based Processing of Sources in Motion,'' chaired by Peter G. Cable and William M. Carey. In the ''Interdisciplinary: Hot Topics in Acoustics'' session, George V. Frisk, with considerable input from John A. Colosi and Eric I. Thorsos, described two ''Hot Topics in Underwater Acoustics.'' They were ''Ocean Acoustic Wave Propagation in Random Media at Basin Scales'' and ''High-Frequency Sediment Acoustics.'' The Atlanta meeting also featured a lively and well-attended graduate student reception, held jointly with Acoustical Oceanography. Winners of the Best Student Paper Award competition were Purnima Ratilal (Massachusetts Institute of Technology)—first prize; Roger W. Schwenke (Pennsylvania State University)Msecond prize; and Chris J. Higham (University of Auckland)—third prize.

Finally, it was announced that John S. Perkins will be the new Underwater Acoustics Technical Committee Chair, starting immediately following the Atlanta meeting.

GEORGE V. FRISK
Chair 1997–2000

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.

2000

21–23 September	Eighth Annual Conference on the Management of the Tinnitus Patient, Iowa City, IA [Richard Tyler; Tel.: 391-356-2471; E-mail: tyler@uiowa.edu; WWW: www.medicine.uiowa.edu/otolaryngology/news/news].
22–25 October	IEEE Ultrasonics Symposium, San Juan, Puerto Rico [R. Almar, 896 Buttonwood Ln., Altamonte Springs, FL 32714; Fax: 407-290-5181; WWW: www.uffcsymp2000.org].
4–8 December	Joint Meeting: 140th Meeting of the Acoustical Society of America/NOISE-CON 2000, Newport Beach, CA [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

2001

4–8 February	Midwinter Meeting, Association for Research in Otolaryngology, St. Petersburg, FL [ARO Office, 19 Mantua Rd., Mt. Royal, NJ 08061; Tel.: 856-423-7222; Fax: 856-423-3420; E-mail: meetings@aro.org; WWW: www.aro.org/mwm/mwm.html]. Deadline for receipt of abstracts: 2 October 2000.
22–25 March	''New Frontiers in the Amelioration of Hearing Loss,'' St. Louis, MO [Sarah Uffman, CID Department of Research, 4560 Clayton Ave., St. Louis, MO 63110; Tel.: 314-977-0278; Fax: 314-977-0030; E-mail: suffman@cid.wustl.edu].
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]. Deadline for submitting abstracts: 14 July 2001.
4–8 June	141st Meeting of the Acoustical Society of America, Chicago, IL [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].
7–10 October	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.ieee-uffc.org/2001].

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

The Technical and Administrative Committees and Technical Groups listed below have been appointed by the Executive Council. These appointments, with such changes as may be made by the President from time to time, will be in effect until the Spring meeting of the Society in 2001.

Technical Committees 2000–2001

Acoustical Oceanography

	Term to
James F. Lynch, *Chair* to 2001	2001
Mohsen Badiey	2003
N. Ross Chapman	2003
Timothy F. Duda	2003
Oleg A. Godin	2003
Bruce M. Howe	2003
Iwao Nakano	2003
Marshall H. Orr	2003
David R. Palmer	2003
John R. Potter	2003
Konstantin D. Sabinin	2003
Hanne Sagen	2003
Alexandra I. Tolstoy	2003
Peter Traykovski	2003
Svein Vagle	2003
John A. Colosi	2002
Grant B. Deane	2002
Brian D. Dushaw	2002
David M. Farmer	2002
Alex E. Hay	2002
Kevin D. Heaney	2002

Stephen A. Reynolds	2002
David M. Rubenstein	2002
Holly A. Burch	2001
Marshall Hall	2001
Darrell R. Jackson	2001
Jeffrey A. Nystuen	2001
Lev A. Ostrovsky	2001
Robert Pinkel	2001
Peter F. Worcester	2001
Manell E. Zakharia	2001

Ex officio:
John C. Burgess, Associate Editor of JASA
David L. Bradley, Associate Editor of JASA
Stanley L. Chin-Bing, Associate Editor of JASA
Michael J. Buckingham, member of Medals and Awards Committee
Christopher Feuillade, member of Membership Committee
Er-Chang Shang, member of ASACOS

Animal Bioacoustics

	Term to
Mardi C. Hastings, *Chair* to 2003	2003
Ann E. Bowles	2003
John R. Buck	2003
Christopher W. Clark	2003
William C. Cummings	2003
Steven P. Dear	2003
Charles R. Greene	2003
Lynette A. Hart	2003
Van D. Holliday	2003
Sam H. Ridgway	2003
Richard W. Mankin	2003
William C. Burgess	2002
James J. Finneran	2002
Adam S. Frankel	2002
Darlene R. Ketten	2002
Larry L. Pater	2002
Peter L. Tyack	2002
Robert Hickling	2001
David A. Helweg	2001
Martin L. Lenhardt	2001
Peter M. Narins	2001
John R. Potter	2001
Daniel R. Raichel	2001

Ex officio:
Whitlow W. L. Au, Associate Editor of JASA
David L. Bradley, Associate Editor of JASA
Stanley A. Chin-Bing, Associate Editor of JASA
Whitlow W. L. Au, member of Medals and Awards Committee
Sam H. Ridgway, member of Membership Committee
Ann E. Bowles, member of ASACOS

Architectural Acoustics

	Term to
Ronald R. Freiheit, *Chair* to 2001	2001
Leslie D. Blomberg	2003
David Braslau	2003
Paul T. Calamia	2003
William J. Cavanaugh	2003
Daniel Clayton	2003
Murray R. Hodgson	2003
Bertram Y. Kinzey, Jr.	2003
Richard J. Peppin	2003
Scott D. Pfeiffer	2003
Roy L. Richards	2003
Kenneth P. Roy	2003
Ludwig W. Sepmeyer	2003
Gary W. Siebein	2003
Louis C. Sutherland	2003
Gregory C. Tocci	2003
Lily M. Wang	2003
Christopher N. Blair	2002
John S. Bradley	2002
Christopher N. Brooks	2002
Angelo J. Campanella	2002
Robert C. Coffeen	2002
Peter D'Antonio	2002
Timothy J. Foulkes	2002
J. Christopher Jaffe	2002
Mendel Kleiner	2002
David Lubman	2002
Michael T. Nixon	2002
Jack E. Randorff	2002
H. Stanley Roller	2002
Noral D. Stewart	2002
Bennett M. Brooks	2001
Steven M. Brown	2001
Richard H. Campbell	2001
M. David Egan	2001
Richard M. Guernsey	2001
Mark A. Holden	2001
K. Anthony Hoover	2001
Dana S. Hougland	2001
David W. Kahn	2001
Gary S. Madaras	2001
Charles T. Moritz	2001
Paul B. Ostergaard	2001
Dennis A. Paoletti	2001
David J. Prince	2001
Neil A. Shaw	2001
Richard H. Talaske	2001
Ewart A. Wetherill	2001
George E. Winzer	2001
Michael R. Yantis	2001

Ex officio:
Courtney B. Burroughs, Associate Editor of JASA
J. David Quirt, Associate Editor of JASA
Steven M. Brown, member of Medals and Awards Committee
Gregory C. Tocci, member of Membership Committee
George E. Winzer, member of ASACOS

Biomedical Ultrasound/Bioresponse to Vibration

	Term to
E. Carr Everbach, *Chair* to 2002	2002
Shira L. Broschat	2003
Robin O. Cleveland	2003
Ibrahim M. Hallaj	2003

Peng Jiang	2003
Joie P. Jones	2003
Inder R. Makin	2003
Janet M. Weisenberger	2003
Junru Wu	2003
Suk Wang Yoon	2003
Robert E. Apfel	2002
Michalakis A. Averkiou	2002
Michael R. Bailey	2002
Andrew J. Coleman	2002
Floyd Dunn	2002
John Erdreich	2002
Christy K. Holland	2002
Wesley L. Nyborg	2002
Joseph E. Piercy	2002
K. Kirk Shung	2002
Paul E. Barbone	2001
Anthony J. Brammer	2001
Diane Dalecki	2001
J. Brian Fowlkes	2001
Leon Frizzell	2001
Alan K. Goble	2001
Mark Hollins	2001
Kullervo Hynynen	2001
T. Douglas Mast	2001
Doug Miller	2001
Pierre Mourad	2001
William D. O'Brien	2001

Ex officio:
Floyd Dunn, Associate Editor of JASA
Laurel H. Carney, Associate Editor of JASA
Sid P. Bacon, Associate Editor of JASA
Wesley L. Nyborg, member of the Medals and Awards Committee
Anthony J. Brammer, member of the Membership Committee
Robin O. Cleveland, member of ASACOS

Engineering Acoustics

	Term to
Stephen C. Thompson, *Chair* to 2003	2003
Kim C. Benjamin	2003
Stanley L. Ehrlich	2003
Gary W. Elko	2003
Robert D. Finch	2003
Guillermo C. Gaunaurd	2003
Dehua Huang	2003
Sung Hwan Ko	2003
Theodore J. Mapes	2003
Victor Nedzelnitsky	2003
Yongrae Roh	2003
Ahmet Selamet	2003
James E. West	2003
George S. K. Wong	2003
Steven R. Baker	2002
David A. Brown	2002
Stephen C. Butler	2002
Robert D. Corsaro	2002
Stephen E. Forsythe	2002
Brian H. Houston	2002
W. Jack Hughes	2002
Arnie L. Van Buren	2002
Mahlon D. Burkhard	2001
James Christoff	2001
Dennis F. Jones	2001
Jan F. Lindberg	2001
Yushieh Ma	2001
Elizabeth A. McLaughlin	2001
Alan Powell	2001
Roger T. Richards	2001
Harold C. Robinson	2001
Kenneth D. Rolt	2001
Neil A. Shaw	2001
James Tressler	2001

Ex officio:
Ronald A. Roy, Associate Editor of JASA
Steven G. Kargl, Associate Editor of JASA
Stanley L. Ehrlich, Associate Editor of JASA
Robert D. Finch, member of Medals and Awards Committee
Thomas R. Howarth, member of Membership Committee
Mahlon D. Burkhard, member of ASACOS

Musical Acoustics

	Term to
James P. Cottingham, *Chair* to 2002	2002
Anders G. Askenfelt	2003
James W. Beauchamp	2003
Xavier Boutillon	2003
Matthew Brooke	2003
Murray D. Campbell	2003
Rene E. Causse	2003
Nicholas J. Giordano	2003
J. M. Harrison	2003
William M. Hartmann	2003
Norman C. Pickering	2003
James M. Pyne	2003
Daniel A. Russell	2003
Punita G. Singh	2003
Paul A. Wheeler	2003
Shigeru Yoshikawa	2003
George A. Bissinger	2002
Annabel J. Cohen	2002
Diana Deutsch	2002
Uwe J. Hansen	2002
Roger J. Hanson	2002
Peter L. Hoekje	2002
Douglas H. Keefe	2002
Ian M. Lindevald	2002
Stephen McAdams	2002
Gary P. Scavone	2002
Mark Pierre Verge	2002
Gabriel Weinreich	2002
R. Dean Ayers	2001
Judith C. Brown	2001
Courtney B. Burroughs	2001
Robert D. Collier	2001
Stephen F. Duncan	2001
George F. Emerson	2001
James H. Irwin	2001
Bozena Kostek	2001

Barry Larkin	2001
Thomas D. Rossing	2001
Julius O. Smith	2001
William J. Strong	2001

Ex officio:
Stanley L. Ehrlich, Associate Editor of JASA
R. Dean Ayers, Associate Editor of JASA
Roger J. Hanson, member of Medals and Awards Committee
Uwe J. Hansen, member of Membership Committee
Ian M. Lindevald, member of ASACOS

Noise

	Term to
Bennett M. Brooks, *Chair* to 2003	2003
Jorge P. Arenas	2003
Keith Attenborough	2003
Martin J. Beam	2003
Sergio Beristain	2003
Tarun K. Bhatt	2003
Leslie D. Blomberg	2003
Robert D. Bruce	2003
John C. Burgess	2003
Angelo J. Campanella	2003
William J. Cavanaugh	2003
Brian V. Chapnik	2003
Kejian Chen	2003
Kuo-Tsai Chen	2003
Li Cheng	2003
Raymond Cheng	2003
Robert J. Comparin	2003
James T. Cowling	2003
Malcolm J. Crocker	2003
Jamie Delannoy	2003
T. James DuBois	2003
John J. Earshen	2003
Jesse J. Ehnert	2003
Douglas K. Eilar	2003
Tony F. W. Embleton	2003
John Erdreich	2003
William J. Gastmeier	2003
Robert D. Hellweg	2003
David K. Holger	2003
Robert M. Hoover	2003
Patrick M. Hurdle	2003
Kazuo Ikegaya	2003
Daniel L. Johnson	2003
Jian Kang	2003
Seong-Woo Kang	2003
Tim Kelsall	2003
Tor S. Kihlman	2003
Sonoko Kuwano	2003
Mark A. Lang	2003
Chantal Laroche	2003
Yun-Hui Liu	2003
Robert Lotz	2003
Bjorn E. L. Lundquist	2003
Alan H. Marsh	2003
Henk M. E. Miedema	2003
Ruisen Ming	2003
Ikuharu Morioka	2003
Ralph Muehleisen	2003
Georgy L. Ossipov	2003
Kenneth J. Plotkin	2003
Joseph Pope	2003
Manuel Recuero	2003
Brigitte Schulte-Fortkamp	2003
Ben H. Sharp	2003
Lawrence Shotland	2003
Benjamin Soenarko	2003
Scott D. Sommerfeldt	2003
Kerrie G. Standlee	2003
Jing Tian	2003
Jean Tourret	2003
Ana M. Verzini	2003
Paul T. Weirich	2003
Kwangsee Allen Woo	2003
Barry R. Wyerman	2003
Serdar H. Yonak	2003
Robert W. Young	2003
James C. Yu	2003
Elliott H. Berger	2002
Stephen H. Bly	2002
Ann E. Bowles	2002
Frank H. Brittain	2002
Sanford Fidell	2002
Lee D. Hager	2002
Murray R. Hodgson	2002
Peter C. Laux	2002
Jerry G. Lilly	2002
David Lubman	2002
George A. Luz	2002
Sally A. McInerny	2002
Luc Mongeau	2002
Michael T. Nixon	2002
Matthew A. Nobile	2002
Robert A. Putnam	2002
Mary M. Prince	2002
Jack E. Randorff	2002
Stephen I. Roth	2002
Paul D. Schomer	2002
Hideki Tachibana	2002
James K. Thompson	2002
Nancy S. Timmerman	2002
Laura A. Wilber	2002
Martin Alexander	2001
John P. Barry	2001
Leo L. Beranek	2001
Arno S. Bommer	2001
Anthony J. Brammer	2001
James O. Buntin	2001
Robert J. Cook	2001
Kenneth A. Cunefare	2001
Joseph Cuschieri	2001
Paul R. Donavan	2001
William D. Gallagher	2001
Gerald C. Lauchle	2001
Anna Maria Monslave	2001
Miomir Mijic	2001
Thomas Norris	2001
John P. Seiler	2001
Noral D. Stewart	2001
Michael R. Stinson	2001

Alice H. Suter	2001
Louis C. Sutherland	2001
Jiri Tichy	2001
Dennis Walton	2001

Ex officio:
Michael R. Stinson, Associate Editor of JASA
Stanley L. Ehrlich, Associate Editor of JASA
Elliott H. Berger, member of Medals and Awards Committee
Daniel L. Johnson, member of Membership Committee
Richard J. Peppin, member of ASACOS

Physical Acoustics

	Term to
Robert M. Keolian, *Chair* to 2002	2002
Henry E. Bass	2003
Yves H. Berthelot	2003
James P. Chambers	2003
David I. Havelock	2003
Cila Herman	2003
Murray S. Korman	2003
Philip L. Marston	2003
Thomas J. Matula	2003
Philip S. Spoor	2003
Larry A. Wilen	2003
D. Keith Wilson	2003
Wayne M. Wright	2003
Evgenia A. Zabolotskaya	2003
Michael R. Bailey	2002
Robin O. Cleveland	2002
Robert A. Hiller	2002
Thomas J. Hofler	2002
R. Glynn Holt	2002
Christopher C. Lawrenson	2002
Timothy G. Leighton	2002
Bart Lipkens	2002
G. Douglas Meegan	2002
Ralph T. Muehleisen	2002
Penelope Menounou	2002
Richard Raspet	2002
John S. Stroud	2002
Richard L. Weaver	2002
David T. Blackstock	2001
David A. Brown	2001
Gilles A. Daigle	2001
Bruce Denardo	2001
Steven L. Garrett	2001
Logan E. Hargrove	2001
Craig J. Hickey	2001
Andres Larraza	2001
William Moss	2001
George Mozurkewich	2001
Vladimir Ostashev	2001
Victor W. Sparrow	2001

Ex officio:
Ronald A. Roy, Associate Editor of JASA
Steven G. Kargl, Associate Editor of JASA
Mack A. Breazeale, Associate Editor of JASA
Mark F. Hamilton, Associate Editor of JASA
Courtney B. Burroughs, Associate Editor of JASA
Floyd Dunn, Associate Editor of JASA
Dale E. Chimenti, Associate Editor of JASA
Andrew N. Norris, Associate Editor of JASA
Paul J. Remington, Associate Editor of JASA
Louis C. Sutherland, Associate Editor of JASA
James M. Sabatier, member of Medals and Awards Committee
Steven L. Garrett, member of Membership Committee
Sameer I. Madanshetty, member of ASACOS

Psychological and Physiological Acoustics

	Term to
Neal F. Viemeister, *Chair* to 2002	2002
Pierre Divenyi	2003
Richard Freyman	2003
Wesley Grantham	2003
Virginia Richards	2003
Barbara Shinn-Cunningham	2003
Patrick M. Zurek	2003
Huanping Dai	2002
Ann Clock Eddins	2002
Richard R. Fay	2002
Peggy B. Nelson	2002
Robert S. Schlauch	2002
Stanley E. Sheft	2002
Prudence Allen	2001
Robert P. Carlyon	2001
David F. Dolan	2001
Brent W. Edwards	2001
Nina Kraus	2001
John J. Rosowski	2001

Ex officio:
Laurel H. Carney, Associate Editor of JASA
Sid P. Bacon, Associate Editor of JASA
Leslie R. Bernstein, Associate Editor of JASA
Marjorie R. Leek, Associate Editor of JASA
Brenda L. Lonsbury-Martin, Associate Editor of JASA
Marjorie R. Leek, member of Medals and Awards Committee
Constantine Trahiotis, member of Membership Committee
Peggy B. Nelson, member of ASACOS

Speech Communication

	Term to
Emily A. Tobey, *Chair* to 2001	2001
Mariko Aoki	2003
Shari R. Baum	2003
Suzanne E. Boyce	2003
Christopher S. Campbell	2003
Randy L. Diehl	2003
David Dorado	2003
Bruce Gerratt	2003
John H. L. Hansen	2003
Jean-Paul Haton	2003
Sarah Hawkins	2003
Mark S. Hedrick	2003
Sun-Ah Jun	2003
Klaus J. Kohler	2003
Qiguang Lin	2003
Nancy S. McGarr	2003
Ian E. Rogers	2003

Robert Ruiz	2003
Caroline L. Smith	2003
Greg S. Turner	2003
Nagalapura Viswanath	2003
Carol Y. Espy-Wilson	2002
Marios S. Fourakis	2002
John W. Hawks	2002
William F. Katz	2002
Patricia A. Keating	2002
Paul A. Luce	2002
Deborah M. Rekart	2002
Juergen Schroeter	2002
Janet W. Stack	2002
Mario A. Svirsky	2002
Rosalie M. Uchanski	2002
Gary G. Weismer	2002
Fredericka Bell-Berti	2001
H. Timothy Bunnell	2001
Astrid Schmidt-Nielsen	2001
Winifred Strange	2001
Douglas H. Whalen	2001

Ex officio:
Keith R. Kluender, Assoicate Editor of JASA
Anders Lofqvist, Associate Editor of JASA
Douglas D. O'Shaughnessy, Associate Editor of JASA
Christopher W. Turner, Associate Editor of JASA
Winifred Strange, member of Medals and Awards Committee
Maureen L. Stone, member of Membership Committee
Diane Kewley-Port, member of ASACOS

Structural Acoustics and Vibration

	Term to
Scott D. Sommerfeldt, *Chair* to	2003
Benjamin Bard	2003
Alain C. Berry	2003
Joseph Dickey	2003
David Feit	2003
Allison Flatau	2003
Guillermo C. Gaunaurd	2003
Karl Grosh	2003
Sabih I. Hayek	2003
Francis Kirschner	2003
Jean R. Nicolas	2003
Allan D. Pierce	2003
Mauro Pierucci	2003
Vasundara V. Varadan	2003
Earl G. Williams	2003
Sean F. Wu	2003
John A. Burkhardt	2002
Kenneth D. Frampton	2002
Sunil Mehta	2002
Jeffrey S. Vipperman	2002
Richard L. Weaver	2002
Kuangcheng Wu	2002
Wen H. Lin	2001
Philip Marston	2001
James G. McDaniel	2001
Angie Sarkissian	2001

Ex officio:
Paul J. Remington, Associate Editor of JASA
Courtney B. Burroughs, Associate Editor of JASA and member of Membership Committee
Gideon Maidanik, member of Medals and Awards Committee
Louis A. Herstein, member of ASACOS

Underwater Acoustics

	Term to
John S. Perkins, *Chair* to	2003
Ralph N. Baer	2003
Kyle M. Becker	2003
John Buck	2003
Nicholas P. Chotiros	2003
Dezhang Chu	2003
David R. Dowling	2003
Steven Finette	2003
Paul C. Hines	2003
Anatoly N. Ivankin	2003
Finn B. Jensen	2003
Sunny Khosla	2003
John J. McCoy	2003
B. Edward McDonald	2003
Kazuhiko Ohta	2003
John R. Preston	2003
Catherine Stamoulis	2003
Ralph A. Stephen	2003
Frederick D. Tappert	2003
Alexander G. Voronovich	2003
James H. Wilson	2003
Lisa M. Zurk	2003
Paul B. Baxley	2002
Shira L. Broschat	2002
Douglas H. Cato	2002
Peter H. Dahl	2002
Grant B. Deane	2002
Gerald L. D'Spain	2002
Garry H. Heard	2002
Charles W. Holland	2002
Samuel W. Marshall	2002
Kevin B. Smith	2002
Dajun Tang	2002
Christopher T. Tindle	2002
Alexandra I. Tolstoy	2002
Stephen N. Wolf	2002
Michael G. Brown	2001
Dennis B. Creamer	2001
Christian P. de Moustier	2001
Stanley E. Dosso	2001
Stewart A. L. Glegg	2001
Zoi-Heleni Michalopoulou	2001
Marshall H. Orr	2001
Gregory J. Orris	2001
James C. Preisig	2001
Daniel Rouseff	2001
William L. Siegmann	2001

Ex officio:
John C. Burgess, Associate Editor of JASA
Stanley A. Chin-Bing, Associate Editor of JASA

David L. Bradley, Associate Editor of JASA and member of Medals and Awards Committee
Peter H. Rogers, member of Membership Committee
Arnie L. Van Buren, member of ASACOS

Interdisciplinary Technical Group on Signal Processing in Acoustics

	Term to
James V. Candy, *Chair* to 2003	2003
Gregory P. Clark	2003
Geoffrey S. Edelson	2003
Stanley L. Ehrlich	2003
Brian Ferguson	2003
William M. Hartmann	2003
Robert Huber	2003
John Impagliazzo	2003
Hua Lee	2003
Patrick J. Loughlin	2003
James C. Preisig	2003
Edmund J. Sullivan	2003
David J. Evans	2002
Charles F. Gaumond	2002
Joseph Pope	2002
Leon H. Sibul	2002
Randall L. Smith	2002
Martin Barlett	2001
David H. Chambers	2001
Elmer Hixson	2001
Ning Xiang	2001

Ex officio:
John C. Burgess, Associate Editor of JASA

Administrative Committees 2000–2001

Archives and History

	Term to
Henry E. Bass, *Chair* to 2001	2001
Mark C. Hedrick	2003
John W. Kopec	2003
Wesley L. Nyborg	2003
Richard J. Peppin	2003
Leo L. Beranek	2002
William J. Cavanaugh	2002
Logan E. Hargrove	2002
James C. Saunders	2001
Harry Schecter	2001
Rosalie M. Uchanski	2001

Books^{+}

	Term to
Mohsen Badiey, *Chair* to 2002	2002
David L. Bradley	2003
Christopher Feuillade	2003
Jerry H. Ginsberg	2003
Philip L. Marston	2003
Joseph Pope	2003
Robert A. Walkling	2003
Stephen N. Wolf	2003
Nancy S. McGarr	2002
Jeffrey A. Nystuen	2002
Neil A. Shaw	2002
Emily A. Tobey	2002
Stanley L. Chin-Bing	2001
Gordon E. Martin	2001
Victor W. Sparrow	2001

Ex officio:
Philip L. Marston, Associate Editor of JASA for Book Reviews

College of Fellows

William J. Cavanaugh, *Chair*

Peter G. Cable
M. David Egan
Robert D. Frisina
Marjorie R. Leek
Richard H. Lyon
Daniel R. Raichel
Astrid Schmidt-Nielsen
Jiri Tichy

All Fellows of the Society are members of the College of Fellows

Education in Acoustics

	Term to
Uwe J. Hansen, *Chair* to 2003	2003
Wlliam A. Ahroon	2003
Anthony A. Atchley	2003
Fredericka Bell-Berti	2003
Suzanne E. Boyce	2003
Robert D. Celmer	2003
Annabel J. Cohen	2003
E. Carr Everbach	2003
Thomas B. Gabrielson	2003
Katherine S. Harris	2003
Elizabeth S. Ivey	2003
Joie P. Jones	2003
Maria B. Mody	2003
Amy T. Neel	2003
P. K. Raju	2003
Deborah M. Rekart	2003
Daniel A. Russell	2003
M. Roman Serbyn	2003
Victor W. Sparrow	2003
Emily A. Tobey	2003
Courtney B. Burroughs	2002
Robin O. Cleveland	2002
Kenneth A. Cunefare	2002
D. Michael Daly	2002
Mary Florentine	2002
Roger J. Hanson	2002
Logan E. Hargrove	2002
Mardi C. Hastings	2002
Peter L. Hoekje	2002
Darrell R. Jackson	2002
Michel T. T. Jackson	2002
Murray F. Korman	2002

Luc Mongeau	2002
John S. Robertson	2002
James M. Sabatier	2002
Neil A. Shaw	2002
Kevin B. Smith	2002
Ralph A. Stephen	2002
James E. West	2002
Wayne M. Wright	2002
David A. Brown	2001
Robert D. Collier	2001
Corinne M. Darvennes	2001
Margaritis S. Fourakis	2001
Carole E. Gelfer	2001
Douglas R. Jones	2001
Sharon Y. Manuel	2001
Philip L. Marston	2001
Ana Maria Monsalve	2001
Ralph Muehleisen	2001
Andrew A. Piacsek	2001
Daniel R. Raichel	2001
Sally G. Revoile	2001
Thomas D. Rossing	2001
Ronald A. Roy	2001
Dawn R. Schuette	2001
Scott D. Sommerfeldt	2001
William Thompson, Jr.	2001
Robert A. Walkling	2001
George S.K. Wong	2001

Ethics and Grievances

	Term to
Robert E. Apfel, *Chair* to 2003	2003
William J. Cavanaugh	2003
Mardi C. Hastings	2001
David Feit	2001

Investments Committee

	Term to
John V. Bouyoucos, *Chair* to 2001	2001
Ira J. Hirsh	2002
Ira Dyer	2002
Kenneth M. Eldred	2001
David Feit, Treasurer, *ex officio*	

Long Range Planning

	Term to
Janet M. Weisenberger, *Chair* to 2002	2002
Lawrence A. Crum	2003
Robert J. Dooling	2003
Keith R. Kluender	2003
Mardi C. Hastings	2002
Scott D. Pfeiffer	2002
Stephen C. Thompson	2002
Nancy S. Timmerman	2002
Fredericka Bell-Berti	2001
Ervin R. Hafter	2001
Louis C. Sutherland	2001
William M. Hartmann, President-Elect, *ex officio*	

Medals and Awards

		Term to
D. Vance Holliday, *Chair* to 2001		2001
Elliott H. Berger	Noise	2003
Marjorie R. Leek	Psychological and Physiological Acoustics	2003
Wesley L. Nyborg	Biomedical Ultrasound/ Bioresponse to Vibration	2003
James M. Sabatier	Physical Acoustics	2003
Michael R. Stinson		
Steven M. Brown	Architectural Acoustics	2002
David L. Bradley	Underwater Acoustics	2002
Whitlow W. L. Au	Animal Bioacoustics	2002
Gideon Maidanik	Structural Acoustics and Vibration	2002
Michael J. Buckingham	Acoustical Oceanography	2001
Robert D. Finch	Engineering Acoustics	2001
Roger J. Hansen	Musical Acoustics	2001
Winifred Strange	Speech Communication	2001

Meetings

Dana S. Hougland, *Chair* to 2002
Yves H. Berthelot, Spring 2000, Atlanta
Joseph M. Cuschieri, Fall 2001, Ft. Lauderdale
Gilles A. Daigle, Vice President
Mardi C. Hastings, Fall 1999, Columbus
Elaine Moran, ASA Office Manager, *ex officio*
Scott D. Pfeiffer, Spring 2001, Chicago
Charles E. Schmid, Executive Director, *ex officio*
Sigfrid D. Soli, Fall 2000, Newport Beach
Jiri Tichy, Spring 1999, Berlin
Janet M. Weisenberger, Vice-President Elect

Membership

		Term to
Joseph W. Dickey, *Chair* to 2003		2003
Uwe J. Hansen	Musical Acoustics	2003
Thomas R. Howarth	Engineering Acoustics	2003
Sam H. Ridgway	Animal Bioacoustics	2003
Maureen Stone	Speech Communication	2003
Contantine Trahiotis	Psychological and Physiological Acoustics	2003
Anthony J. Brammer	Biomedical Ultrasound/ Bioresponse to Vibration	2002
Courtney B. Burroughs	Structural Acoustics and Vibration	2002
Burton G. Hurdle	Foreign Members	2002
Daniel L. Johnson	Noise	2002
Christopher Feuillade	Acoustical Oceanography	2001
Steven L. Garrett	Physical Acoustics	2001
Peter H. Rogers	Underwater Acoustics	2001
Gregory C. Tocci	Architectural Acoustics	2001

Public Relations

	Term to
Paul A. Baxley, *Chair* to 2003	2003
William J. Cavanaugh	2003
N. Ross Chapman	2003
Stanley E. Dosso	2003
Blas Espinoza-Varas	2003
Jack E. Randoff	2003
Thomas D. Rossing	2003
James C. Yu	2003
John Erdreich	2002
Logan E. Hargrove	2002
Burton G. Hurdle	2002
Ronald T. Kessel	2002
James F. Lynch	2002
Duncan E. McGehee	2002
Joanne L. Miller	2002
Joseph Pope	2002
Barbara J. Sotirin	2002
E. Carr Everbach	2001
Charles Gaumond	2001
Christy K. Holland	2001
David Lubman	2001
Andrew A. Piacsek	2001

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

Publication Policy

	Term to
Floyd Dunn, *Chair* to 2003	2003
Jont B. Allen	2003
Robert C. Bilger	2002
James F. Lynch	2002
James H. Miller	2002
Allan J. Zuckerwar	2002
Sigfrid D. Soli	2001
Richard Stern	2001
Michael R. Stinson	2001

William M. Hartmann, President-Elect, *ex officio*
Allan D. Pierce, Editor-in-Chief, *ex officio*

Regional Chapters

Roger T. Richards, *Chair* to 2001

Elmer L. Hixson	Austin
Angelo J. Campanella	Central Ohio
Courtney B. Burroughs	Central Pennsylvania
John W. Kopec	Chicago
Ernest M. Weiler	Cincinnati
Edwin H. Toothman	Delaware Valley
Gary W. Siebein	Florida
Yves H. Berthelot	Georgia
Timothy J. Foulkes	Greater Boston
Robert M. Hoover	Houston
Michael J. Anderson	Inland Northwest
Neil A. Shaw	Los Angeles
Hari S. Paul	Madras, India
Marehalli G. Prasad	Metropolitan New York
Elizabeth A. McLaughlin	Narragansett
Larry H. Royster	North Carolina
Peter F. Assmann	North Texas
James R. Angerer	Northwest
R. Dean Ayers	Orange County
Paul Baxley	San Diego
David Braslau	Upper Midwest
VACANT	Washington, DC
Thomas M. Disch	Wisconsin

Uwe J. Hansen, Chair, Education in Acoustics, *ex officio*

Rules and Governance

	Term to
Tony F. W. Embleton, *Chair* to 2002	2002
Elaine Moran	2003
Charles E. Schmid	2003
William M. Hartmann	2002
Richard H. Lyon	2002
William J. Cavanaugh	2001
Floyd Dunn	2001

Special Fellowships

	Term to
Wayne M. Wright, *Chair* to 2002	2002
Fredericka Bell-Berti	2003
James E. West	2003
Uwe J. Hansen	2002
John Erdreich	2001
Constantine Trahiotis	2001

Standards

Executive Committee
Daniel L. Johnson, Chair (Standards Director)
Paul D. Schomer, Vice Chair
Avril Brenig, Standards Manager, *ex officio*

S1 Representation
George S. K. Wong, Chair S1 and ASA representative on S1
Theodore J. Kuemmel, Vice Chair S1 and ASA alternate representative on S1

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

S3 Representation
R. F. Burkard, Chair S3 and ASA representative on S3
J. Franks, Vice Chair S3 and ASA alternate representative on S3

S12 Representation
Paul D. Schomer, Chair S2
Robert D. Hellweg, Vice Chair S12
Bennett M. Brooks, ASA representative on S12
William J. Galloway, ASA alternate representative on S12

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

ASA Technical Committee Representatives
Gilles A. Daigle, Chair of ASA Technical Council, *ex officio*
Er Chang Shang, Acoustical Oceanography
Ann E. Bowles, Animal Bioacoustics
George E. Winzer, Architectural Acoustics
Robin O. Cleveland, Biomedical Ultrasound/Bioresponse to Vibration
Mahlon D. Burkhard, Engineering Acoustics
Ian M. Lindevald, Musical Acoustics
Richard J. Peppin, Noise
Sameer I. Madanshetty, Physical Acoustics
Peggy B. Nelson, Psychological and Physiological Acoustics
Diane Kewley-Port, Speech Communication
Louis A. Herstein, Structural Acoustics and Vibration
Arnie L. Van Buren, Underwater Acoustics

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

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

Associate Editors for Standards News—JASA (ex officio)
Avril Brenig
George S. K. Wong

Tutorials

	Term to
Joseph Pope, *Chair* to 2001	2001
Fredericka Bell Berti	2003
Robin O. Cleveland	2003
George V. Frisk	2003
Yves H. Berthelot	2002
Uwe J. Hansen	2002
Ralph T. Muehleisen	2002
Ann E. Bowles	2001
Douglas D. O'Shaughnessy	2001
Gary W. Siebein	2001
Beverly A. Wright	2001

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

Women in Acoustics

	Term to
Nancy S. Timmerman, *Co-chair*	2003
Zoi-Heleni Michalpoulou, *Co-chair*	2003
Ina Rea Bicknell	2003
Margaret Cheesman	2003
Bozena Kostek	2003
Alexandra I. Tolstoy	2003
Peta White	2003
Uwe J. Hansen	2002
Mardi C. Hastings	2002
Michelle Langlais	2002
Peggy B. Nelson	2002
Linda M. Carroll	2001
Lawrence A. Crum	2001
Corinne M. Darvennes	2001
Laura K. Smith	2001

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).]*

The Psychology of Music (1999 second edition)

Diana Deutsch (Editor)

Academic Press, London.
807 pp., ISBN 0122135652, Price: Hard cover, $129.95; Soft cover, $69.95.

This edition of *The Psychology of Music* follows the first edition by 16 years (*The Psychology of Music*, 1982, also edited by D. Deutsch). Much has transpired in this interdisciplinary field in the interim, and in many respects, this timely volume fills the need for an update. The aim of this book is to ''interpret musical phenomena in terms of mental function . . . '' in ways that provide a comprehensive reference source for musicians as well as cognitive psychologists. Toward this end, the book comprises 18 chapters on topics ranging from psychoacoustics and neurological basis of sound to perception and performance of it. In this brief review, it is impossible to do justice to all of the authors of these chapters; instead, I highlight the most interesting chapters (in my opinion) with the aim of suggesting the strengths of this book.

The volume opens appropriately with a chapter by John R. Pierce on the nature of musical sound. Pierce begins delightfully by inviting the reader into the world of sound, sound sources, and the complexity of a sound scene. The chapter continues in this vein to remind readers of the basics of sound, including pitch, timbre, and spectra, temporal resolution of the ear and so on. It departs from the reader's expected formula for such a tutorial by offering interesting historical hypotheses, such a Galileo's rhythmic theory of consonance, and later introducing techniques for describing and reproducing musical sounds in general (the latter including not only linear prediction and Fourier techniques, but wavelet representations and newer approaches for plucking apart deterministic and stochastic aspects of a sound scene). This chapter certainly fits the bill for a tutorial for graduate students.

The chapters by Manford R. Schroeder and Norman M. Weinberger were likewise extremely valuable resources about important, but often inaccessible, topics. In Schroeder's chapter (Chapter 2) on concert halls: From magic to number theory, the reader is swept along by the author's fascinating account of concert hall acoustics. His descriptions of inadequacies in computer ray studies of reverberation times in various concert halls and the rise of ingenious computer wave studies and Monte Carlo simulations of complex transmission functions is quite nice. Also interesting, from the master of numerical magic (Schroeder, 1991), is Schroeder's description of elegant solutions to problems posed by modern (wide and low-ceilinged) concert halls which involves (of course) certain special frequency ratios. Weinberger's Chapter 3 on neurological aspects of musical sound is similarly instructive. For instance, many readers (including myself) who dip into the physiological literature aperiodically find that it is discouraging that much of the neurophysiological research on the response of the auditory system to ''music'' is based on impoverished stimuli, namely isolated pure tones. Weinberger brings together research to indicate that this practice is changing. Especially heartening is his coverage of recent research on the organization of the auditory cortex aimed at discovering whether it is based on frequency or pitch. He describes the nice interplay between human behavioral studies and theories (e.g., of virtual pitch of complex sounds) and animal studies (cats, birds, and monkeys do indeed hear virtual pitch). Also exciting is his review of research on responses of single neurons (and aggregates) to relative pitch information in a stimulus and (relatedly) the neurological encodings of musical intervals. For a nice tutorial on what's happening these days in the ''musical'' nervous system, I recommend this chapter.

Chapters 4 and 5 consider perception of musical tones (authors Rasch and Plomp) and explorations of musical timbre (authors Risset and Wessel). Both are useful. Rasch and Plomp review basic aspects of auditory psychophysics as well as issues of consonance and dissonance. The chapter by Risset and Wessel seemed long (56 pages) and detailed. It is pitched at a higher level of discourse than other chapters in this volume.

Sundberg's chapter on the perception of singing was most enjoyable. Certainly he presents the most succinct summary of research on various issues in perception of singing to date. His presentation captures one's attention thoroughly as he describes different strategies that trained singers (opera singers) employ to enhance resonances in the singing voice. Other interesting topics covered include vibrato and modes of expressive singing. This is an excellent chapter.

Chapters by Burns (''Intervals, Scales, and Tuning'') and Ward (''Absolute Pitch'') are well positioned within the volume. Burns' chapter is long and densely filled with information not simply on musical scales, their nature, and discussion of ''naturalness'' versus learnability, but also with detailed coverage of recent research on perception of musical intervals, categorical perception, and consonance. Paradigmatic distinctions (e.g., magnitude estimation versus absolute identification) as well as theoretical ones (trace-context theory, Terhardt's theory, etc.) are also succinctly discussed. Some relevant work on pitch circularity is missing and the critique of recent work by Schellenberg and Trehub (Schellenberg and Trehub, 1994a, 1996) seems misplaced, especially in light of Schellenberg and Trehub (1994b, not cited by Burns). Nevertheless, the chapter as a whole is valuable. Likewise, Ward's chapter is useful for his informative and readable answer to the question: ''Why doesn't everyone have absolute pitch (AP)?'' The measurement and stability of AP along with other issues are considered.

Deutsch's two chapters cover, respectively, basic grouping mechanisms in music and processing of pitch combinations. The former competently covers the fundamental aspects of grouping, which are conventionally interpreted in terms of Gestalt grouping principles. In the latter chapter, Deutsch reviews her own research on the tritone paradox and pitch memory, as well as feature abstraction and higher-order abstractions of pitch relationships.

Bharucha's chapter, ''Neural Nets, Temporal Composites, and Tonality'' (Chapter 11), presents a clear and instructional thesis on the application of neural nets to musical memory, learning, and the generation of musical (largely pitch) expectancies. The author pulls together both his earlier connectionist theorizing on harmony (e.g., MUSACT in Bharucha, 1987) with more recent work on the acquisition of schematic (style, tonal classes) and veridical (melodic sequences) expectancies to propose how neural nets ''learn'' musical sequences. A recurrent theme involves temporal composites, namely the mental representations of chords, keys, and even sequences acquired through temporal integration of elements presented to listeners at different times. Bharucha clearly explains some basics of neural net modeling (e.g., back propagation), illustrating how they apply to sequence learning. Only in this regard does one search for a comment on the relative merits of the author's models and other sequence learning models in the connectionist tradition that rely on a simple recurrent network architecture (e.g.,

 0001-4966/2000/108(3)/879/4/$17.00

Cleermans and McClelland, 1991). Finally, however, this is a solid tutorial that would be valuable for readers seeking an entry into neural networks in music.

Although the next chapter by Eugene Narmour seemingly offers a contrast to Bharucha's focus on neural networks and learning, in fact there are interesting comparisons because Narmour's focus in Chapter 12 is on the structure and hierarchical aspects of extraopus style. Bharucha distinguished schematic from veridical internal structures, whereas Narmour distinguishes extraopus structures that rest on top-down schemas form bottom-up primitives that realize hard-wired Gestalt rules. Both theorists emphasize that schemas are learned and that frequency and recency of exposure may be the primary determinants of acquiring musical schemas. They differ with regard to veridical expectancies about a particular melodic sequence, with Bharucha willing to attribute more to contextually modulated associative learning, while Narmour relies more on the listener's acquired sensitivity to inherent structural properties.

Narmour's chapter is a challenge for the reader, but it is worth the effort. It presents the author's theoretical perspective on the higher-order structures that underlie theme and variation (among other things) in music. It will help if the reader is familiar with Narmour's work; although he is careful to define his notation in footnotes and captions, Narmour's notation can be daunting to those without some preparation. The aim of the chapter is to demonstrate, through various analyses of musical texts, ways in which hierarchical levels of structure across different pieces capture similar aspects of musical style at a fairly high level. Key to understanding this is Narmour's treatment of hierarchical levels (issues of meter, beat level, and higher durational properties emerge). In a nutshell, Narmour proposes defining layers of musical structure in terms of the actual durations of various hierarchical levels. Although this is not an entirely novel idea (e.g., Jones, 1976), Narmour's development of it within a sophisticated analysis of musical structure carries much weight. His analysis also lays the groundwork for his speculation on neuronal representations of such structural components and their time spans. I liked this chapter and this speculation; however, I admit that this can be read as bias because these ideas are similar to ones that I have embraced.

Clarke's chapter, entitled ''Rhythm and Timing in Music,'' considers a few trends in studies of timing. Although it begins with a review of the important work of psychologist Paul Fraisse, finally this chapter may be of greater interest to musicians than to psychologists, for it fails to consider a number of psychological studies, empirical and theoretical, that are pertinent to issues of rhythm and timing.

Gabrielsson's chapter on the ''Performance of Music,'' which also focuses on rhythm and timing, is excellent. The chapter is long and comprehensive, covering virtually every aspect of performance in music, from the standard issues such as planning, sight-reading, motor processes, and expressive timing to the role of practice, historical issues (including detailed sections specific to performance on different instruments), and social factors and performance anxiety. No topic related to performance is overlooked. Gabrielsson is careful to include work from a wide range of researchers and theorists and is most even-handed and conscientious about presenting their work. Anyone interested in music performance as a subject of scientific study should read this chapter; it is a marvelous resource. I have only two (minor) reservations. The first is that in spite of the author's lucidity, the reader often finishes a section wondering, what was gained? The author's eclectic approach engages such a variety of divergent results from research on a common problem that a reader is sometimes left bemused. Relatedly, as a whole the chapter is heavy on the empirical side, but somewhat light on presenting traditional and emerging theories that might integrate diverse findings.

Dowling's chapter on the ''Development of Music Perception and Cognition'' is a model of a competent and thorough review of this field. He considers research on music perception and cognition in various age groups from infancy to adulthood. He clearly lays out parallels between findings with infants and adults in their responses to certain aspects of musical structure when these parallels are found to exist as they do in responses pitch, rhythm, and pitch contour change. He also identifies the principle difference between infant listeners and adults, namely differentiation of tonal from atonal melodies. Dowling also considers the development of rhythmic skill and discusses absolute pitch capabilities. Although this is an excellent review, its heavy emphasis on empirical studies with relatively little examination of theoretical models is perhaps telling. It leads to the inference that this research area is one of several in this volume that remains largely in a data-gathering mode. This inference is reinforced by the chapter on musical ability by Shuter-Dyson. This chapter provides a useful resource for readers seeking information on measures of musical ability and musical potential.

The final two chapters, one on ''Neurological Aspects of Music Perception and Performance'' (authored by Oscar S. M. Marin and David W. Perry), and the other on ''Comparative Music Perception and Cognition'' (authored by Edward C. Carterette and Roger A. Kendall) are both long and challenging. The first of these gains the reader's gratitude by confirming our suspicion that '' . . . the optimistic view of classical neurology that musical functions are clearly localized has not, in most cases been verified.'' The chapter is dense, reviewing data collected from both experimental work and case studies on a range of topics including auditory agnosia, verbal deafness, and music perception. This is clearly an evolving field. In that respect, this chapter shares with the final chapter on comparative music a tough assignment. Both chapters cover domains that comprise many seemingly unrelated topics. In Chapter 18, Carterette and Kendall address some of the problems of comparative music cognition involving methodology and data analysis. A number of interesting studies on tonality, melodic contour, and so on are reviewed. The chapter continues, ranging widely into species differences (e.g., birdsong) and neurophysiology. In some places, the organizational support for the chapter slows a reader.

These brief sketches suggest the variety of topics, treatments, and perspectives contained in this volume. A number of these are excellent. In general, the placement of chapters based on related topics by the editor was nicely done. With the exception of separating the two neurological chapters (by Weinberger and Marin and Perry), other chapters with closely related themes or domains were well positioned to facilitate reader comprehension. Some topics are over-represented and others under-represented. Redundancy occurs among chapters authored by Rasch and Plomp, and Burns and Deutsch with respect to topics such as pitch intervals, consonance, and absolute pitch; but, the topic of tonality is under-represented in this volume. A significant change in the field since the last edition involves emergence of research on tonality, spearheaded largely by Krumhansl (e.g., Krumhansl, 1990); a more focused treatment of this important topic would have strengthened this volume. A chapter on emerging trends might also have been useful. Even in the chapters presented, older work is often emphasized relative to that breaking in the nineties; to the degree that this occurs, the volume is less than up-to-date in this quickly changing field. Relatedly, to interject a personal bias, for freshness and continuity I think major volumes such as this should attempt to include voices of a few of the younger, but influential, scholars in the field (e.g., Boltz, Drake, Gjerdingen, London, Palmer, Schellenberg, and Trainor, to mention only a few). Finally, some chapters were marred by failure of authors and/or copy editors to carefully proof. For instance, in Chapter 2 (page 31), the relationship of the written text to Fig. 41 is simply not clear. Another example occurs in Chapter 18 (page 774) with the confusing sentence: ''A possibility that birdsongs may learn complex, tonal qualities of birdsong was shown by . . . .'' A number of such mistakes occur, rendering them distracting to the reader.

In summary, there is much of value in this book. Generally, the chapters are aimed at the appropriate level for engaging both experts and novices in the field. Not surprisingly, some chapters are stronger and more readable than others. Similarly, some are more selective and/or theoretically oriented, whereas others are straightforward presentations of empirical findings. To a significant degree, the book succeeds in its goal of providing a comprehensive resource for the field. With respect to the aim of ''interpreting musical phenomena in terms of mental functions . . . ,'' indeed a wealth of interpretation is provided, but (predictably) with little agreement on the nature of this mental function. Finally, the picture of the field that emerges from this book is one of a still-new scientific domain that has made significant progress in the 16 years since the first volume was published. At the same time, it is clear that few overarching themes or paradigms have been developed to provide cohesion for the field as a whole and that as a domain, the psychology of music is still seeking solid ground within the larger fields of both psychology and music.

Bharucha, J. J. (**1987**). ''MUSACT: A connectionist model of musical harmony,'' in *Proceedings of the Cognitive Science Society* (Erlbaum, Hillsdale, NJ).

Cleermans, A., and McClelland, J. L. (**1991**). ''Learning the structure of event sequences,'' J. Exp. Psychol. **120**, 235–253.

Jones, M. R. (**1976**). ''Time, our lost dimension: Toward a new theory of perception, attention, and memory,'' Psychol. Rev. **83**, 323–355.
Krumhansl, C. (**1990**). *Cognitive Foundations of Musical Pitch* (Oxford University Press, New York).
Schellenberg, E. G., and Trehub, S. E. (**1994a**). ''Frequency ratios and the discrimination of pure tone sequences,'' Percept. Psychophys. **56**, 472–478.
Schellenberg, E. G., and Trehub, S. E. (**1994b**). ''Freqeuncy ratios and the perception of tone patterns,'' Psychonom. Bull. Rev. **1**, 191–201.
Schellenberg, E. G., and Trehub, S. E. (**1996**). ''Natural musical intervals: Evidence from infant research,'' Psychol. Sci. **7**, 272–277.
Schroeder, M. (**1991**). *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise* (W. H. Freeman, New York).

MARI RIESS JONES
Department of Psychology
The Ohio State University
Columbus, Ohio 43210

Theoretical Acoustics of Underwater Structures

E. A. Skelton and J. H. James

Imperial College Press, 1997.
Price: $78, ISBN: 1860940854 (distributed by World Scientific Publishing Co.).

This clearly written and fairly comprehensive book is a welcome addition to the somewhat limited collection of books devoted to the vibroacoustic response of submerged structures, a subject sometimes referred to as *structural acoustics*. This important book provides an account of the linear acoustics of basic isotropic and anisotropic structures excited by time-harmonic and transient mechanical forces and acoustic sources. The first book dedicated to this specialized subject *Sound, Structures, and their Interaction* by M. C. Junger and D. Feit was first published close to 30 years ago in 1972, and is now in its third edition.

The present book makes extensive reference to the former in 7 of the 14 chapters of the book, and in some respects can be considered as complementary to it especially with regard to its inclusion of numerical approaches that are now so much in use. There are many very useful numerical results presented here using formulations and equations first presented in the Junger–Feit book. The numerical examples provide physical insight and can be used as benchmark results for computations of sound radiation and sound scattering. In addition, the book does an excellent job of including the details of more complex structural configurations than dealt with earlier. For example in Chapter 8, the problems of frame and bulkhead stiffened and orthogonally stiffened water-loaded plates are treated together with numerical results for far-field radiated pressure. One notable deficiency however, is the lack of material related to frequency-wave-number analysis of the structural response field. This is now one of the more ubiquitous analysis tools used by structural acousticians, and it is unfortunate that there is very little mention of this. I did find one figure in Chapter 9 showing the dispersion curves for a water-filled cylindrical shell *in vacuo*. In general, this reviewer found that the visual quality of the problem illustrations were somewhat lacking.

The first three chapters are introductory in nature. Chapter 1 presents the mathematical tools necessary for solving problems in theoretical acoustics and also introduces numerical integration procedures useful for the finite element approaches presented in later chapters. The dynamical response, both transient and time-harmonic, of mechanical systems is treated in Chapter 2, while Chapter 3 deals with the acoustic medium and the fundamental solutions in different coordinate systems.

The few canonical problems of acoustics, such as scattering of sound by soft and hard spherical, cylindrical and planar boundaries are treated in the next chapter. For those boundaries, not treatable by separation of variable solutions, numerical techniques such as finite element formulations are necessary, and these are introduced in Chapter 5. Chapter 6 presents the equations of linear elasticity for isotropic materials. This is useful for the book's later treatment of the structural acoustics of systems using fiber reinforced composite components. The use of elasticity theory in describing the response of the structural components is a major departure from the Junger–Feit approach. In the latter, the use of the term *structural* implies a description of the mechanical system using the structural equations of motion as opposed to the elastodynamic equations of motion.

Following these introductory chapters, the author's begin their treatment of underwater structure response and radiation in Chapter 7 by presenting the solution for that of an elastic spherical shell immersed in an acoustic medium. This is referred to as ''the baseline problem of fluid-structure interaction.'' This configuration is the only one for which closed form expressions for both the structural response and the acoustic field can be obtained, and thus provides a benchmark problem with which computational approaches can be compared and validated. The next chapter considers the most frequently studied problem of structural acoustics, that is, the force-excited fluid-loaded plate of inifinite extent with or without attached dynamical systems. This problem embodies all the physical mechanisms of acoustic radiation, and the results can be obtained with minimal computational effort. The following chapter covers the structural acoustic problems of thin cylindrical shells. The current book uses a slightly different version of thin shell theory than that used by Junger–Feit and also considers the problem of rib-stiffened shells, not considered elsewhere. The following three chapters follow the pattern of the previous three, except that the structures are now layered, and the response is governed by the equations of elasticity rather than thin shell theory. Chapter 14, the final chapter of the book, deals with finite axisymmetric structures that may have anisotropic properties. The approach outlined in this chapter makes extensive use of finite element formulations.

In summary this book will serve as an excellent reference for graduate students, research scientists, and engineers who are interested in the structural acoustics of marine structures. As stated earlier, it can be viewed as an excellent complement to the Junger–Feit book that, although still in print (published by ASA) is now close to 30 years old, but lacks the many numerical results presented in the current book.

DAVID FEIT
Signatures Directorate
Carderock Division Naval Surface Warfare Center
9500 MacArthur Boulevard
W. Bethesda, Maryland 20817

Acoustic Cavitation and Sonoluminescence

J. R. Blake

Philosophical Transactions of the Royal Society: Mathematical, Physical and Engineering Sciences Volume: 357; Issue: 1751.

This collection of articles is very welcome both to experts, and to the general audience. The research in the field of acoustic cavitation and in the related effects of single and multibubble sonoluminescence (SBSL and MBSL in what follows) has been developing very rapidly during the last decade or so. *Acoustic Cavitation and Sonoluminescence* provides an overview of recent research in the field, leading the reader from the simplest case of a single bubble, all the way up to applications, such as acoustic cleaning, crystallization, or synthesis of biomaterials. Covering the details of all aspects of such a big field in less than 200 pages is, of course, impossible. However, the issue does a very good job in outlining the main directions of the recent research, with the right balance of experiment and theory.

The first article, by A. Prosperetti and Y. Hao, gives a basic picture of the dynamics of a single, spherical bubble in acoustic field. The basic physical principles are covered, as well as the very recent finding that nitrogen and oxygen tend to escape from an oscillating bubble into surrounding liquid, leading to purely noble (argon) bubbles. The authors outline the various factors that are of relevance to bubble dynamics and SBSL, such as stability of a bubble with respect to the perturbations from a spherical shape, mass and heat flow. They give a very useful overview of the approximations needed to keep the model tractable. One of the main assumptions—that of the spatial uniformity of the gas in the bubble—is recognized by the authors as a still open question, related to the basic principles leading to light emission.

The second article, by T. J. Matula, tells us about experimental research of the dynamics of a single bubble and SBSL, and its connection to light emission from bubble clouds (MBSL). The reader can learn about state of the art experimental techniques which proved extremely valuable in understanding the basic picture. At the end of the article, comparison between SBSL and MBSL is made, emphasizing both the similarities and the differences between these two effects.

The next article, by J. R. Blake *et al.*, extends to the more complicated case of a bubble oscillating close to a physical boundary, and also to multibubble situation. Using boundary integral methods, the authors show how the interaction of a bubble with a boundary, or with other bubbles, can lead to deviations from a spherical shape, jet formation, and other effects. The authors recognize both the importance of this research to applications, and the fact that significant new research is needed to understand the effects such as dynamics of a bubble cloud (this problem is addressed later in the issue).

The article by C.-D. Ohl *et al.* is concerned with the different, but closely related effect of laser induced cavitation. Much larger bubbles are produced in this manner—a bubble could reach millimeter scale, while the acoustic cavitation usually deals with bubbles of a few microns. The increased bubble size and the corresponding changes in bubble dynamics allow for more precise insight into the interior of a bubble, possibly shedding new light to acoustic cavitation as well. The observation that the size of the light emitting region is limited to the very center of a bubble could be very relevant to the light emission in SBSL. One still open question in SBSL, namely are there converging shock waves in a collapsing bubble, leading to ''hot spot'' and optical radiation, appears to be supported by these results, even though on very different length scales. The second part of the article outlines experimental results concerning multibubbles, and serves as a useful introduction to a latter (page 313) article by Parlitz *et al.*

The article by B. K. Bourne and J. E. Field on the subject of dynamics of bubble clouds, presents a different problem of shock-induced luminescence. The authors discuss experimental results for asymmetric collapse of cavities when acted upon by a plane shock wave. In particular, applications of this process to the problem of explosive ignition is discussed in some detail. Possible analogy to the light emission process in SBSL is discussed; however, the authors correctly recognize that there is a significant difference both in the geometry of the presented problem and the physical mechanism leading to the collapse of the cavities/bubbles in these two cases.

The work by Parlitz *et al.* considers theoretical and modeling issues related to the dynamics of bubble clouds. This work is complementary to the earlier article by Blake *et al.*, which concentrates on the shape oscillations of bubbles. The authors show that homogeneous bubble density is unstable, and use this result to explain experimentally observed structure of bubble clouds. It is pointed out that an extension of this approach, which would include neglected effects (e.g., shape oscillations of the bubbles, liquid streaming, etc.) is needed in order to obtain better insight into this complex phenomenon.

The final two articles, by K. S. Suslick *et al.*, and by T. J. Mason, give concise, but well documented overview of the applications of acoustic cavitation. After reviewing the similarities and the differences between SBSL and MBSL, the article by K. S. Suslick *et al.* concentrates on one of the important applications of acoustic cavitation–chemical synthesis. The interested reader will find a number of results concerning the synthesis of both inorganic compounds (such as amorphous metals), and biomaterials (protein microspheres). A more general overview is given in the article by T. J. Mason, where a variety of industrial applications are outlined, in such diverse fields as welding, atomization, cleaning, chemical processing, and many others.

This collection of articles will prove very valuable to a number of researchers interested in gaining insight into acoustic cavitation and sonoluminescence. The careful reader will notice, however, that one aspect is almost completely missing in this issue—the discussion of physical origins of single bubble sonoluminescence. Perhaps this is appropriate, since there is still no concensus concerning the final stages of bubble collapse, and related physical and chemical processes leading to SBSL. My opinion is that this collection of articles will help us obtain answers to this and other questions in the exciting field of acoustic cavitation.

DR. LJUBINKO LOU KONDIC
Department of Mathematical Sciences
New Jersey Institute of Technology
Newark, New Jersey 07102

BOOKS RECEIVED

Electronic Concepts: An Introduction. J. H. Krenz. Cambridge University Press, 2000. 454 pp. $49.95 *hc* ISBN 0521662826.

Fundamentals of Modeling and Analyzing Engineering Systems. Philip D. Cha, James J. Rosenberg, and Clive L. Dym. Cambridge University Press, 2000. 466 pp. $100.00 *hc* ($44.95 *pb*) ISBN 052159443X (0521594634 *pb*).

Elastic Waves in Solids I: Free and Guided Propagation. Daniel Royer and Eugène Dieulesaint. Springer-Verlag New York, Inc., 1996. 374 pp. $82.00 *hc* ISBN 3540659323.

Elastic Waves in Solids II: Generation, Acousto-optic Interaction, Applications. Daniel Royer and Eugène Dieulesaint. Springer-Verlag New York, Inc., 1999. 446 pp. $89.95 *hc* ISBN 3540659315.

Autoparametric Resonance in Mechanical Systems. A. Tondl, T. Ruijgrok, F. Verhulst, and R. Nabergoj. Cambridge University Press, 2000. 196 pp. $59.95 *hc* ISBN 0521650798.

Introduction to Experimental Nonlinear Dynamics: A Case Study in Mechanical Vibration. Lawrence N. Virgin. Cambridge University Press, 2000. 256 pp. $74.95 *hc* ($32.95 *pb*) ISBN 0521662869 (0521779316 *pb*).

Speech Processing and Synthesis Toolboxes. D. G. Childers. John Wiley & Sons, 2000. INCLUDES SOFTWARE. 483 pp. $90.95 *hc* ISBN 0471349593.

Advances in Phonetics. *Proceedings of the International Phonetic Sciences Conference (IPS)*. A. Braun, ed. Franz Steiner Verlag, Stuttgart, 1999. 144 pp. DM 76.00 *pb* ISBN 3515076557.

Phonetics—The Science of Speech. M. J. Ball and J. Rahilly. Oxford University Press, New York, 2000. 239 pp. $65.00 *hc* ($19.95 *pb*) ISBN 0340700092 (0340700106 *pb*).

Acoustics for Audiologists. Edgar Villchur. Singular Publishing Group, 2000. 125 pp. $53.95 *pb* ISBN 0769300642.

Acoustics Applied to Music. Howard Pollard. Gleebooks (distributor), Glebe, Australia, 1999. 218 pp. $29.95 *pb* ISBN 0646380656.

The Design of Animal Communication. M. D. Hauser and M. Konishi, eds. MIT Press, 1999. 701 pp. $65.00 *hc* ISBN 0262082772.

Auditory Worlds: Sensory Analysis and Perception in Animals and Man. G. A. Manley, H. Fastl, M. Kossl, H. Oeckinghaus, and G. Klump, eds. Wiley-VCH, 2000. 359 pp. DM 198.00 ISBN 3527275878.

Comparative Hearing: Birds and Reptiles. Robert J. Dooling, Richard R. Fay, and Arthur N. Popper, eds. Springer-Verlag New York, Inc., 2000. 380 pp. $98.00 *hc* ISBN 0387946845.

Comparative Hearing: Fish and Amphibians. Richard R. Fay and Arthur N. Popper, eds. Springer-Verlag New York, Inc., 1998. 438 pp. $89.95 *hc* ISBN 0387984704.

Comparative Hearing: Insects. Ronald R. Hoy, Arthur N. Popper, and Richard R. Fay, eds. Springer-Verlag New York, Inc., 1998. 341 pp. $89.95 *hc* ISBN 0387946829.

Development of the Auditory System. Edwin W. Rubel, Arthur N. Popper, and Richard R. Fay, eds. Springer-Verlag New York, Inc., 1998. 431 pp. $89.95 *hc* ISBN 0387949844.

Hearing by Whales and Dolphins. Whitlow W. L. Au, Arthur N. Popper, and Richard R. Fay, eds. Springer-Verlag New York, Inc., 2000. 485 pp. $129.00 *hc* ISBN 0387949062.

The Science and Applications of Acoustics. Daniel R. Raichel. Springer-Verlag New York, Inc., 2000. 598 pp. $69.95 *hc* ISBN 0387989072.

Now Hear This: The Life of Hugh S. Knowles, Acoustical Engineer and Enterpreneur. Susan Goodwillie. Francis Pr, 1999. 215 pp. $24.00 *hc* ISBN 0966505123.

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.

Reviewers for this issue:

KEVIN P. SHEPHERD, *M.S. 463, NASA Langley Research Center, Hampton, Virginia 23681*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*

6,012,334

43.38.Md VIBRATION WAVE DETECTING METHOD AND VIBRATION WAVE DETECTOR

Shigeru Ando *et al.*, assignors to Sumimoto Metal Industries Limited
11 January 2000 (Class 73/651); filed in Japan 26 May 1997

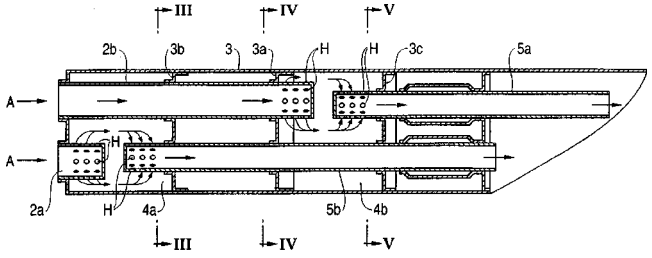
The object of the invention described in this patent is to provide a single piece of mechanical hardware that can determine the frequency spectra of acoustic signals rapidly and accurately, primarily to assist in automated speech recognition. The device described in this patent consists in essence of an array of cantilever reeds, each tuned to a predetermined frequency, and each interacting with an electrode to generate an electric signal that is integrated over a selected time interval. The reeds may be tuned so that their resonances are distributed linearly in the mel scale or in the Bark scale.—EEU

5,979,583

43.50.Gf MUFFLER FOR MOTORCYCLE

Hideo Amino *et al.*, assignors to Honda Giken Kogyo Kabushiki Kaisha
9 November 1999 (Class 180/219); filed in Japan 25 July 1996

This muffler for a multi-cylinder motorcycle engine provides independent exhaust sound for each cylinder. The parallel exhaust paths each contain an expansion chamber. The forward cylinder of a V-type two cylinder engine is attached to the lower inlet in the figure.—KPS

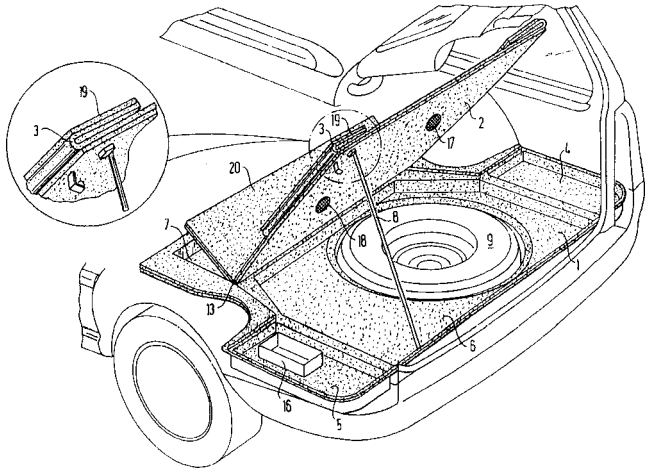


5,979,962

43.50.Gf LOAD FLOOR LINING HAVING INTEGRATED SOUND INSULATION

Erhard Valentin *et al.*, assignors to Stankiewicz GmbH
9 November 1999 (Class 296/37.1); filed in Germany 10 June 1994

A double-wall floor lining for an automobile trunk is described which consists of a hinged, flat upper wall and a lower surface which conforms to the vehicle shape. Enhanced transmission loss which is achieved by the double-wall construction would seem to be potentially threatened by filling of the storage spaces which are located between the walls. It is suggested that the space between these walls could also be used to house the electronic components for an active control system.—KPS



5,980,343

43.50.Gf EXHAUST SYSTEM FOR MARINE VESSELS

Adam Rolinski, assignor to S2 Yachts, Incorporated
9 November 1999 (Class 440/89); filed 25 February 1998

This marine muffler utilizes a two-pathway exhaust and water discharge system. At low engine speeds exhaust gases and cooling water are discharged through an outlet above the water line. At higher speeds, the exhaust gases are discharged through an underwater outlet and the cooling water is discharged through the above water outlet. Thus noise reduction is achieved at higher speeds and the problem of ''burping'' at low speeds is avoided.—KPS

5,981,046

43.50.Gf SOUND ABSORBING COMPONENT

Syohei Masui *et al.*, assignors to Sumitomo Chemical Company, Limited; Kawasaki Steel Corporation; K-Plasheet Corporation
9 November 1999 (Class 428/213); filed in Japan 13 March 1995

Thin, sound-absorbing material systems intended for use in automobile engine compartments are described. A fiber-reinforced thermoplastic resin, expanded such that it contains at least 50% voids (by volume), is coupled with a resin molded body with an air layer in between. Another variation consists of the resin molded body being replaced by a thermoplastic foamed

 0001-4966/2000/108(3)/883/6/$17.00

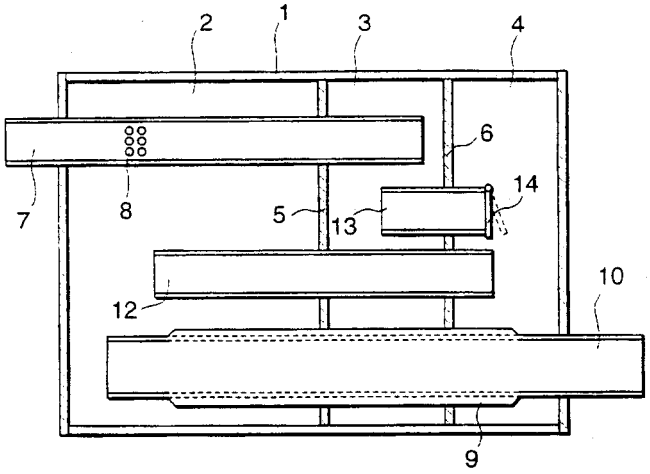
body. Various geometrical arrangements are described and normal incidence sound absorption coefficients are provided.—KPS

5,984,045

43.50.Gf ENGINE EXHAUST NOISE SUPPRESSOR

Kazushige Maeda and Akira Sasaki, assignors to Nissan Motor Company, Limited
16 November 1999 (Class 181/254); filed in Japan 14 February 1997

There are two exhaust paths in this muffler design. The first path consists of the inlet tube **7**, small holes **8**, expansion chamber **2**, and exit tube **10**. The second path consists of additional chambers **3** and **4** and connecting pipes as shown. This second path is activated at high engine speeds



by the valve **14**. Since this second path has two degrees of freedom, the resonance frequencies can be chosen to give noise reduction over a wide frequency range.—KPS

5,996,733

43.50.Gf DUAL FREQUENCY SIDE BRANCH RESONATOR

Jon A. DeTuncq and Steven M. Gleason, assignors to Thermo King Corporation
7 December 1999 (Class 181/250); filed 20 November 1998

This resonator is intended as a compact means for attenuating two frequency-components of the air inlet or outlet noise of a machine, such as a refrigeration compressor. The resonator described in this patent consists in essence of a tube that communicates at one of its ends with the pipe that carried the airflow to or from the machine. The length of the tube between its inlet and its other end is made equal to a quarter of the wavelength of the lower-frequency tone that is to be attenuated, and the far end is closed off by an end cap. A plate with an orifice is installed in the tube at a distance from the inlet that is equal to one-quarter of the wavelength of the higher-frequency tone that is to be attenuated.—EEU

5,996,734

43.50.Gf AIR INTAKE SILENCER FOR A TWO-STROKE ENGINE

Edward Lam *et al.*, assignors to Outboard Marine Corporation
7 December 1999 (Class 181/255); filed 11 March 1998

The silencer described in this patent consists of a housing that defines a passage with a relatively small height and with a length that corresponds to a quarter of the wavelength of the tonal sound that is to be suppressed. The width of the passage is made large enough to provide an adequate amount of air to the engine. Because the passage essentially is a narrow slot, it is claimed to be filled with acoustic boundary layers that attenuate noise significantly over a relatively wide frequency range.—EEU

6,007,307

43.50.Gf AIR PUMP WITH NOISE REDUCTION PARTITIONS IN THE HOUSING

Yoshiteru Sonoda, assignor to Fujikin, Incorporated
28 December 1999 (Class 417/312); filed in Japan 26 November 1996

This patent relates to air pumps of the type which provide fresh air to tropical fish tanks. An expansion-type silencing chamber (muffler) is integral to the pump housing and large reductions in sound levels are demonstrated.—KPS

5,974,872

43.50.Lj PITCH ARRANGEMENT EVALUATION SYSTEM OF TREAD PATTERN OF PNEUMATIC TIRE

Hisaya Morishita *et al.*, assignors to The Yokohama Rubber Company, Limited
2 November 1999 (Class 73/146); filed in Japan 17 October 1995

A method to evaluate noise characteristics of the tread patterns on pneumatic tires is presented. The acoustic signature is measured and a narrow-band spectrum computed. The least-squares method is then used to fit a quartic regression equation to the amplitude spectrum. The magnitude of the resulting correlation coefficient is used as an indicator of ''low pattern noise'' and is said to correlate highly with drivers' evaluations. The essence of the method seems to indicate that a smooth spectrum is preferable to one with narrow spectral peaks, and the quartic regression coefficient is simply a measure of deviation from ''smooth.''—KPS

5,974,882

43.50.Lj IDENTIFICATION OF RESONANT FREQUENCIES OF VIBRATION OF ROTATING BLADES

Steven Heath, assignor to Rolls-Royce public limited company
2 November 1999 (Class 73/579); filed in the United Kingdom 29 August 1996

A method to determine resonant frequencies of rotating blades, such as those found in turbomachinery, is described in which measurements of blade positions are made at two circumferential locations. It is shown that by rotating the blades at several frequencies, the blade resonant frequencies can be deduced. This method is contrasted with one in which a multitude of sensors can be used to determine the same quantity.—KPS

5,975,609

43.50.Lj SOUND INSULATING LAYER WITH INTEGRAL BOOT

Michael T. Campbell, assignor to Cascade Engineering, Incorporated
2 November 1999 (Class 296/39.3); filed 25 July 1997

An acoustical barrier for mounting to the firewall of an automobile contains an integral boot to provide an acoustical seal around the steering column. The boot is designed to be acoustically effective and to also allow easy installation of the steering column.—KPS

5,978,728

43.50.Lj DETONATION DETECTION METHOD FOR INTERNAL COMBUSTION ENGINES

Roberto Fontana *et al.*, assignors to C. R. F. Societa Consortile per Azioni
2 November 1999 (Class 701/111); filed in Italy 1 December 1995

A method is described to detect detonation (pinging) in an internal combustion engine consisting of a vibration transducer attached to the engine block and a signal processing technique which relies on calculation of standard deviations of the filtered, rectified, and integrated vibration signal. A detonation coefficient is thus calculated which describes the likelihood or propensity for said engine to exhibit detonation.—KPS

5,988,982

43.50.Lj ALTERING VIBRATION FREQUENCIES OF WORKPIECES, SUCH AS GAS TURBINE ENGINE BLADES

Allan H. Clauer, assignor to LSP Technologies, Incorporated
23 November 1999 (Class 416/241R); filed 9 September 1997

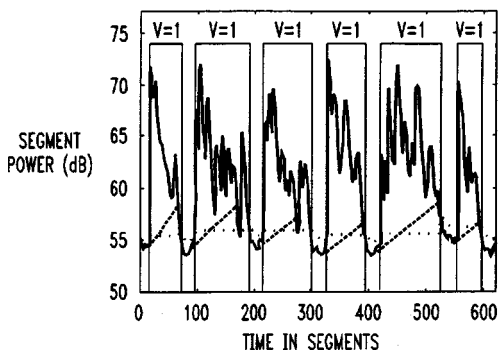
A method of modifying the vibratory resonance characteristics of gas turbine blades is described. Based on the mode shape, locations on the blade having maximum curvature are identified, and are then subjected to shock laser peening. This creates residual stresses in the material and results in a shift in the resonance frequency.—KPS

5,991,718

43.72.Ar SYSTEM AND METHOD FOR NOISE THRESHOLD ADAPTATION FOR VOICE ACTIVITY DETECTION IN NONSTATIONARY NOISE ENVIRONMENTS

David Malah, assignor to AT&T Corporation
23 November 1999 (Class 704/233); filed 27 February 1998

The patented method of speech presence detection improves the reliability of detection speech in noise by determining the most advantageous instants at which the background noise level is estimated. During a speech signal, two energy levels are tracked, the usual upper envelope and a new



''lower envelope'' level. This allows the system to determine more quickly when the speech stops and true background noise is present in the input.—DLR

5,991,724

43.72.Ew APPARATUS AND METHOD FOR CHANGING REPRODUCTION SPEED OF SPEECH SOUND AND RECORDING MEDIUM

Hideki Kojima and Shinta Kimura, assignors to Fujitsu Limited
23 November 1999 (Class 704/266); filed in Japan 19 March 1997

This speech rate changing device is designed to replay a speech recording at a maximum rate, perhaps even beyond the point where words are comprehensible. As the high-speed audio signal is received, an analysis is performed to locate words spoken with emphatic loudness or an increased pitch level. Such words are played back at a lower, perceptible rate, after which the playback again becomes a blur.—DLR

5,991,611

43.72.Gy VOLUME CONTROL DEVICE FOR RECEIVER OF BLOCK CODED SPEECH SIGNALS

Nicolas Palumbo, assignor to Alcatel Mobile Phones
23 November 1999 (Class 455/219); filed in France 16 August 1995

This patent refers to a type of speech transmission called ''block coding.'' Here, multiple blocks (or frames) are sent as a transmission cluster. The individual frames within the cluster may be marked as silence or empty. Prior methods of controlling the playback gain during silence portions have the result that volume is inappropriately reduced during the playback of desired background sounds. This patent alters the volume control logic to remedy that situation.—DLR

5,991,642

43.72.Gy MOBILE COMMUNICATION SYSTEM HAVING A CONTROL STATION WHICH SELECTS SPEECH CODING SCHEMES FOR A MOBILE STATION

Yasuyuki Watanabe *et al.*, assignors to NTT Mobile Communications Network, Incorporated
23 November 1999 (Class 455/560); filed in Japan 10 January 1995

Cell phone systems often have available a variety of speech coding schemes, each optimized for some combination of transmitter, receiver, and local cell noise conditions. This patent describes a method by which the ideal speech coding standard can be selected and smoothly switched during a call. The choice of coding scheme is made by the central control station and passed to the mobile unit when a switch is required.—DLR

5,991,716

43.72.Gy TRANSCODER WITH PREVENTION OF TANDEM CODING OF SPEECH

Matti Lehtimäki, assignor to Nokia Telecommunication OY
23 November 1999 (Class 704/212); filed in Finland 13 April 1995

In a system for telephone communication to a mobile station, the speech coding used for the mobile link portion of the path is typically a very low bitrate method. The base station typically decodes the signal and routes the acoustical signal on to another network. In the case where the receiver is another mobile unit, this results in a second serial (tandem) speech coding, which results in unacceptable speech quality degradation. This patent provides a means for the base station to detect such a double encoding path and to send the original encoded parameters on to the second mobile station.—DLR

5,991,717

43.72.Gy ANALYSIS-BY-SYNTHESIS LINEAR PREDICTIVE SPEECH CODER WITH RESTRICTED-POSITION MULTIPULSE AND TRANSFORMED BINARY PULSE EXCITATION

Tor Björn Minde and Peter Alexander Mustel, assignors to Telefonaktiebolaget LM Ericsson
23 November 1999 (Class 704/223); filed in Sweden 22 March 1995

Many varieties exist of code-exited linear prediction (CELP) vocoding. The method used to code the excitation typically consumes large amounts of channel bandwidth or computing resources or both. The vocoder described here uses a combination of multipulse coding and transformed binary pulse coding and has the ability to select a trade-off between the two techniques. The combined excitation generator replaces the fixed codebook of a typical fixed/adaptive vocoder.—DLR

5,991,723

43.72.Gy METHOD AND APPARATUS FOR TRANSLATING TEXT AND SPEECH TRANSFERRED OVER A TELEPHONY OR SIMILAR NETWORK

Paul Andrew Duffin, assignor to International Business Machines Corporation
23 November 1999 (Class 704/260); filed in European Patent Office 17 January 1998

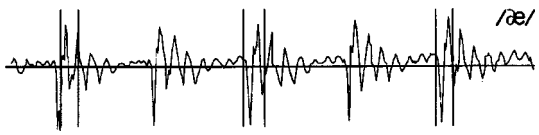
This patent describes an application of typical speech synthesis and recognition equipment to convert the text interface as used by a deaf telephone caller into a standard audio interface for a hearing user. There is no discussion of the detection or handling of recognition errors nor of the synthesis quality.—DLR

5,991,725

43.72.Gy SYSTEM AND METHOD FOR ENHANCED SPEECH QUALITY IN VOICE STORAGE AND RETRIEVAL SYSTEMS

Saf Asghar and Mark Ireton, assignors to Advanced Micro Devices, Incorporated
23 November 1999 (Class 704/270); filed 7 March 1995

Speech storage and playback systems have an advantage over transmission systems (i.e., the typical vocoder) in that all of the data for an entire phrase (or more) is available to the algorithm at the time of playback. This



patent presents a method of using phrase-length parameter behavior to perform a more accurate and bit-conservative smoothing than would be possible for frame-by-frame (real-time) playback.—DLR

6,006,179

43.72.Gy AUDIO CODEC USING ADAPTIVE SPARSE VECTOR QUANTIZATION WITH SUBBAND VECTOR CLASSIFICATION

Shuwu Wu and John Mantegna, assignors to America Online, Incorporated
21 December 1999 (Class 704/222); filed 28 October 1997

This patent describes an audio codec suitable for low-bitrate speech or music encoding. It uses a critically sampled subband filter bank with at least 64 bands and preferably as many as 256 bands. A spectral perceptual masking step zeroes the signal in masked regions. An adaptive sparse vector quantizer performs six different quantization methods, choosing the most favorable for bit reduction. Finally, a soft clipping arrangement rounds the clipped peaks, reducing distortion normally due to hard clipping.—DLR

5,991,712

43.72.Ne METHOD, APPARATUS, AND PRODUCT FOR AUTOMATIC GENERATION OF LEXICAL FEATURES FOR SPEECH RECOGNITION SYSTEMS

Paul A. Martin, assignor to Sun Microsystems, Incorporated
23 November 1999 (Class 704/9); filed 5 December 1996

This patent describes a method for improving the accuracy of a speech recognition system by taking into account certain properties of the candidate words. During training, word co-occurrences are observed, such as adjective/noun combinations. A scheme of optionally considering or ignoring the observed restrictions allows novel combinations to be processed.—DLR

5,991,719

43.72.Ne SEMANTIC RECOGNITION SYSTEM

Masatomo Yazaki *et al.*, assignors to Fujitsu Limited
23 November 1999 (Class 704/251); filed in Japan 27 April 1998

A speech recognition system used for an application such as controlling a process typically combines a word recognizer with a semantic component which acts on particular recognized words to perform specific actions. This patented system provides a language-independent semantic action component by encoding recognized words into standardized number codes, which are passed on to the semantic component. A word-to-number translation table is provided for a variety of input languages.—DLR

5,991,720

43.72.Ne SPEECH RECOGNITION SYSTEM EMPLOYING MULTIPLE GRAMMAR NETWORKS

Michael Galler and Jean-Claude Junqua, assignors to Matsushita Electric Industrial Company, Limited
23 November 1999 (Class 704/256); filed 6 May 1996

This speech recognition system is designed to make use of multiple grammar networks, each optimized for a possible input speech pattern. The system as described here is applied to a telephone dialing task in which the user may either speak or spell a name to be dialed. Recognition results in the form of *N*-best lists from the various grammars are combined to produce a single list of competing matches. These are used to construct a new recognition space in which the final best choice is determined by Viterbi decoding.—DLR

5,991,721

43.72.Ne APPARATUS AND METHOD FOR PROCESSING NATURAL LANGUAGE AND APPARATUS AND METHOD FOR SPEECH RECOGNITION

Yasuharu Asano *et al.*, assignors to Sony Corporation
23 November 1999 (Class 704/257); filed in Japan 31 May 1995

This speech recognition system replaces the typical language grammar with a large set of phrases collected from a speaker of the language. An input speech signal is recognized by searching the phrase database for the most similar spoken phrase. Although not mentioned in the patent title, abstract, or claims, the text makes several references to a language translation function. The phrase storage scheme might make more sense in that context.—DLR

5,991,726

43.72.Ne SPEECH RECOGNITION DEVICES

Peter Immarco *et al.*, Boca Raton, Florida
23 November 1999 (Class 704/270); filed 9 May 1997

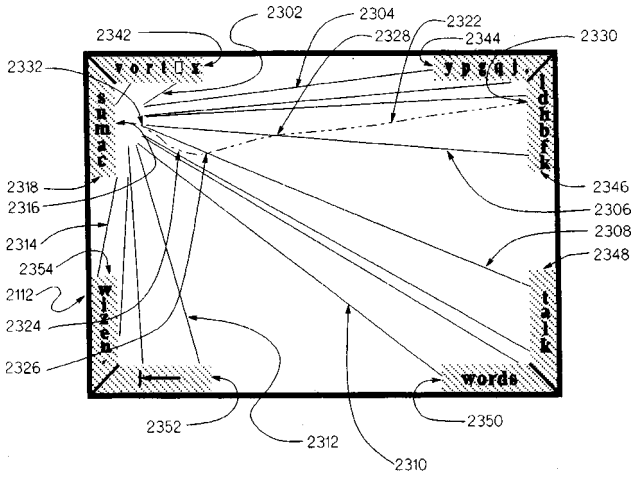
This patent describes a collection of enhancements for use with a voice recognition system in controlling an industrial environment or machine setup. A proximity detector detects operator presence, allowing automatic shutdown and reactivation. A voice detection switch includes a short speech memory store, assuring that an initial utterance is not lost due to switch activation delays. A variety of audio and video responses give the operator a varied environment, enhancing responses and attention span.—DLR

5,999,895

43.72.Ne SOUND OPERATED MENU METHOD AND APPARATUS

Donald K. Forest, Rockville, Maryland
7 December 1999 (Class 704/1); filed 27 March 1995

The patent describes a very general computer interface device, primarily intended for use by persons with any of several physical disabilities, but also said to be applicable to a variety of input tasks. The figure shows one of many possible configurations of the display screen. A sound input capability



is equally flexible, for example, allowing various kinds of sounds, not necessarily phonetically distinct, to be trained as unique item selector keys.—DL

6,003,003

43.72.Ne SPEECH RECOGNITION SYSTEM HAVING A QUANTIZER USING A SINGLE ROBUST CODEBOOK DESIGNED TO MULTIPLE SIGNAL TO NOISE RATIOS

Safdar M. Asghar and Lin Cong, assignors to Advanced Micro Devices, Incorporated
14 December 1999 (Class 704/243); filed 27 June 1997

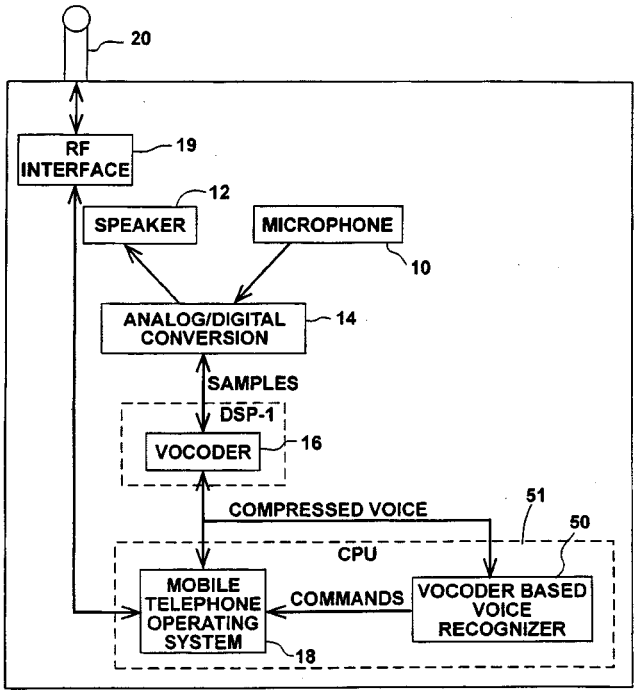
The patent covers a collection of speech recognition techniques, combined in various ways to produce alternate recognizer embodiments. The common element seems to be a set of methods for dealing with varying noise levels in the input signal. The techniques include binary and fuzzy vector and matrix quantizers, with single and multiple codebooks, hidden Markov models, and neural network classifiers.—DLR

6,003,004

43.72.Ne SPEECH RECOGNITION METHOD AND SYSTEM USING COMPRESSED SPEECH DATA

Yehudah Hershkovits and Gabriel Ilan, assignors to Advanced Recognition Technologies, Incorporated
14 December 1999 (Class 704/253); filed 8 January 1998

When a speech recognizer is added to a vocoder, such as in a cell phone, a typical arrangement is to duplicate the signal analysis functions. This patent describes a recognizer designed to operate directly from the



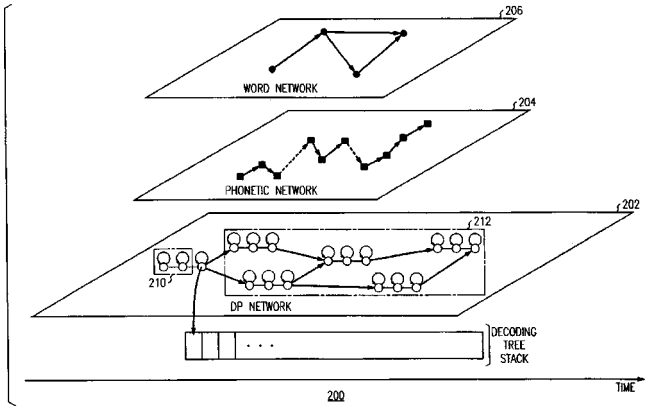
vocoder spectrum and energy data. This allows the recognizer to be implemented in the CPU chip rather than requiring its own separate DSP chip.—DLR

6,006,181

43.72.Ne METHOD AND APPARATUS FOR CONTINUOUS SPEECH RECOGNITION USING A LAYERED, SELF-ADJUSTING DECODER NETWORK

Eric Rolse Buhrke and Wu Chou, assignors to Lucent Technologies, Incorporated
21 December 1999 (Class 704/231); filed 12 September 1997

The speech recognition system described here uses well-known dynamic programming methods for matching the sequence of feature vectors at an acoustic analysis stage. A second stage uses hidden Markov models to match the phonetic segment sequence. A third stage implements the grammar constraints to recover the most probable word sequence. Most of the techniques used are known prior art, described as a series of layers, as might be appropriate for a model of speech recognition in the brain.—DLR

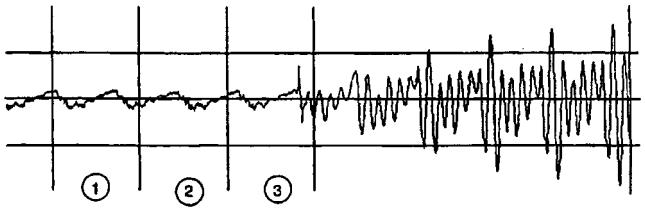


6,006,185

43.72.Ne SYSTEM AND DEVICE FOR ADVANCED VOICE RECOGNITION WORD SPOTTING

Peter Immarco, Boca Raton, Florida
21 December 1999 (Class 704/251); filed 9 May 1997

This speech recognition system emphasizes the process of phonetic segmentation as a major portion of the recognizer. Three sequential windows are located, each of length preferably shorter than a pitch period. A computation which appears to be an absolute magnitude difference function



(AMDF) is used to determine which window best represents a segment boundary. Waveform samples thus segmented are passed to a typical, commercial, isolated-word recognizer for further processing.—DLR

A three-dimensional, two-way, parabolic equation model for acoustic backscattering in a cylindrical coordinate system

Dong Zhu[a] and Leif Bjørnø
Department of Industrial Acoustics, Technical University of Denmark, Building 425, DK-2800 Lyngby, Denmark

(Received 5 January 1999; accepted for publication 20 March 2000)

A new PE model for solving three-dimensional, forward and backward sound propagation in a cylindrical coordinate system is presented. The model marches a wave field in the radial direction including the azimuthal diffraction effects, and solves for a backscattered field based on a three-dimensional, single scattering approach. A periodic sidewall boundary condition is applied for computations in a 360-degree sector, while an approximate sidewall boundary condition is used for calculation in a sector less than 360 degrees. These two sidewall boundary conditions are verified by the numerical results. The major drawback of using the cylindrical coordinate system, when the backscattering solution is valid within a limited area, is analyzed using a geometrical-optical interpretation. The model may be useful for studying three-dimensional backscattering phenomena comprising azimuthal diffraction effects. © *2000 Acoustical Society of America.*
[S0001-4966(00)00607-X]

PACS numbers: 43.20.Bi, 43.30.Bp [DLB]

I. INTRODUCTION

Acoustic backscattering is an important issue in underwater acoustics, particularly in active sonar applications. However, it is normally difficult to model the backscattering in the ocean accurately, because fully three-dimensional effects and two-way propagation are involved. In recent years, the parabolic equation (PE) method[1] has been widely used to solve various sound propagation problems including acoustic backscattering. Backscattering from scatterers of arbitrary shape in three dimensions in a Cartesian coordinate system has been recently solved using the PE technique combining a direct computation of the scattered field based on nonhomogeneous boundary conditions on the scattering objects.[2] In this method the size of objects can range from a few wavelengths to hundreds of wavelengths. Another earlier method for solving three-dimensional backscattering in a Cartesian coordinate system was developed on the basis of the PE technique and the cross-range wave number spectrum.[3] The cross-range transform reduces the three-dimensional problem to a two-dimensional problem that can be solved using the existing two-dimensional PE models. However, the spectral PE solution breaks down for energy propagating nearly parallel to the cross-range direction. For three-dimensional scattering problems in a cylindrical coordinate system, a few PE models were developed by solving a PE using the finite difference method[4] and using the alternating direction method.[5] A coupled mode PE method including mode coupling terms is presented for large-scale scattering problems involving coupling of energy between modes and azimuths.[6] An analytical solution for the three-dimensional acoustic scattering from a nonpenetrable cylindrical island in shallow water is derived using a normal-mode, double-series expansion.[7] The analytical solution is numerically efficient for relatively low frequencies. However, the models above for scattering problems in a cylindrical coordinate system were one-way models (i.e., for forward propagation only). In active sonar simulations where signals are transmitted and received at the same point, it is better to use a cylindrical coordinate system rather than a Cartesian coordinate system. Although a hybrid three-dimensional, two-way PE model[8] may be applicable to model three-dimensional backscattering in a cylindrical coordinate system in some cases, it cannot handle fully three-dimensional effects. Therefore, a PE model that can solve fully three-dimensional backscattering in the cylindrical coordinate system is required, and is presented in this article.

The three-dimensional, two-way PE model is mainly based on a radius-marching algorithm and a three-dimensional, single scattering approach. The radius-marching algorithm originates from a three-dimensional, one-way PE model[4] that achieves a wide-angle capability using a quadratic operator approximation.[4] The three-dimensional, single scattering approach arises from the idea introduced in a two-dimensional, two-way PE model.[9] The three-dimensional computing domain is discretized as a sequence of stepwise radius-independent regions. The radius-marching algorithm marches a wave field forward and backward in the radial direction within each region. Based on the continuity conditions of acoustic pressure and normal particle velocity across the vertical interface between regions, a specific boundary equation is established to obtain the backscattered field. For simplicity of the solution technique, a linear operator approximation is presently used to form the specific boundary equation. In future works other wide-angle approximations may be implemented to improve the model.

The outgoing wave equation is essentially based on a paraxial approximation, a far-field approximation and an as-

[a]Now at Ødegaard & Danneskiold-Samsøe A/S, Titangade 15, DK-2200 Copenhagen N, Denmark.

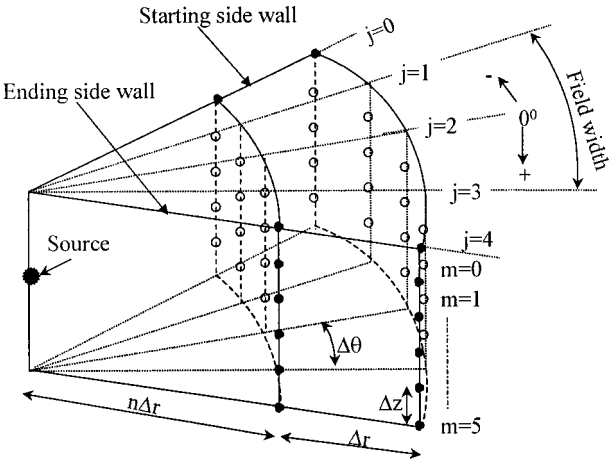


FIG. 1. The three-dimensional grid mesh for $J=3$ and $M=4$.

sumption that backscattering is negligible.[1] The paraxial approximation limits sound propagation in the directions having small angles with respect to the horizontal. However, this narrow-angle limitation has been relaxed nowadays using a high-order Padé approximation.[9] The far-field assumption (i.e., $k_0 r \gg 0$) causes the solution to be valid only beyond a certain distance from the source. The third assumption is not needed in the two-way PE where backscattering has to be included. As the solution of outgoing and incoming waves can be expressed as combinations of Hankel functions,[12] the solution breaks down at the source position because of the singularity of the Hankel functions. Should the solution be regular at the source position, the incident field had to be modified by the presence of the scatterer so that a part of the outgoing field combined with the incoming field could produce a regular result. This is, in fact, a multiple scattering effect that is beyond the scope of this article. Nevertheless, the use of the far-field assumption has implied that the valid solution domain is defined in such a way that a small neighborhood of the source is excluded. Consequently, the solution is not singular within the valid solution domain.

Different from the two-dimensional modeling, the three-dimensional computations in the cylindrical coordinate system involve two additional sidewall boundaries, i.e., a starting sidewall and an ending sidewall. The starting sidewall refers to the radius–depth plane at the starting azimuth angle of the computing domain. The ending sidewall means the radius–depth plane at the ending azimuth angle of the computing domain. If computations are performed in a 360-degree sector (i.e., a cylinder space), the physically correct sidewall boundary conditions are periodic because the starting sidewall overlaps the ending sidewall, while if a scattering object locates within a very narrow sector, the computing domain may be truncated into a sector less than 360 degrees. As a result, the computational time can be reduced greatly. However, this type of computation requires absorbing sidewall boundary conditions that are very difficult to implement. Therefore, approximate sidewall boundary conditions are presently used. The absorbing sidewall boundary conditions may be implemented in further works.

There is a major drawback of using the cylindrical coordinate system to model three-dimensional backscattering, which will be discussed later in this article. Because the radius-marching algorithm always marches the wave field back to the source, the backscattered energy propagating in other directions may not be fully handled. The backscattering solution is thus valid within a limited area. This drawback is analyzed and explained using a geometric-optical interpretation.

It is clear that the single scattering approach is valid for problems involving a single interface,[9] but it is not clear how well this approach works for problems involving multiple interfaces. Therefore, the present computations are focused on the backscattering from those objects that have a single cylindrical vertical surface. Some interesting three-dimensional backscattering phenomena are also discussed.

II. THE THREE-DIMENSIONAL, RADIUS-MARCHING ALGORITHM

The derivation is performed in a cylindrical coordinate system (r,θ,z) and begins with the outgoing wave equation $(+)$ and the incoming wave equation $(-)$:

$$\frac{\partial}{\partial r}u^+ = ik_0(Q-1)u^+, \tag{1a}$$

$$\frac{\partial}{\partial r}u^- = -ik_0(Q-1)u^-, \tag{1b}$$

where k_0 is a reference wave number. The wave fields u^+ and u^- are related to the acoustic pressures p^+ and p^- via the Hankel functions, $\mathrm{H}_0^{(1)}$ and $\mathrm{H}_0^{(2)}$, which are given in the forms of their far-field asymptotic expressions $(k_0 r \gg 1)$:

$$p^+(r,\theta,z) = u^+(r,\theta,z)\mathrm{H}_0^{(1)}(k_0 r) \approx u^+(r,\theta,z)\sqrt{\frac{2}{\pi k_0 r}}\,e^{ik_0 r - i\pi/4}, \tag{2a}$$

$$p^-(r,\theta,z) = u^-(r,\theta,z)\mathrm{H}_0^{(2)}(k_0 r) \approx u^-(r,\theta,z)\sqrt{\frac{2}{\pi k_0 r}}\,e^{-ik_0 r + i\pi/4}. \tag{2b}$$

Note that the solutions of Eq. (2a) and Eq. (2b) are not singular in the far field (i.e., $r \gg 1/k_0$). The three-dimensional operator $Q \equiv \sqrt{1+X+Y}$ is defined by

$$X = \xi^2(r,\theta,z) - 1 + \frac{1}{k_0^2}\rho\frac{\partial}{\partial z}\left(\frac{1}{\rho}\frac{\partial}{\partial z}\right),$$
$$Y = \frac{1}{k_0^2 r^2}\rho\frac{\partial}{\partial \theta}\left(\frac{1}{\rho}\frac{\partial}{\partial \theta}\right), \tag{3}$$

where $\xi = c_0/c(r,\theta,z)$ represents the index of refraction, c_0 is a reference sound speed, and $\rho = \rho(\theta,z)$ denotes the density of the medium. Since Y is proportional to $1/r^2$, the azimuthal coupling in relation to Y reduces quickly when the wave field is marched out in radius. At very large radii, the azimuthal coupling becomes unimportant.

In a cylindrical coordinate system, a three-dimensional environment is discretized in the z direction, r direction, and θ direction. A general three-dimensional grid mesh is shown in Fig. 1, where the computing domain is a sector. The steps

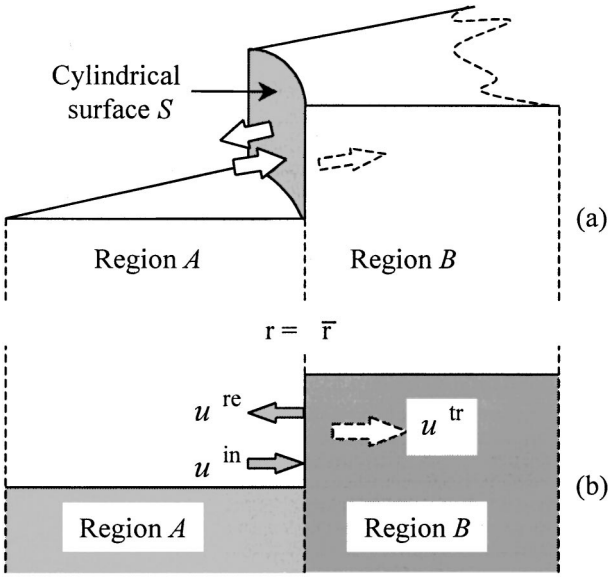


FIG. 2. A three-dimensional view (a) and a side view (b) of the geometry of a step object.

in range, in azimuth, and in depth are denoted by Δr, $\Delta\theta$, and Δz, respectively. The range index is denoted by n. The sector boundaries are identified by $j=1-J$ ($J=3$ in the example in Fig. 1). In particular, the starting sidewall boundary is denoted by $j=0$, and the ending sidewall boundary is denoted by $j=J+1$. The depth index is denoted by $m=1-M$ ($M=4$ in Fig. 1). The sea surface is indicated by $m=0$, and the horizontal interface at the maximum depth is indicated by $m=M+1$. Thus, u_{mj}^{n} denotes the value of a point at the position of $r=n\Delta r$, $\theta=j\Delta\theta$, and $z=m\Delta z$.

A range-dependent environment, where the water–seabed interface may vary in the radial direction as well as in the azimuthal direction, is discretized forming a sequence of stepwise, range-independent regions shown in Fig. 2. Two regions A and B are separated by a cylindrical interface (S) at $r=\bar{r}$. Sound is assumed to propagate from A to B across S. The incident field, the transmitted field, and the reflected field are denoted by u^{in}, u^{tr}, and u^{re}, respectively. Note that the azimuthal discretization may cause discrepancies between the real (original) environment and the discrete environment. Analyses of the errors arising from the azimuthal discretization can be found in Ref. 8.

The radius-marching algorithm is based on the quadratic approximation[4,10]

$$Q\equiv\sqrt{1+X+Y}\cong 1+\tfrac{1}{2}X-\tfrac{1}{8}X^2+\tfrac{1}{2}Y. \tag{4}$$

The radius-marching algorithm can be expressed as[3]

$$\left[1+\frac{1-ik_0\Delta r}{4}X\right]\left[1-\frac{ik_0\Delta r}{4}Y\right]u^{n+1} =\left[1+\frac{1+ik_0\Delta r}{4}X\right]\left[1+\frac{ik_0\Delta r}{4}Y\right]u^{n}, \tag{5}$$

where u^n denotes the known wave field at the present range (range n), and u^{n+1} denotes the unknown wave field at the subsequent range (range $n+1$). Since in Eq. (5) the couplings between X and Y are separated, Eq. (5) can be solved efficiently by solving two tridiagonal system of equations using the lower-upper triangular (LU) decomposition method.[11]

III. THE THREE-DIMENSIONAL, SINGLE SCATTERING APPROACH

As shown in Fig. 2, across a cylindrical interface S at $r=\bar{r}$, the acoustic pressure and the normal components of the particle velocity satisfy the continuity conditions:

$$u^{\mathrm{in}}H_0^{(1)}(k_0\bar{r})+u^{\mathrm{re}}H_0^{(2)}(k_0\bar{r})=u^{\mathrm{tr}}H_0^{(1)}(k_0\bar{r}), \tag{6a}$$

$$\frac{1}{\rho_A}\frac{\partial(u^{\mathrm{in}}H_0^{(1)})}{\partial r}\bigg|_{\bar{r}}+\frac{1}{\rho_A}\frac{\partial(u^{\mathrm{re}}H_0^{(2)})}{\partial r}\bigg|_{\bar{r}}=\frac{1}{\rho_B}\frac{\partial(u^{\mathrm{tr}}H_0^{(1)})}{\partial r}\bigg|_{\bar{r}}. \tag{6b}$$

Replacing the derivatives in Eq. (6b) using Eqs. (1a) and (1b), the combination of Eqs. (6a) and (6b) results in a three-dimensional boundary equation:

$$\left[L_A+L_B-\frac{i}{2k_0\bar{r}}\left(\frac{1}{\rho_A}-\frac{1}{\rho_B}\right)\right]u^{\mathrm{tr}}=2L_Au^{\mathrm{in}}, \tag{7}$$

where

$$L_A=\frac{\sqrt{1+X_A+Y_A}}{\rho_A}\approx\frac{1+0.5X_A+0.5Y_A}{\rho_A},$$
$$L_B=\frac{\sqrt{1+X_B+Y_B}}{\rho_B}\approx\frac{1+0.5X_B+0.5Y_B}{\rho_B}. \tag{8}$$

A low-order expansion is used here to approximate the square-root operator in Eq. (8). Although this is a narrow-angle approximation for propagation in azimuth-depth plane, it gives the simplest finite difference scheme to the operator L. Other approximations such as a rational-function approximation may be used; however, discretizing Eq. (7) will be much more difficult and complicated because operators are involved in both the numerator and denominator. Once u^{tr} is obtained, the reflected field u^{re} can be readily found:

$$u^{\mathrm{re}}=-i\cdot e^{i2k_0\bar{r}}(u^{\mathrm{tr}}-u^{\mathrm{in}}). \tag{9}$$

The continuity conditions in the radial direction are fulfilled by satisfying Eq. (7), while the continuity conditions in the azimuthal direction are fulfilled automatically during the radius-marching process that includes the azimuthal coupling and azimuthal dependence.

IV. THE SIDEWALL BOUNDARY CONDITIONS

Solving Eq. (7) forms the crucial part in the development of the present model. The solution technique greatly depends on how the operator L is approximated. Since Eq. (8) only involves X and Y in the numerator, the operator L can be discretized using the finite difference scheme and Varga's procedure[10] that satisfies the azimuthal continuity. Using the notations shown in Fig. 1, the operation Lu can be expressed as

$$Lu_{m,j}\equiv B_{m,j}u_{m,j}+D_{m,j}u_{m+1,j}+E_{m,j}u_{m-1,j}+F_{m,j}u_{m,j+1} +G_{m,j}u_{m,j-1}, \tag{10}$$

where $B_{m,j}$, $D_{m,j}$, $E_{m,j}$, $F_{m,j}$, and $G_{m,j}$ are the constant coefficients given in the Appendix. Unlike Eq. (5), Eq. (7) is solved at one range only. Therefore, the superscript n is omitted in Eq. (10).

To rewrite Eq. (10) in the form of a product of matrices, proper boundary conditions for the starting sidewall and the ending sidewall are needed. Two types of sidewall boundary conditions are implemented corresponding to whether the computing domain is 360 degrees or less than 360 degrees in azimuth. To clearly illustrate how these sidewall boundary conditions are implemented, a simple example given in Fig. 3 is used.

In general the computing domain is discretized as sectors shown in Fig. 3(a). The sector boundaries are counted from the starting sidewall ($j=0$) to the ending sidewall ($j=4$). Imagine to rotate the ending sidewall ($j=4$) clockwise and the starting sidewall ($j=0$) counter-clockwise until one sector is overlapped, the computing domain becomes a circular area shown in Fig. 3(b). The sector boundaries are numbered from $j=0$ to $j=4$ in the direction shown by the arrows. The periodic sidewall boundary conditions applied to Fig. 3(b) assume that $u_{m,0}=u_{m,3}$ and $u_{m,4}=u_{m,1}$ because one sector is overlapped, shown by the dark shadow sector. Using these assumptions, Eq. (10) can be rewritten as

$$\mathbf{L}\cdot\mathbf{U}=\begin{bmatrix} \mathbf{T}_1 & \mathbf{F}_1 & 0 & \cdots & \cdots & 0 & \mathbf{G}_1 \\ \mathbf{G}_2 & \mathbf{T}_2 & \mathbf{F}_2 & \ddots & & & 0 \\ 0 & \mathbf{G}_3 & \mathbf{T}_3 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \mathbf{T}_{J-2} & \mathbf{F}_{J-2} & 0 \\ 0 & & & \ddots & \mathbf{G}_{J-1} & \mathbf{T}_{J-1} & \mathbf{F}_{J-1} \\ \mathbf{F}_J & 0 & \cdots & \cdots & 0 & \mathbf{G}_J & \mathbf{T}_J \end{bmatrix} \cdot \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \vdots \\ \vdots \\ \mathbf{u}_{J-1} \\ \mathbf{u}_J \end{pmatrix} + \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \vdots \\ \vdots \\ \mathbf{v}_{J-1} \\ \mathbf{v}_J \end{pmatrix}, \tag{11a}$$

where the block matrices are defined by

$$\mathbf{T}_i=\begin{bmatrix} B_{1,i} & D_{1,i} & 0 & 0 \\ E_{2,i} & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & D_{M-1,i} \\ 0 & 0 & E_{M,i} & B_{M,i} \end{bmatrix},$$

$$\mathbf{F}_i=\begin{bmatrix} F_{1,i} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & F_{M,i} \end{bmatrix}, \tag{11b}$$

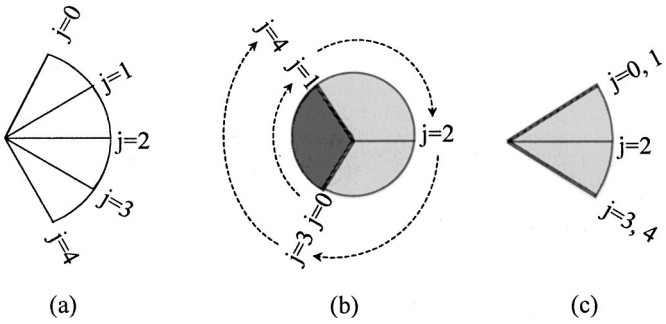


FIG. 3. Top view of sector boundaries. (a) General mesh. (b) Computing domain around 360 degrees. (c) Computing domain less than 360 degrees.

$$\mathbf{G}_i=\begin{bmatrix} G_{1,i} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & G_{M,i} \end{bmatrix}.$$

Each $\mathbf{T}_i$ is a tridiagonal matrix, while $\mathbf{F}_i$ and $\mathbf{G}_i$ are diagonal matrices. The elements $B_{m,j}$, $D_{m,j}$, $E_{m,j}$, $F_{m,j}$, and $G_{m,j}$ are given in the Appendix. The vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ contain the values of the field points at the ith sector boundary. They are given by

$$\mathbf{u}_i=\begin{pmatrix} u_{1,i} \\ u_{2,i} \\ \vdots \\ u_{M-1,i} \\ u_{M,i} \end{pmatrix}, \quad \mathbf{v}_i=\begin{pmatrix} E_{1,i}\cdot u_{0,i} \\ 0 \\ \vdots \\ 0 \\ D_{M,i}\cdot u_{M+1,i} \end{pmatrix}, \tag{11c}$$

where $u_{1,i}$ to $u_{M,i}$ represent the values of the field points in the z direction, $u_{0,i}$ denotes the value on the sea surface, and $u_{M+1,i}$ denotes the value at the maximum depth. The vector $\mathbf{v}_i$ will vanish if a pressure-release boundary condition is assumed on the sea surface and at the maximum depth.

If the computing domain is a sector less than 360 degrees, absorbing boundary conditions should be applied to the sidewall boundaries to prevent artificial reflections. However, it is difficult to implement the absorbing boundary conditions. Presently, approximate sidewall boundary conditions shown in Fig. 3(c) are used, which set the starting sidewall ($j=0$) the same as the sector boundary of $j=1$, and set the ending sidewall ($j=4$) the same as the sector boundary of $j=3$, namely, $u_{m,0}=u_{m,1}$ and $u_{m,4}=u_{m,3}$. Based on the approximate sidewall boundary conditions, a new sparse matrix equation is derived from Eq. (10). Slightly different from Eq. (11a), the new sparse matrix does not involve the block matrices $\mathbf{G}_1$ and $\mathbf{F}_J$ at the corners. In addition, the new sparse matrix has two block matrices $\mathbf{T}_1$ and $\mathbf{T}_J$ defined differently from Eq. (11b). The new $\mathbf{T}_1$ and $\mathbf{T}_J$ are defined by

$$\mathbf{T}_1=\begin{bmatrix} B_{1,1}+G_{1,1} & D_{1,1} & 0 & 0 \\ E_{2,1} & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & D_{M-1,1} \\ 0 & 0 & E_{M,1} & B_{M,1}+G_{M,1} \end{bmatrix}, \tag{12}$$

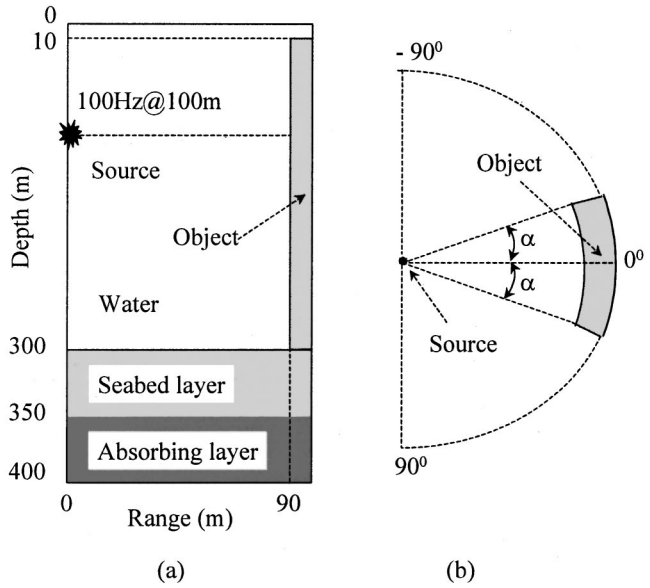


FIG. 4. The geometry of the single sector case. (a) The side view. (b) The top view.

$$\mathbf{T}_J=\begin{bmatrix} B_{1,J}+F_{1,J} & D_{1,J} & 0 & 0 \\ E_{2,J} & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & D_{M-1,J} \\ 0 & 0 & E_{M,J} & B_{M,J}+F_{M,J} \end{bmatrix}.$$

Note that only the main diagonals in **T** in Eq. (12) are different from Eq. (11b).

Using either of the sidewall boundary conditions, Eq. (7) can finally be expressed in the form of a sparse matrix equation. As shown in Eqs. (11a) and (11b), the size of the matrix L is of the order of $M^2\times J^2$, where M is the total number of points in the z direction, and J is the number of sectors in the θ direction. The matrix L may be very huge but has only very few nonzero diagonals. In the case of Fig. 3(b), L has $5MJ-2J$ nonzero elements. In the case of Fig. 3(b), L has only $5MJ-2(M+J)$ nonzero elements. The sparse matrix equation is solved using a *row-index* method,[11] which stores only nonzero elements during the computation.

V. EXAMPLES

An example shown in Fig. 4 is used to verify the two different sidewall boundary conditions discussed above. As shown in the side view of Fig. 4(a), the water is 300 deep with the sound speed $c_w=1500$ m/s, density $\rho_w=1000$ kg/m^3, and attenuation $\alpha_w=0.0$ dB/λ. The 50-m seabed layer is described by $c_b=1700$ m/s, $\rho_b=1500$ kg/m^3, and $\alpha_b=0.5$ dB/λ. The absorbing layer extended from 350 to 400 m has the same parameters as in the seabed layer except $\alpha_b=30.0$ dB/λ. An omnidirectional point source is assumed locating at 100 m below the sea surface and transmitting continuous wave signals with a frequency of 100 Hz. A cylindrical sector is used as the scattering object because it has only one vertical cylindrical interface. To concentrate on the effects of backscattering from the object, the influence from the interaction of sound with the seabed should be reduced greatly. Therefore, the object is placed at the range of 90 m that is much shorter than the water depth to reduce the bottom influence. The object is assumed situating from $-\alpha$ to α in the azimuthal direction shown in Fig. 4(b). In this example α equals 2 degrees. To use the two different sidewall boundary conditions, the computing domain is chosen to be a 360-degree sector in one test case and a 90-degree sector in another test case. The trial runs show that the parameters of $\Delta r=1$ m, $\Delta z=1$ m, and $\Delta\theta=1$ degree may be appropriate to obtain the convergent results.

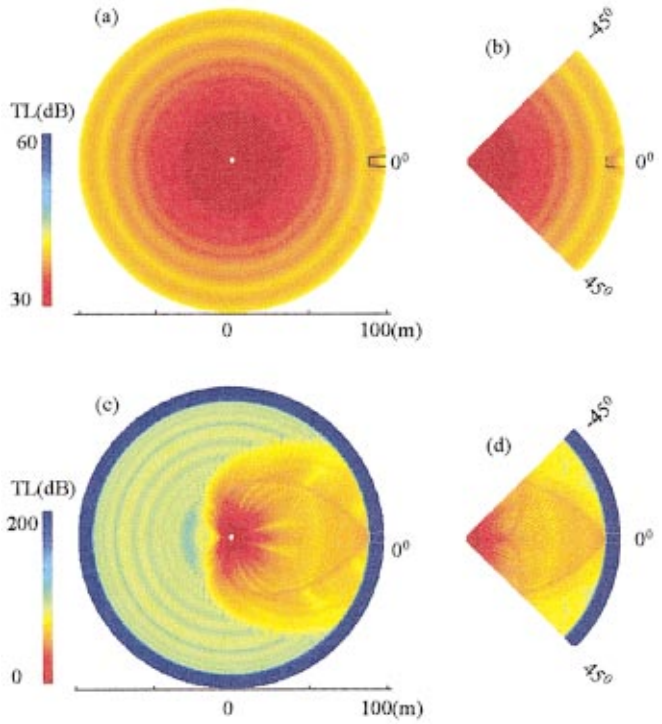


FIG. 5. The top views of the forward transmission losses, (a) and (b), and the backward transmission losses, (c) and (d). The computing domain is 360 degrees in (a) and (c), but 90 degrees in (b) and (d). All results are created at the receiving depth of 100 m, and the same object within an azimuthal angle of 4 degrees is used.

The top views of the forward and backward transmission losses obtained at the receiving depth of 100 m are shown in Fig. 5, where the top plots display the forward transmission losses, and the lower plots are the backward transmission losses. The left plots correspond to the 360-degree computations, while the right plots show the results calculated in a sector area from -45 degrees to 45 degrees. The comparison between Figs. 5(a) and 5(b) and the comparison between Figs. 5(c) and 5(d) both display very good agreement in the scattering patterns. Significant diffraction effects are found in Figs. 5(c) and 5(d), where energy is backscattered from the object in a manner similar to sound radiation from a point source. Quantitative comparisons are given in Figs. 6 and 7. Figure 6 shows the transmission losses versus azimuthal angles at a given range. The plotting depth is 100 m, and the plotting range is 100 m for the forward propagation and 70 m for the backward propagation. The solid curve corresponds to the computation from -180 degrees to 180 degrees, and the dashed curve corresponds to the computation from -45 degrees to 45 degrees. The forward and backward transmission losses created at the azimuthal angle of 0 degrees and at

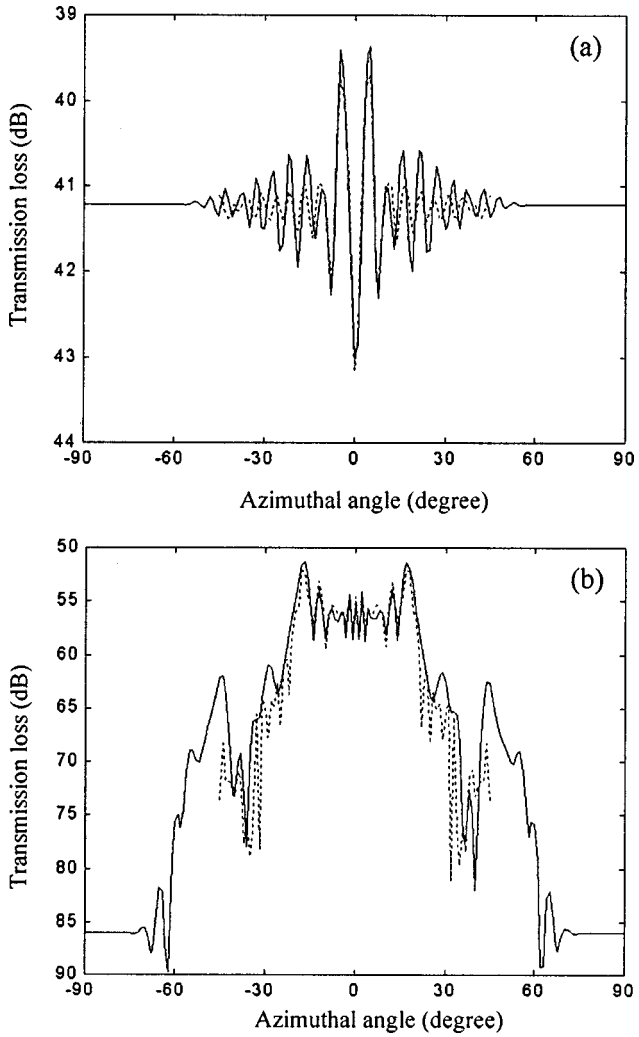


FIG. 6. Comparisons of the transmission losses for (a) the forward propagation plotted at the radius of 100 m, and (b) the backward propagation plotted at the radius of 70 m. The solid and dashed curves correspond to the computation over 360 degrees and the calculation within a 90-degree sector, respectively.

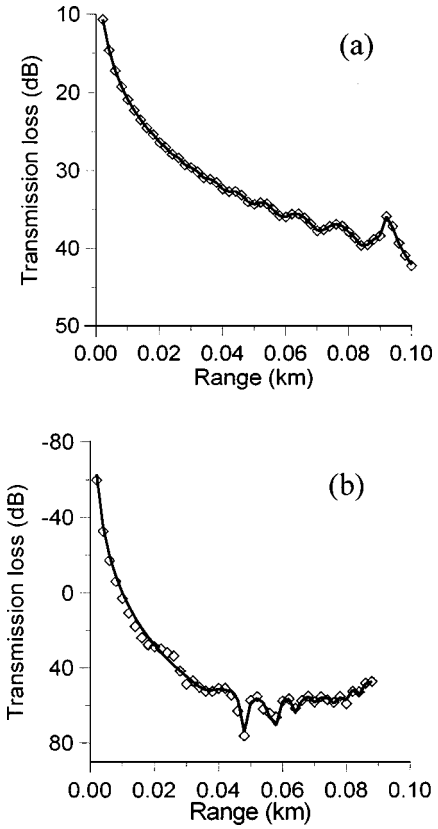


FIG. 7. The forward transmission loss (a) and the backward transmission loss (b) created at azimuthal angle of 0 degrees and at the depth of 100 m. Comparisons between the solid curve for the 360-degree computation and the squares for the 90-degree calculation show an excellent agreement.

the depth of 100 m are shown in Fig. 7, where the solid curve corresponds to the 360-degree computation and the squares are the result of the 90-degree calculation. Excellent agreement is found in all the comparisons in Figs. 6 and 7, indicating that both of the sidewall boundary conditions shown in Fig. 3 can be applied to obtain identical results. However, the computational time due to the use of two different sidewall boundary conditions is very different. In the example above, the two-way run of the 360-degree computation takes about 10 min on a 400-MHz PC, while the two-way run of the 90-degree calculation uses only about 1 min on the same computer. Therefore, it is not always preferable to perform 360-degree calculation using the periodic sidewall boundary condition if computations in a sector area less than 360 degrees can cover the backscattered field that is of interest, and the solution accuracy is acceptable.

The side views of the forward and backward transmission losses created at an azimuthal angle of 0 degrees for the 360-degree and the 90-degree computations are given in Fig. 8. Excellent agreement is observed in the comparison of the forward propagating field between Figs. 8(a) and 8(b), and in the comparison of the backward propagating field between Figs. 8(c) and 8(d). It is also noted that the absorbing layer is efficient enough to reduce the bottom reflection by attenuating significantly the sound energy that penetrates into the bottom. Therefore, the backscattered fields given in Fig. 5 are mostly due to the backscattering from the cylindrical sector object with minimized bottom influences. The results thus are more useful and accurate for studying the characteristics of the backscattering from objects.

Backscattering effects are of course dependent on the source frequency as well. To illustrate the frequency dependence, the source frequency is increased from 100 to 300 Hz in the previous example. The side views of the forward-propagating field and the backward-propagating field created at azimuthal angle of 0 degrees are shown in Figs. 9(a) and 9(b). In Fig. 9(a) the major change from Fig. 8(a) is that the sound energy radiating from the point source is trapped within a beam, which ensonifies the object very well. Figure 9(b) shows significant energy being backscattered from the object, where the lower edge of the backscattered field is along the specular direction of the lower edge of the incident field. Bottom influence can be ignored in this case.

Figure 10 shows the top view of the backward transmission loss obtained at the receiving depth of 100 m. The backscattering pattern shows a beam that is narrower than the one found in Fig. 5(c), meaning that the diffraction effects are weaker in Fig. 10. It is noted that the backscattered energy goes all the way back to the source. However, this is not

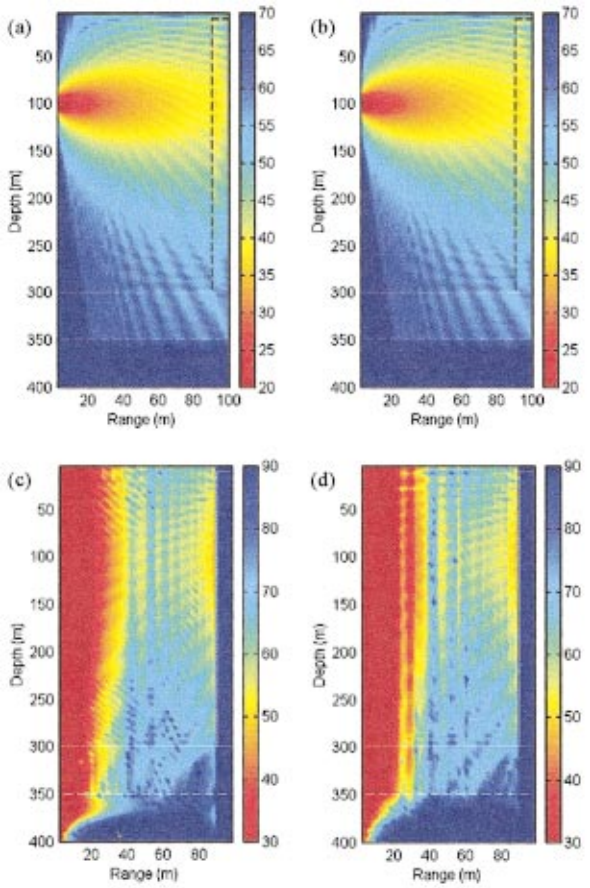


FIG. 8. The side views of the forward transmission losses, (a) and (b), and the backward transmission losses, (c) and (d), created at the azimuthal angle of 0 degrees. The plots to the left correspond to the 360-degree computations, and the plots to the right are the results of the 90-degree calculations. The horizontal white solid lines indicate the water–seabed interface, the horizontal white dashed lines show the absorbing layer interface, and the vertical black and white lines outline the cylindrical interface of the object.

what is exactly observed in practice. Therefore, the solution cannot be taken seriously inside some range around the source. This is also the main drawback of using a cylindrical coordinate system for three-dimensional, two-way propagation modeling. A geometrical-optical interpretation displayed

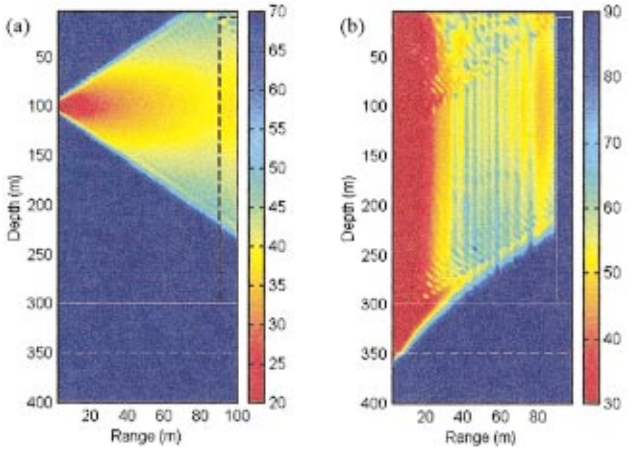


FIG. 9. The side views of the forward transmission loss (a) and the backward transmission loss (b) created at the azimuthal angle of 0 degrees for the frequency of 300 Hz.

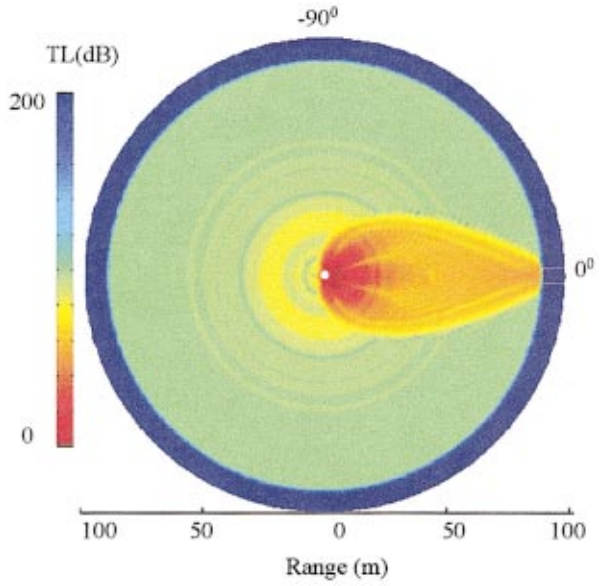


FIG. 10. The top view of the backward transmission loss created at the receiving depth of 100 m. The source frequency is 300 Hz.

in Fig. 11 may be helpful to roughly understand the phenomenon. The scatterer is assumed to be a very narrow object (*S*) behaving like an omnidirectional point source that can radiate sound in all directions. The wave field is marched back to the source (*O*) through the radii indicated by the dotted circles. An incoming propagation is defined as the sound propagation from large radii to small radii. For example, ray *SA* is considered as an incoming propagation, while ray *AB* is considered as an outgoing propagation. Since the backward radius-marching algorithm handles only the incoming propagation, the outgoing propagation is omitted during the backward marching process. Therefore, all the possible diffracting rays (indicated by the arrows) from the scatterer *S* are ''truncated'' with only the incoming propagation parts

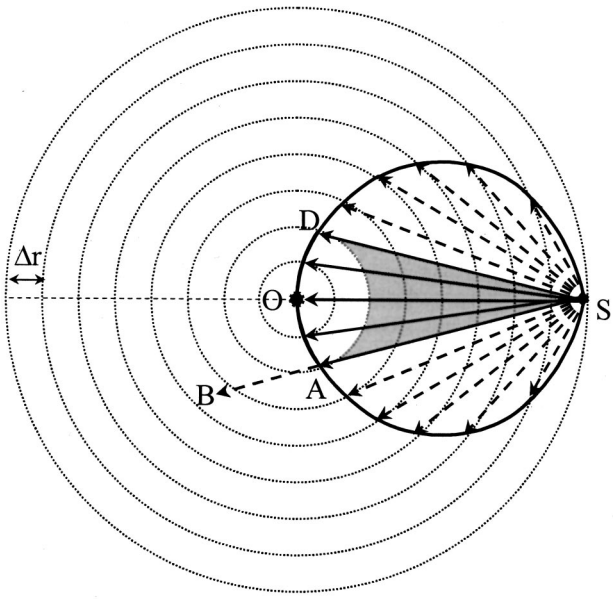


FIG. 11. A geometric-optical interpretation of the limited region of backscattering from a point scatterer.

remaining during the backward marching. Those incoming rays representing all the valid solutions are bounded within a limited region shown by the thick elliptic curves. Furthermore, if the scatterer S is not an omnidirectional point source but has certain directivity, the possible diffracting rays (now indicated by the solid arrows only) from S are limited within a certain angle, which may be due to the PE's angular limitation. The far-field assumption also leads to a bound at a certain radius. The valid backscattering solutions represented by those solid arrows finally are bounded within a small area indicated by the shadow zone in Fig. 11, which is similar to the backscattering pattern in Fig. 10.

VI. DISCUSSIONS AND CONCLUSIONS

A new three-dimensional, two-way PE model has been developed in a cylindrical coordinate system to solve three-dimensional backscattering problems including azimuthal diffraction effects. The correct periodic sidewall boundary conditions are implemented for computations in a sector around 360 degrees. When the computing domain is less than 360 degrees, approximate sidewall boundary conditions are applied to give fairly good results. The accurate absorbing sidewall boundary conditions may be implemented in future works. Although the Hankel function is singular at the source position, the solution does not have a singularity because the use of the far-field assumption removes a small neighborhood of the source from the valid solution domain. The major drawback is that the backscattering solution is valid within a limited area, which is due to the cylindrical marching process as well as the single scattering approach.

Nevertheless, the three-dimensional, two-way PE model can be useful to demonstrate some interesting three-dimensional backscattering phenomena that cannot be seen using two-dimensional models. For example, the results show that the backscattering from a narrow-angle object is in a manner similar to sound radiation from a point source. The azimuthal diffraction effects at higher frequencies are weaker than that at lower frequencies.

ACKNOWLEDGMENTS

The work was partly financed by the EU/MAST projects POSEIDON (MAS3-CT95-0038), LOTUS (MAS3-CT97-0099), and SIGMA (MAS3-CT97-0100).

APPENDIX: DERIVATION OF THE SPARSE MATRIX EQUATION

With the operator L defined in Eq. (8), the operation Lu can be expressed as

$$Lu=\left[\frac{A_0+A_1X+A_2Y}{\rho}\right]u\equiv[L_1+L_2+L_3]u, \tag{A1}$$

where $A_0=1$, $A_1=A_2=0.5$. Inserting Eq. (3) into Eq. (A1) leads to

$$L_1=A_0\frac{1}{\rho}+A_1\frac{\xi^2(r,\theta,z)-1}{\rho},\quad L_2=\frac{A_1}{k_0^2}\frac{\partial}{\partial z}\left(\frac{1}{\rho}\frac{\partial}{\partial z}\right),$$
$$L_3=\frac{A_2}{(k_0\overline{r})^2}\frac{\partial}{\partial\theta}\left(\frac{1}{\rho}\frac{\partial}{\partial\theta}\right). \tag{A2}$$

The varying density comprises the inhomogeneity of the medium. The derivation uses the notations shown in Fig. 1. Using the continuity conditions on the mth horizontal interface ($z_m=z_0+m\Delta z$) and the finite difference scheme, L_1 can be expressed as

$$L_1u_{m,j}=\frac{A_0}{2}\left(\frac{1}{\rho_{m+1/2,j}}+\frac{1}{\rho_{m-1/2,j}}\right)u_{m,j}+\frac{A_1}{2}\left[\frac{\xi^2_{m+1/2,j}-1}{\rho_{m+1/2,j}}+\frac{\xi^2_{m-1/2,j}-1}{\rho_{m-1/2,j}}\right]u_{m,j}. \tag{A3}$$

For expressing L_2, the first-order derivative can easily be handled using a forward finite difference scheme. But the major concern is how to treat $(\partial/\partial z)((1/\rho)\partial/\partial z)$ because $(1/\rho)\partial/\partial z$ is discontinuous on the horizontal interface. Here, Varga's procedure[10] is applied:

$$\frac{\partial}{\partial z}\left(\frac{1}{\rho}\frac{\partial}{\partial z}\right)u_{m,j}=\frac{1}{\Delta z}\left[\left(\frac{1}{\rho}\frac{\partial}{\partial z}\right)_{m+1/2,j}-\left(\frac{1}{\rho}\frac{\partial}{\partial z}\right)_{m-1/2,j}\right]u_{m,j}=\frac{\dfrac{1}{\rho_{m+1/2,j}}(u_{m+1,j}-u^1_{m,j})-\dfrac{1}{\rho_{m-1/2,j}}(u^2_{m,j}-u_{m-1,j})}{(\Delta z)^2}, \tag{A4}$$

where $u^1_{m,j}$ denotes the field on the horizontal interface z_m in medium 1, and $u^2_{m,j}$ denotes the field on the horizontal interface z_m in medium 2. The continuity conditions imply $u^1_{m,j}=u^2_{m,j}=u_{m,j}$. Using the scheme in Eq. (A4), L_2 and L_3 can be expressed as

$$L_2u_{m,j}=\frac{A_1}{(k_0\Delta z)^2}\left[-\left(\frac{1}{\rho_{m+1/2,j}}+\frac{1}{\rho_{m-1/2,j}}\right)u_{m,j}+\frac{1}{\rho_{m+1/2,j}}u_{m+1,j}+\frac{1}{\rho_{m-1/2,j}}u_{m-1,j}\right], \tag{A5}$$

$$L_3u_{m,j}=\frac{A_2}{(k_0\overline{r}\Delta\theta)^2}\left[-\left(\frac{1}{\rho_{m,j+1/2}}+\frac{1}{\rho_{m,j-1/2}}\right)u_{m,j}+\frac{1}{\rho_{m,j+1/2}}u_{m,j+1}+\frac{1}{\rho_{m,j-1/2}}u_{m,j-1}\right]. \tag{A6}$$

Note that the densities in Eqs. (A3) and (A5) are evaluated at the depths of $z_0+(m+\frac{1}{2})\Delta z$ and $z_0+(m-\frac{1}{2})\Delta z$ on both sides of the mth horizontal interface while the densities in Eq. (A6) are evaluated at the azimuthal angles of $(j+\frac{1}{2})\Delta\theta$ and $(j-\frac{1}{2})\Delta\theta$ on both sides of the jth vertical interface. In-

serting Eqs. (A3), (A5), and (A6) into (A1) results in Eq. (10), where the coefficients are given by

$$B_{m,j}=\left(\frac{A_0}{2}-\frac{A_1}{(k_0\Delta z)^2}\right)\left(\frac{1}{\rho_{m+1/2,j}}+\frac{1}{\rho_{m-1/2,j}}\right)+\frac{A_1}{2}\left[\frac{\xi^2_{m+1/2,j}-1}{\rho_{m+1/2,j}}+\frac{\xi^2_{m-1/2,j}-1}{\rho_{m-1/2,j}}\right]-\frac{A_2}{(k_0\bar{r}\bar{\Delta}\theta)^2}\left(\frac{1}{\rho_{m,j+1/2}}+\frac{1}{\rho_{m,j-1/2}}\right), \tag{A7}$$

$$D_{m,j}=\frac{A_1}{(k_0\Delta z)^2}\frac{1}{\rho_{m+1/2,j}},\quad E_{m,j}=\frac{A_1}{(k_0\Delta z)^2}\frac{1}{\rho_{m-1/2,j}}, \tag{A8}$$

$$F_{m,j}=\frac{A_2}{(k_0\bar{r}\bar{\Delta}\theta)^2}\frac{1}{\rho_{m,j+1/2}},\quad G_{m,j}=\frac{A_2}{(k_0\bar{r}\bar{\Delta}\theta)^2}\frac{1}{\rho_{m,j-1/2}}. \tag{A9}$$

For transforming Eq. (10) into a matrix form, a simple example shown in Fig. 3 is used, where $m=0-5$ and $j=0-4$. Following the order of j and m, Eq. (10) can be expanded as

$$Lu_{m,j}=\begin{cases}B_{1,1}\cdot u_{1,1}+D_{1,1}u_{2,1}+E_{1,1}u_{0,1}+F_{1,1}u_{1,2}+G_{1,1}u_{1,0}\\ B_{2,1}u_{2,1}+D_{2,1}u_{3,1}+E_{2,1}u_{1,1}+F_{2,1}u_{2,2}+G_{2,1}u_{2,0}\\ B_{3,1}u_{3,1}+D_{3,1}u_{4,1}+E_{3,1}u_{2,1}+F_{3,1}u_{3,2}+G_{3,1}u_{3,0}\\ B_{4,1}u_{4,1}+D_{4,1}u_{5,1}+E_{4,1}u_{3,1}+F_{4,1}u_{4,2}+G_{4,1}u_{4,0}\\ B_{1,2}\cdot u_{1,2}+D_{1,2}u_{2,2}+E_{1,2}u_{0,2}+F_{1,2}u_{1,3}+G_{1,2}u_{1,1}\\ B_{2,2}u_{2,2}+D_{2,2}u_{3,2}+E_{2,2}u_{1,2}+F_{2,2}u_{2,3}+G_{2,2}u_{2,1}\\ B_{3,2}u_{3,2}+D_{3,2}u_{4,2}+E_{3,2}u_{2,2}+F_{3,2}u_{3,3}+G_{3,2}u_{3,1}\\ B_{4,2}u_{4,2}+D_{4,2}u_{5,2}+E_{4,2}u_{3,2}+F_{4,2}u_{4,3}+G_{4,2}u_{4,1}\\ B_{1,3}\cdot u_{1,3}+D_{1,3}u_{2,3}+E_{1,3}u_{0,3}+F_{1,3}u_{1,4}+G_{1,3}u_{1,2}\\ B_{2,3}u_{2,3}+D_{2,3}u_{3,3}+E_{2,3}u_{1,3}+F_{2,3}u_{2,4}+G_{2,3}u_{2,2}\\ B_{3,3}u_{3,3}+D_{3,3}u_{4,3}+E_{3,3}u_{2,3}+F_{3,3}u_{3,4}+G_{3,3}u_{3,2}\\ B_{4,3}u_{4,3}+D_{4,3}u_{5,3}+E_{4,3}u_{3,3}+F_{4,3}u_{4,4}+G_{4,3}u_{4,2}.\end{cases} \tag{A10}$$

Equation (A10) may be written in the form of a product of matrices if u is written as a vector following the order of j. However, of special concern is how to treat the starting sidewall boundary ($u_{m,0}$) and the ending sidewall boundary ($u_{m,4}$). Two types of sidewall boundary conditions are shown in Fig. 3.

In the case of Fig. 3(b), the periodic sidewall boundary conditions assume $u_{m,0}=u_{m,3}$ and $u_{m,4}=u_{m,1}$. Equation (A10) can be reorganized and expressed in the form of a sparse matrix:

$$Lu=\left[\begin{array}{cccc:cccc:cccc}
B_{1,1} & D_{1,1} & 0 & 0 & F_{1,1} & 0 & 0 & 0 & G_{1,1} & 0 & 0 & 0\\
E_{2,1} & B_{2,1} & D_{2,1} & 0 & 0 & F_{2,1} & 0 & 0 & 0 & G_{2,1} & 0 & 0\\
0 & E_{3,1} & B_{3,1} & D_{3,1} & 0 & 0 & F_{3,1} & 0 & 0 & 0 & G_{3,1} & 0\\
0 & 0 & E_{4,1} & B_{4,1} & 0 & 0 & 0 & F_{4,1} & 0 & 0 & 0 & G_{4,1}\\
\hdashline
G_{1,2} & 0 & 0 & 0 & B_{1,2} & D_{1,2} & 0 & 0 & F_{1,2} & 0 & 0 & 0\\
0 & G_{2,2} & 0 & 0 & E_{2,2} & B_{2,2} & D_{2,2} & 0 & 0 & F_{2,2} & 0 & 0\\
0 & 0 & G_{3,2} & 0 & 0 & E_{3,2} & B_{3,2} & D_{3,2} & 0 & 0 & F_{3,2} & 0\\
0 & 0 & 0 & G_{4,2} & 0 & 0 & E_{4,2} & B_{4,2} & 0 & 0 & 0 & F_{4,2}\\
\hdashline
F_{1,3} & 0 & 0 & 0 & G_{1,3} & 0 & 0 & 0 & B_{1,3} & D_{1,3} & 0 & 0\\
0 & F_{2,3} & 0 & 0 & 0 & G_{2,3} & 0 & 0 & E_{2,3} & B_{2,3} & D_{2,3} & 0\\
0 & 0 & F_{3,3} & 0 & 0 & 0 & G_{3,3} & 0 & 0 & E_{3,3} & B_{3,3} & D_{3,3}\\
0 & 0 & 0 & F_{4,3} & 0 & 0 & 0 & G_{4,3} & 0 & 0 & E_{4,3} & B_{4,3}
\end{array}\right]\left\langle\begin{array}{c}u_{1,1}\\u_{2,1}\\u_{3,1}\\u_{4,1}\\ \hdashline u_{1,2}\\u_{2,2}\\u_{3,2}\\u_{4,2}\\ \hdashline u_{1,3}\\u_{2,3}\\u_{3,3}\\u_{4,3}\end{array}\right\rangle+\left\langle\begin{array}{c}E_{1,1}u_{0,1}\\0\\0\\D_{4,1}u_{5,1}\\ \hdashline E_{1,2}u_{0,2}\\0\\0\\D_{4,2}u_{5,2}\\ \hdashline E_{1,3}u_{0,3}\\0\\0\\D_{4,3}u_{5,3}\end{array}\right\rangle, \tag{A11}$$

where the values on the surface ($u_{0,j}$) and the values at the maximum depth ($u_{5,j}$) are extracted and written in a vector. Following the order of j, the matrix in Eq. (A11) may be divided into block matrices as indicated by the dashed lines. This sparse matrix has only seven diagonals and can be divided into J^2 block matrices (each block has M^2 elements). Two diagonal blocks $G_{m,1}$ and $F_{m,3}$ occurring at the corners are due to the use of the periodic sidewall boundary conditions. Equation (A11) can be written in the form of Eqs. (11a) and (11b). If $J=1$, the sparse matrix reduces to a tridiagonal matrix, which has been encountered in two-dimensional models.

In the case of Fig. 3(c), approximate sidewall boundary conditions set $u_{m,0}=u_{m,1}$ and $u_{m,4}=u_{m,3}$. Following the same procedure as in the previous case, Eq. (A10) can also be transformed into a sparse matrix similar to Eq. (A11). The new sparse matrix has only five diagonals without the block matrices $G_{m,1}$ and $F_{m,3}$ occurring at the corners. The different block matrices are given in Eq. (12).

[1] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP, New York, 1994), Chap. 6.

[2] M. F. Levy and A. A. Zaporozhets, ''Target scattering calculations with the parabolic equation method,'' J. Acoust. Soc. Am. **103**, 735–741 (1998).

[3] G. J. Orris and M. D. Collins, ''The spectral parabolic equation and three-dimensional backscattering,'' J. Acoust. Soc. Am. **96**, 1725–1731 (1994).

[4] D. Lee, G. Botseas, and W. L. Siegmann, ''Examination of three-dimensional effects using a propagation model with azimuth-coupling capability (FOR3D),'' J. Acoust. Soc. Am. **91**, 3192–3202 (1992).
[5] M. D. Collins and S. A. Chin-Bing, ''A three-dimensional parabolic equation model that includes the effects of rough boundaries,'' J. Acoust. Soc. Am. **87**, 1104–1109 (1990).
[6] A. T. Abawi, W. A. Kuperman, and M. D. Collins, ''The coupled mode parabolic equation,'' J. Acoust. Soc. Am. **102**, 233–238 (1997).
[7] G. A. Athanassoulis and A. M. Prospathopoulos, ''Three-dimensional acoustic scattering of a source-generated field from a cylindrical isliand,'' J. Acoust. Soc. Am. **100**, 206–218 (1996).
[8] D. Zhu and L. Bjørnø, ''A hybrid 3-D, two-way IFD PE model for 3-D acoustic backscattering,'' J. Comput. Acoust. **7**(2), 133–145 (1999).
[9] M. D. Collins and R. B. Evans, ''A two-way parabolic equation for acoustic backscattering in the ocean,'' J. Acoust. Soc. Am. **91**, 1357–1368 (1992).
[10] D. Lee and M. H. Schultz, *Numerical Ocean Acoustic Propagation in Three Dimensions* (World Scientific, Singapore, 1995), Chaps. 4 and 5.
[11] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes*, 2nd ed. (Cambridge U. P., Cambridge, 1992), Chap. 2.
[12] R. B. Evans, ''A coupled mode solution for acoustic propagation in a waveguide with stepwise depth variations of a penetrable bottom,'' J. Acoust. Soc. Am. **74**, 188–195 (1983).

Quantitative imaging using a time-domain eigenfunction method

Feng Lin
Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627

Adrian I. Nachman
Department of Mathematics, University of Rochester, Rochester, New York 14627

Robert C. Waag[a)]
Departments of Electrical and Computer Engineering and Radiology, University of Rochester, Rochester, New York 14627

(Received 10 August 1999; accepted for publication 29 February 2000)

An inverse scattering method that uses eigenfunctions of a scattering operator at a single frequency is extended to include the full range of frequencies present in the incident pulse waveform. The resulting so-called time-domain eigenfunction method is shown to yield a modulated version of the scattering potential. The potential is recovered by a demodulation process using cross correlation with a reference. Including an adaptive delay in the reference is shown to compensate partially for the linearization of the Born approximation and to extend its valid range. The k-space window of the time-domain solution is expressed in terms of the incident waveform and shown to be smoother than that of a single-frequency solution. The time-domain method is examined using both calculated and measured data. In the calculations, an exact solution for scattering from one or multiple nonconcentric cylinders is used to obtain the scattered field. In the measurements, a novel ring-transducer system was employed to obtain the incident and total fields. The results of simulations and experiments show that the method is robust and accurate for the size of objects considered and that the point resolution approaches one-half the wavelength at the pulse center frequency. © *2000 Acoustical Society of America.* [S0001-4966(00)01806-3]

PACS numbers: 43.20.Fn, 43.60.Pt, 43.35.Wa, 43.80.Qf [ANN]

INTRODUCTION

Interest in computing quantitative ultrasonic images from transmission and reflection data obtained for many directions of incidence has been stimulated by the implementation and clinical utility of systems that employ x-rays to produce tomograms from a number of projections. X-ray tomographic algorithms were first applied to ultrasonic data in pioneering work.[1,2] The use of x-ray algorithms requires the oversimplifying approximation of a straight-line path for ultrasonic propagation. In later work,[3–10] the wave equation was used to model propagation. Recent research has focused on incorporation of practical ultrasonic beams and measurement configurations[11,12] and on algorithms that incorporate multiple scattering effects.[13–17] However, such techniques have not yet been implemented in practice; they await an appropriate combination of theoretical advances, mathematical and engineering expertise, and experimental apparatus.

Recent progress in theory provides a foundation for the development of efficient new inverse scattering methods. This theoretical progress has shown that the eigenfunctions and eigenvalues of an operator associated with scattering measurements can be used to focus on distributed scattering objects and to form images efficiently.[18,19] A ring-transducer system has been implemented for measurements.[20–22] The ring transducer consists of 2048 elements, each of which can be used independently as a transmitter or receiver. This fixed transducer configuration avoids signal degradation from phase jitter and excessive scanning time associated with moving transducers. The control electronics associated with the ring transducer provides the capability to program arbitrary transmit waveforms. The ring-transducer elements are spaced at intervals of less than one-half of the wavelength at the nominal center frequency of 2.5 MHz to assure sampling that provides all the available spatial-frequency information in the scattered wave.

Quantitative ultrasonic imaging requires measurement of high-frequency scattering data to achieve resolution comparable to magnetic resonance imaging and x-ray computed tomography. However, traditional fixed-frequency ultrasonic imaging algorithms are hampered by instabilities and by slow or erroneous convergence at high frequencies. To avoid these problems, imaging algorithms implemented in the time domain are attractive because the entire bandwidth of a signal can be used to provide information for the reconstruction process. A time-domain eigenfunction method has consequently been developed and is presented in this paper. The time-domain method produces a reconstruction that is modulated by the incident waveform. An adaptive demodulation algorithm is employed to compensate partially for the linearization introduced by the Born approximation and thus extend the valid range of the approximation. Compared to a single-frequency reconstruction, the time-domain reconstruction has a more desirable k-space window. The window can

[a)]Electronic mail: waag@ece.rochester.edu

be further adjusted by specification of the temporal incident waveform or a postprocessing k-space weight.

This paper describes results first reported in Ref. 23. Independently, a different approach for time-domain inversion using a Born approximation has also been reported.[24] That approach is implemented directly in the time domain and employs a method similar to shift-and-add for b-scan imaging rather than the scattering operator eigenfunctions on which the approach in this paper is based.

In Sec. I, the eigenfunction method of Ref. 19 is extended to the time domain and the adaptive demodulation process is described. The reconstruction of radially symmetric objects is examined to gain insight about the algorithm. The computational method and experimental procedure employed in this investigation are described in Sec. II. In Sec. III are reconstructions of cylindrical objects of various internal complexity from calculated and measured scattering data. The results are discussed in Sec. IV. Conclusions are presented in Sec. V.

I. THEORY

A. Configuration and statement of the problem

A two-dimensional circular scattering configuration is considered in this paper. An incident plane wave P_i traveling in the direction $\boldsymbol{\alpha}$ is scattered by an inhomogeneous region and the scattered field P_s is measured in direction $\boldsymbol{\theta}$. For each harmonic of frequency f, the total field of acoustic pressure $P=P_i+P_s$ is given by the Lippman–Schwinger equation[25]

$$P(\mathbf{x},\boldsymbol{\alpha},f)=e^{jk\boldsymbol{\alpha}\cdot\mathbf{x}}-k^2\int G_0(\mathbf{x}-\mathbf{y},f)q(\mathbf{y})P(\mathbf{y},\boldsymbol{\alpha},f)dy, \quad (1)$$

where $G_0(\mathbf{x}-\mathbf{y},f)$ is the Green's function for the Helmholtz equation in a homogeneous medium,

$$q(\mathbf{x})=1-c_0^2/c^2(\mathbf{x}) \quad (2)$$

is the scattering potential of the inhomogeneous object, c_0 and $c(\mathbf{x})$ are the sound speeds in the background and the inhomogeneous region, respectively, and k is the wave number defined as $2\pi f/c_0$. The definition of q in this paper differs by a factor of k^2 from that in Ref. 19 because this factor is explicitly included in the integral of Eq. (1). The scattering potential defined here is a function of sound speed in the medium. Density variations and absorption are assumed to be insignificant.

The inverse scattering problem is: Given $P(r\boldsymbol{\theta},\alpha,f)$ for $-\pi\leqslant\theta,\alpha<\pi$ around a ring of sufficiently large radius r, reconstruct the scattering potential $q(\mathbf{x})$.

B. Single-frequency solution

The results in Refs. 19 and 26 are summarized here for convenience.

The scattering amplitude $A(\theta,\alpha)$, i.e., the far-field pattern of the scattered pressure, is defined in two dimensions as

$$A(\theta,\alpha)=\lim_{r\to\infty} j\sqrt{8\pi jkr}e^{-jkr}P_s(r,\theta,\alpha). \quad (3)$$

If instead of a single plane wave, the incident pressure is a superposition of plane waves traveling in different directions, the incident pressure may be expressed

$$P_i(\mathbf{x})=\int f(\alpha)e^{jk\boldsymbol{\alpha}\cdot\mathbf{x}}d\alpha, \quad (4)$$

and the far-field pattern of the resulting scattered pressure is then

$$Af(\theta)=\int A(\theta,\alpha)f(\alpha)d\alpha. \quad (5)$$

Equation (5) defines the far-field operator A that maps the incident wave distribution f to the corresponding far-field pattern. Eigenvalues λ_i and orthonormal eigenfunctions f_i of the operator A satisfy

$$Af_i(\theta)=\lambda_i f_i(\theta). \quad (6)$$

The total pressure associated with the incident wave distribution f_i in a medium containing the scattering potential $q(\mathbf{x})$ is called the retransmitted field F_i and defined by

$$F_i(\mathbf{x})=\int f_i(\alpha)P(\mathbf{x},\alpha)d\alpha, \quad (7)$$

in which P is the total pressure associated with an incident plane wave. For the special case of a homogeneous medium, the retransmitted field simplifies to

$$E_i(\mathbf{x})=\int f_i(\alpha)e^{jk\boldsymbol{\alpha}\cdot\mathbf{x}}d\alpha. \quad (8)$$

A regularized solution of the inverse scattering problem may be found by minimizing the L^2 norm of $q(\mathbf{x})$ with respect to an appropriate weight $W(\mathbf{x})$, subject to the constraint that the far-field scattering pattern of q be equal to the measured pattern. In Ref. 19, the method of Lagrange multipliers was employed to obtain a solution of this constrained minimization problem and to show that the minimizer q_M has the form

$$q_M(\mathbf{x})=\frac{1}{W(\mathbf{x})}\sum_l\sum_m Q_{lm}F_l(\mathbf{x})\overline{F}_m^*(\mathbf{x}), \quad (9)$$

in which the coefficients Q_{lm} are determined from the system of equations

$$\delta_{ij}\lambda_i=k^2\sum_l\sum_m\left[\int\frac{F_i(\mathbf{x})E_j^*(\mathbf{x})F_l(\mathbf{x})\overline{F}_m^*(\mathbf{x})}{W(\mathbf{x})}dx\right]Q_{lm}, \quad (10)$$

and $\overline{F}_m^*(\mathbf{x})$ is defined by the relation

$$\overline{F}_m^*(\mathbf{x})=\int f_m^*(\alpha)P(\mathbf{x},\alpha+\pi)d\alpha. \quad (11)$$

The field $\overline{F}_m^*(\mathbf{x})$ is the complex conjugate of the retransmitted field corresponding to an incoming condition at infinity.

The above system is still nonlinear. However, in the case of weak scattering, the Born approximation can be used. The F fields are then reduced to E fields and an explicit analytic expression for the coefficients Q_{lm} was found in Ref. 19 for the case $W(\mathbf{x})\equiv 1$ to be

$$Q_{lm}=\frac{1}{8\pi^2}\int\int |\sin(\theta-\alpha)|A(\theta,\alpha)f_l^*(\theta)f_m(\alpha)d\alpha\, d\theta. \tag{12}$$

An even simpler and more efficient procedure, first announced in Ref. 26, has been found for determination of the coefficients. In this procedure, motivated by Eq. (12), the kernel of a modified far-field scattering operator B is defined by

$$B(\theta,\alpha)=|\sin(\theta-\alpha)|A(\theta,\alpha), \tag{13}$$

and use is made of the observation in Ref. 19 that Eqs. (9) and (12) are valid for any choice of the orthonormal basis $\{f_m\}$. If such a basis consisting of eigenfunctions of the operator B can be found, then the corresponding coefficients Q_{lm} are zero for $l\neq m$ and $Q_{mm}=\tilde{\lambda}_m$, the eigenvalue associated with the normalized eigenfunction $\tilde{f}_m$ of operator B. The solution under the Born approximation can, therefore, be written

$$q_B(\mathbf{x})=\frac{1}{8\pi^2}\sum_m \tilde{\lambda}_m|\tilde{E}_m(\mathbf{x})|^2, \tag{14}$$

where $\tilde{E}_m$ is the retransmitted field associated with $\tilde{f}_m$.

The frequency-domain eigenfunction reconstruction under the Born approximation has been shown in Ref. 19 to be equivalent to the filtered backpropagation formula

$$q_B(\mathbf{x})=\frac{1}{8\pi^2}\int\int |\sin(\alpha-\theta)|A(\theta,\alpha)e^{jk\mathbf{x}\cdot(\boldsymbol{\theta}-\boldsymbol{\alpha})}d\alpha\, d\theta. \tag{15}$$

The above solution yields an approximation to the low-pass filtered version of the true potential q, i.e., in k space,

$$q_B(\mathbf{k})\approx q(\mathbf{k})\Pi(|\mathbf{k}|-4\pi|f|/c_0), \tag{16}$$

where $\mathbf{k}$ is the (two-dimensional) spatial-frequency vector,

$$\Pi(|\mathbf{k}|-a)=\begin{cases}1, & |\mathbf{k}|\leqslant a\\ 0, & |\mathbf{k}|>a\end{cases}, \tag{17}$$

is a circular window, and $q(\mathbf{k})$ is the spatial Fourier transform of the true scattering potential. In Eq. (16) and in the following analysis, the notation uses the quantum-mechanics convention in which a symbol represents a conceptual object and the argument indicates the domain in which an object is evaluated. Thus, $q(\mathbf{x})$ is the scattering potential in the space domain as a function of the spatial-position vector $\mathbf{x}$ and $q(\mathbf{k})$ is the two-dimensional Fourier transform as a function of the spatial-frequency vector $\mathbf{k}$. The sharp cutoff of the circular window causes fluctuations around the discontinuities in the inhomogeneous region.

C. Time-domain solution

The time-domain solution begins with consideration of a frequency-weighted incident plane wave transmitted in the direction $\boldsymbol{\alpha}$ that is expressed

$$P_i(\mathbf{x},\boldsymbol{\alpha},f)=g(f)e^{jk\boldsymbol{\alpha}\cdot\mathbf{x}}, \tag{18}$$

where $g(f)$ is the temporal Fourier transform of a pulse waveform $g(t)$. The corresponding operator B constructed from the scattered pressure $P_s(r,\theta,\alpha,f)$ is then frequency dependent with kernel

$$B_T(\alpha,\theta,f)=\lim_{r\to\infty} j\sqrt{8\pi jkr}e^{-jkr}|\sin(\theta-\alpha)|P_s(r,\theta,\alpha,f). \tag{19}$$

The reconstructed scattering potential under the Born approximation is obtained from the eigenvalues $\tilde{\lambda}_{Tm}$ and retransmitted fields $\tilde{E}_{Tm}$ associated with eigenfunctions of the operator B_T using

$$q_{TB}(\mathbf{x},f)=\frac{1}{8\pi^2}\sum_m \tilde{\lambda}_{Tm}(f)|\tilde{E}_{Tm}(\mathbf{x},f)|^2, \tag{20}$$

or, in the time domain,

$$q_{TB}(\mathbf{x},t)=\frac{1}{8\pi^2}\sum_m \tilde{\lambda}_{Tm}(t)\times\tilde{E}_{Tm}(\mathbf{x},t)\times\tilde{E}_{Tm}(\mathbf{x},-t), \tag{21}$$

where $q_{TB}(\mathbf{x},t)$, $\tilde{\lambda}_{Tm}(t)$, and $\tilde{E}_{Tm}(\mathbf{x},t)$ are the inverse temporal Fourier transforms of $q_{TB}(\mathbf{x},f)$, $\tilde{\lambda}_{Tm}(f)$, and $\tilde{E}_{Tm}(\mathbf{x},f)$, respectively.

Substitution of Eqs. (3), (13), and (18) in Eq. (19) yields

$$B_T(\alpha,\theta,f)=g(f)B(\alpha,\theta,f). \tag{22}$$

The relation between the eigenvalues and retransmitted fields of the operators B_T and B can then be expressed

$$\tilde{\lambda}_{Tm}(f)=g(f)\tilde{\lambda}_m(f), \quad \tilde{E}_{Tm}(\mathbf{x},f)=\tilde{E}_m(\mathbf{x},f). \tag{23}$$

Thus,

$$q_{TB}(\mathbf{x},f)=g(f)q_B(\mathbf{x},f). \tag{24}$$

The potential $q_B(\mathbf{x},f)$ is the single-frequency reconstruction at f, so in the spatial-frequency domain

$$q_{TB}(\mathbf{k},f)\approx g(f)q(\mathbf{k})\Pi(|\mathbf{k}|-4\pi|f|/c_0). \tag{25}$$

The corresponding time-domain quantity is

$$q_{TB}(\mathbf{k},t)\approx q(\mathbf{k})\gamma(\mathbf{k},t), \tag{26}$$

in which

$$\gamma(\mathbf{k},t)=\left[\int_{-\infty}^{-|\mathbf{k}|c_0/(4\pi)}+\int_{|\mathbf{k}|c_0/(4\pi)}^{+\infty}\right]g(f)e^{-j2\pi ft}df. \tag{27}$$

The scattering potential given by Eq. (21) is thus time dependent and is modulated by the waveform $\gamma(\mathbf{k},t)$. The final step is to recover $q(\mathbf{x})$ from $q_{TB}(\mathbf{x},t)$. A time-independent potential $q'_B(\mathbf{x})$ can be obtained by demodulation or cross correlation. The result is

$$q'_B(\mathbf{k})=\int_{-\infty}^{+\infty} q_{TB}(\mathbf{k},t)\gamma^*(\mathbf{k},t)dt \Big/ \int_{-\infty}^{+\infty}|\gamma(\mathbf{k},t)|^2dt, \tag{28}$$

for those $\mathbf{k}$ for which the integral in the denominator is nonzero. For such $\mathbf{k}$, in view of Eq. (26),

$$q(\mathbf{k})\approx q'_B(\mathbf{k}). \tag{29}$$

For a bandlimited incident pulse $g(t)$, $\gamma(\mathbf{k},t)$ vanishes for large $|\mathbf{k}|$ and Eq. (26) shows that no information about the high spatial-frequency components of q can be obtained from q_{TB}. Instead of introducing an artificial k-space cutoff, the demodulation process may be modified to yield a win-

dow determined by the incident waveform $g(t)$. The modification is suggested by use of Plancherel's theorem in Eq. (28) that together with Eq. (27) shows

$$q'_B(\mathbf{k}) = \int_{-\infty}^{+\infty} q_{TB}(\mathbf{k},t)g(t)dt \Big/ 2\int_{|\mathbf{k}|c_0/(4\pi)}^{+\infty} |g(f)|^2\, df. \quad (30)$$

This leads to the natural definition

$$w(\mathbf{k}) = \int_{|\mathbf{k}|c_0/(4\pi)}^{+\infty} |g(f)|^2\, df \Big/ \int_0^{+\infty} |g(f)|^2\, df. \quad (31)$$

Then, Eqs. (29) and (30) yield

$$q(\mathbf{k})w(\mathbf{k}) \approx q''_B(\mathbf{k}), \quad (32)$$

in which

$$q''_B(\mathbf{x}) = \int_{-\infty}^{+\infty} q_{TB}(\mathbf{x},t)g(t)dt \Big/ \int_{-\infty}^{+\infty} g^2(t)dt. \quad (33)$$

The calculation of the solution q''_B using the time-domain method presented here is more efficient than the corresponding calculation of q'_B and does not pose the same difficulty for large $|\mathbf{k}|$.

For a bandlimited incident waveform $g(t)$, the k-space window of the time-domain solution drops gradually to zero when $|\mathbf{k}|$ becomes large, as opposed to the sharp cutoff in the window of the single-frequency solution. Hence, less ringing is produced in the point-spread function of the time-domain solution. Furthermore, the window of the time-domain solution may be adjusted to minimize ringing in the point-spread function in exchange for an increased width of the central part of the point-spread function. To improve the k-space window, either of two approaches can be employed. One approach is a postprocessing procedure. Since $g(t)$ is known, $w(\mathbf{k})$ may be first calculated using Eq. (31). Then, the scattering potential estimate obtained with the new window can be determined from $q''_B(\mathbf{k})$ using

$$q_C(\mathbf{k}) = q''_B(\mathbf{k}) \frac{w_S(\mathbf{k})}{w(\mathbf{k})}. \quad (34)$$

In this expression, $w_S(\mathbf{k})$ is a standard window, e.g., the Blackman window.[27] The other approach is a preprocessing procedure that specifies the incident waveform satisfy

$$w_S(\mathbf{k}) = \int_{|\mathbf{k}|c_0/(4\pi)}^{+\infty} |g_S(f)|^2\, df \Big/ \int_0^{+\infty} |g_S(f)|^2\, df. \quad (35)$$

From this expression, $g_S(t)$ is found to be the inverse Fourier transform of

$$|g_S(f)| = K\sqrt{\mathrm{sgn}(-f)\frac{dw_S(4\pi f/c_0)}{df}}, \quad (36)$$

where K is any positive constant.

The solution in Eq. (33) for the potential is still under the Born approximation that assumes no change in the incident field and no disturbance in the scattered field during the scattering process. However, the arrival time of the incident field and the scattered field may be altered in the scattering region by variations in wave speed, this shift causes an error in the cross correlation. An adaptive demodulation is, therefore, employed to find the maximum of the cross correlation function of $q_{TB}(\mathbf{x},t)$ and $g(t)$ at each position and, thus, compensate for a wave-speed difference. In this adaptive method, the potential is expressed

$$q_A(\mathbf{x}) = \int_{-\infty}^{+\infty} q_{TB}(\mathbf{x},t)g[t+\Delta t(\mathbf{x})]dt \Big/ \int_{-\infty}^{+\infty} g^2(t)dt, \quad (37)$$

where the offset $\Delta t(\mathbf{x})$ is an estimate of the time shift between $q_{TB}(\mathbf{x},t)$ and $g(t)$. For the reconstructions reported in this paper, the estimates of the time shift were obtained by numerically picking the Δt that produced the highest cross correlation of q_{TB} and g.

D. Radially symmetric objects

For the general case, the eigenvalues and eigenfunctions of a scattering operator can only be calculated numerically. Nevertheless, for some special cases, e.g., the case of radially symmetric scatterers, the scattering operator and the reconstruction can be obtained analytically.

The scattering potential of a radially symmetric object, $q(\mathbf{x})$, depends only on $|\mathbf{x}|$. In this case,

$$P_s(r,\theta,\alpha,f) = P_s(r,\theta+\beta,\alpha+\beta,f) \quad (38)$$

for any angular value β. The scattering amplitude $A(\theta,\alpha)$ then depends only on $(\theta-\alpha)$ and the eigenfunctions of the operator A have the form

$$f_m(\theta) = \frac{1}{\sqrt{2\pi}} e^{jm\theta}, \quad m = 0,\pm 1,\pm 2,\ldots. \quad (39)$$

The eigenvalues are

$$\lambda_m(f) = \int_0^{2\pi} A(\theta,0,f)e^{-jm\theta}\, d\theta, \quad m = 0,\pm 1,\pm 2,\ldots. \quad (40)$$

The scattering operator B is also circularly symmetric. Consequently, the eigenfunctions of B are the same as those of A, and the eigenvalues of B are

$$\tilde{\lambda}_m(f) = \int_0^{2\pi} B(\theta,0,f)e^{-jm\theta}\, d\theta = a_m * \lambda_m(f), \quad (41)$$

where $a_m = 2[1+(-1)^m]/(1-m^2)$, $m = 0,\pm 1,\pm 2,\ldots$, are the Fourier coefficients of $|\sin(\theta)|$. Then, the retransmitted field associated with the eigenfunction f_m becomes

$$E_m(r,\theta,f) = \sqrt{2\pi} j^m e^{jm\theta} J_m(kr), \quad (42)$$

where $J_m(\cdot)$ is a Bessel function of the first kind and order m. Noteworthy is that the operators A and B have the same retransmitted fields and that these fields are independent of the scattering object.

As an example, consider an infinitely long cylinder with radius a, sound speed c_1, and density ρ_1. Assume the cylinder is centered on the z axis and the sound speed and density of background are c_0 and ρ_0, respectively. The scattered field prodced by a plane wave traveling in the direction of $\theta = 0$ with a harmonic time dependence $e^{-j2\pi ft}$ can be expressed as an orthogonal function expansion in which the coefficients are determined by matching boundary conditions on the surface of the cylinder. The far-field asymptotic form of the scattered pressure is[25]

$$P_s(r,\theta,f)=\sqrt{\frac{2}{\pi jkr}}e^{jkr}\sum_{m=-\infty}^{\infty}D_m(f)\cos(m\theta), \quad (43)$$

where

$$D_m=\frac{J_m(k_1a)J'_m(ka)-\alpha J'_m(k_1a)J_m(ka)}{\alpha J'_m(k_1a)H_m^{(1)}(ka)-J_m(k_1a)H'^{(1)}_m(ka)}, \quad (44)$$

in which $k_1=2\pi f/c_1$, $\alpha=(c_0\rho_0)/(c_1\rho_1)$, $H_m^{(1)}(\cdot)$ is a Hankel function of the first kind and order m, and $J'_m(\cdot)$ and $H'^{(1)}_m(\cdot)$ are derivatives of $J_m(\cdot)$ and $H_m^{(1)}(\cdot)$, respectively. The scattering operator A can then be expressed as

$$A(\theta,\alpha,f)=4j\sum_{m=-\infty}^{+\infty}D_m(f)e^{jm(\theta-\alpha)}. \quad (45)$$

The eigenvalues of A are the Fourier coefficients of $A(\theta,0,f)$ with respect to θ. Hence,

$$\lambda_m(f)=8\pi jD_m(f). \quad (46)$$

The eigenvalues of operator B are obtained from Eq. (41) by substituting for $\lambda_m(f)$ the expression given in Eq. (46).

II. METHODS

A. Sampling relations

To implement the foregoing reconstruction method numerically, a discrete version of the algorithm is employed. The sampling theorem in the time domain (with respect to t) and in the space domain (with respect to θ and α) determines the minimum increments needed for reconstruction without aliasing.

In the time domain, the sampling frequency is determined by the maximum frequency $f_{\max}$ of the measurement and must satisfy the relation

$$f_s\geqslant 2\,f_{\max}. \quad (47)$$

In the space domain, the number of transmit views M and the number of receive angles N around the ring are chosen to be the same to make the discrete (matrix) version of the scattering operators A and B square. Then, θ and α have the same increment $\Delta\phi=2\pi/M$. For a cylindrical scattering object with radius a, the lowest spatial-frequency space-sampling rate satisfies the Nyquist criterion that can be expressed as

$$\frac{f}{c_0}\Delta\phi\leqslant\frac{1}{2a}. \quad (48)$$

Thus,

$$M\geqslant 4\pi\frac{f_{\max}}{c_0}a. \quad (49)$$

This indicates the number of views is independent of the size of the measurement ring and only determined by the highest temporal frequency and size of the object.

For the ring transducer, the maximum frequency $f_{\max}$ is 3.5 MHz so that $f_s\geqslant 7.0$ MHz. Assuming $c_0=1.5$ mm/μs and $a=3$ mm yields $M\geqslant 88$. In calculations and measurements of scattering for a 3-mm-radius object, the sampling frequency was 20 MHz and the number of views was chosen to be 128.

TABLE I. Parameters used in the reconstruction of a homogeneous cylinder from calculated data.

Parameter	Symbol	Value
Sound speed of background	c_0	1.5 mm/μs
Density of background	ρ_0	1.0 g/mm^3
Sound speed of the cylinder	c_1	1.515 mm/μs
Density of the cylinder	ρ_1	1.0 g/mm^3
Radius of the cylinder	a	3.0 mm
Radius of measurement ring	r	7500.0 mm
Number of views	M	128
Sampling frequency	f_s	20 MHz
Center frequency of the Gaussian pulse	f_c	2.5 MHz
Bandwidth of the Gaussian pulse	BW	1.7 MHz

B. Computational methods

In the calculations of time-domain scattering, data were found at M far-field receive angles equally spaced between 0 and 360° for each of M transmit angles also equally spaced between 0 and 360°. An orthogonal function expansion [Eq. (43)] of the field was employed for the case of a homogeneous cylinder. An exact solution[28] was used for a nonradial scattering object consisting of multiple nonconcentric cylinders. For arbitary scattering objects, numerical approaches, like the k-space method[29] or a pseudospectral method,[30] may be used to calculate the propagation.

The incident waveform was a Gaussian pulse having the form

$$g(t)=Ce^{-t^2/2\sigma^2}\sin(2\pi f_c t), \quad (50)$$

in which C is the pulse amplitude, f_c is the center frequency of the signal, and σ is the width parameter. This parameter can be written in terms of the −6-dB bandwidth BW as

$$\sigma=\frac{\sqrt{2\ln 2}}{\mathrm{BW}\,\pi}. \quad (51)$$

The values of the parameters used in the calculations of scattering are listed in Table I.

The computations yield a discrete representation of the time-domain scattering operators A and B, which are denoted as the A matrix and B matrix, respectively, in this paper. For each harmonic frequency within the passband, the eigenvalues, eigenfunctions, and retransmitted fields of the B matrix were calculated numerically. The frequency-domain scattering potential was obtained by summing the retransmitted field energies each weighted by the corresponding eigenvalue. A time-domain solution as a function of space and time was found using an inverse Fourier transform. For radially symmetric objects, the eigenvalues and retransmitted fields were obtained analytically using Eqs. (46), (41), and (42). Adaptive demodulation was implemented by cross correlating the aligned time-dependent scattering potential and the incident waveform.

Aliasing caused by the term $|\sin(\theta-\alpha)|$ in the calculation of eigenvalues and eigenvectors of the B matrix may

produce reconstruction errors. The aliasing can be understood by considering a radially symmetric object. Equation (41) indicates that eigenvalues $\tilde{\lambda}_m(f)$ of the operator B are the convolution of the Fourier coefficients a_m of $|\sin(\theta)|$ with the eigenvalues $\lambda_m(f)$ of the operator A. Since the a_m decrease slowly and do not vanish as m increases and extends to infinity, the eigenvalues $\tilde{\lambda}_m(f)$ then also do not vanish. For θ continuous, convolution produces an accurate solution of eigenvalues. However, for the discrete case, even if sampling is within the noted bounds of the sampling relations, the finite number of samples not only limits the highest order $\tilde{\lambda}_m(f)$ to $|m|=M/2-1$, but also produces aliasing of eigenvalues from high orders to low orders. One approach to reduce aliasing is to use more samples of the angles θ and α than required by the sampling conditions. This approach is, however, inefficient because more measurement time and data are required.

For the case of radially symmetric objects, the aliasing error can be kept below any specified threshold without changing the number of angular samples because the $\lambda_m(f)$ may be convolved with the analytic expression for a_m extending to sufficiently high m. The same idea can be applied to nonradial objects by interpolating the B matrix to a larger size before the eigenvalues of the matrix are calculated numerically. The interpolation may be implemented by expressing the two-dimensional Fourier series for the elements in the B matrix as the convolution of terms in the two-dimensional Fourier series for each element in the A matrix and for the factor $|\sin(\theta-\alpha)|$. The result may be written

$$B_{mn}=c_{mn}*A_{mn}, \tag{52}$$

where $m,n=\ldots,-1,0,1,\ldots$, are the Fourier indices of θ and α, the c_{mn} are the Fourier coefficients of the term $|\sin(\theta-\alpha)|$ and extend to high orders of m and n, while the A_{mn} are the Fourier coefficients of the A matrix and have limited extent if the sampling conditions are satisfied. The Fourier coefficients B_{mn} are then calculated by convolving A_{mn} with c_{mn} evaluated for large m and n to obtain the B matrix at a greater number of angles than that of the A matrix. By choice of an appropriately large range for the values of m and n, the eigenvalues of the B matrix may be computed with aliasing less than any desired level. Interpolation that quadrupled the dimensions of the B matrix relative to the dimensions of the original A matrix and provided the equivalent of four times as many views was found to be sufficient and was used for all the reconstructions presented here.

C. Experimental procedure

Three tissue-mimicking cylindrical scattering objects each with a 6-mm diameter were used for experimental measurements. The first is a homogeneous cylinder made of agar. The second is an agar cylinder with tiny glass beads randomly distributed inside. The third is an agar cylinder with randomly positioned glass beads and three nylon filaments aligned parallel to the axis of the cylinder.

The scattering objects were produced in a manner similar to that described elsewhere.[31,32] A 0.4-liter solution of distilled water and 16 grams of agar was heated to 90 °C and stirred until the solution clarified. If the solution was for a phantom with randomly distributed scatterers, 1.6 g of glass beads with a mean diameter of 83 μm were added at this time. The solution was cooled to 60 °C while being stirred magnetically to avoid bubbles and to distribute the glass beads uniformly if they were present. The solution was poured into a mold after 20 ml of the preservative n-propanol was added. The mold was used to position nylon filaments if they were present and was rotated to avoid settling if glass beads were present.

The reconstruction of an object from measured data consisted of five steps.

1. Data acquisition

Spatially limited plane waves were transmitted from the ring transducer from M positions equally spaced around the ring without the scattering object present for the incident field measurement and with the scattering object present for the total field measurement. A backpropagation method was used to obtain the transmit waveforms that produced the desired incident wave.[33] The width of the plane wave was that needed to span the scattering object. A cosine roll-off was added to the two sides of a uniform central region to provide a smooth transition in amplitude and avoid dispersion. For a 3-mm-radius scattering object, a spatially limited plane wave having a 10-mm-wide central region with a 4-mm roll-off at each side was employed as the incident wave. The total field was measured around the ring for each incident view. To compensate for sound-speed changes due to the water temperature variation, the sound speed in the background was tracked using a probe beam during the measurement of both the incident and total fields. The probe beam was also a spatially limited plane wave produced in the same way as the incident wave, but the probe beam was directed to the side of the phantom so that the path of the beam was through the background when the scattering object was present. The sound speed in the background was estimated from knowledge of the arrival time and the travel distance of the probe beam. The speed estimate was used to equalize the time scale of all the waveforms.

2. Incident wave subtraction

An incident field $p_i(a,\alpha,\theta,t)$ compensated for time shift caused by temperature variation between the time of the incident wave measurement and the total field measurement was subtracted from the total field $p(a,\alpha,\theta,t)$ to obtain the scattered field $p_s(a,\alpha,\theta,t)$. Here, a denotes the radius of the measurement ring. Subtraction of the uncompensated field was unsatisfactory because small temperature variations produced sufficient time shift to make the subtraction inaccurate. The subtraction with the shift was expressed

$$p_s(a,\alpha,\theta,t)=p(a,\alpha,\theta,t)-p_i(a,\alpha,\theta,t+\Delta t), \tag{53}$$

in which $\Delta t(\alpha)$ is estimated from the probe beams. Assuming the arrival time shift between the probe beams of the two measurements is τ and the propagation distance of the probe beam is d, then

$$\Delta t = \tau \frac{2a}{d}. \tag{54}$$

3. Incident pulse estimation

To obtain experimentally the incident pulse $g(t)$ used in the demodulation step in the reconstruction of a quantitative image from measured scattering, the measured incident waveform at each view was backpropagated to the center of the ring again following the method in Ref. 33. The incident field at the center was denoted as $p_i(x,0,\alpha,t)$, in which the first two arguments are position in the Cartesian coordinates based on the incident direction α. (The x axis is orthogonal to the direction of α.) The estimate of the incident pulse was then obtained by averaging $p_i(x,0,\alpha,t)$ over space and the direction of incidence. This may be written

$$g(t) = \frac{1}{2\pi D} \int_{-\pi}^{\pi} \int_{-D/2}^{D/2} p_i(x,0,\alpha,t)\, dx\, d\alpha, \tag{55}$$

where D is the width of the spatially limited plane wave. The averaging takes into account lateral nonuniformity and view-to-view variation that are unavoidable in any measurement.

4. Far-field extrapolation

Far-field data for the reconstruction were obtained by propagating the data measured around the 75-mm radius of the ring transducer to a ring of much larger radius (7500 mm). The problem can be stated: Given $p_s(a,\theta,t)$, find $p_s(r,\theta,t)$ for $r>a$. Since each temporal harmonic frequency of the field satisfying the Helmholtz equation in cylindrical coordinates can be written as a series of weighted Hankel functions as done in Ref. 34, the far-field extrapolation of scattered field can be modeled as a temporal–spatial filter with transfer function that may be expressed

$$H(f,m) = \frac{H_m^{(1)}(2\pi f r/c_0)}{H_m^{(1)}(2\pi f a/c_0)}, \tag{56}$$

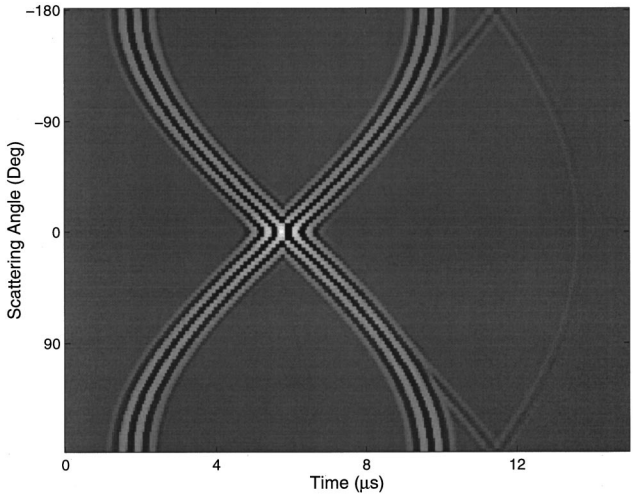


FIG. 1. Calculated far-field scattered pressure from a cylinder. The backward direction corresponds to $\theta=180°$ while the forward direction corresponds to $\theta=0$. Amplitude is shown on a 110-dB bipolar log scale.

where m is the index of Fourier coefficient associated with θ and f is temporal frequency.

5. Image reconstruction

Reconstruction in the time domain was accomplished as described in Sec. II B.

III. RESULTS

A. Model scattering data

A homogeneous cylinder was used for a radial scattering object. The calculated far-field pressure $p_s(r,0,\theta,t)$ scattered by this object is shown in Fig. 1 for the parameters in Table I. The highest energy is in the forward direction. The boundary produces reflections in all directions. Multiple reflections are clearly visible.

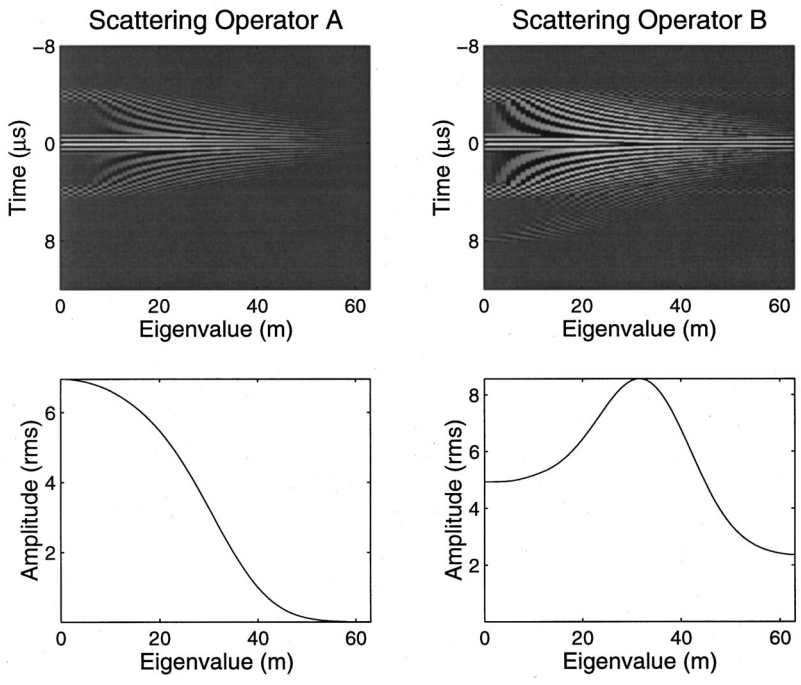


FIG. 2. Eigenvalues of scattering operators A and B. The upper-left and upper-right panels show the time-domain eigenvalues of A and B, respectively, on an 80-dB bipolar log scale. The lower-left and lower-right panels show the temporal rms amplitude of the eigenvalues of A and B, respectively.

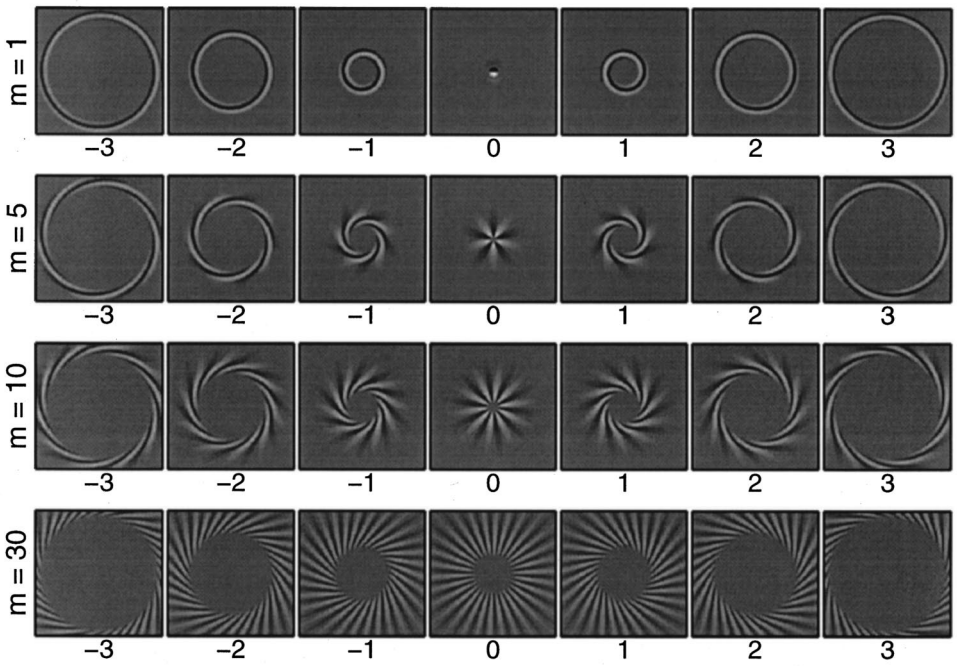


FIG. 3. Examples of retransmitted fields in a 10×10-mm^2 region. Amplitude is shown on a 40-dB bipolar log scale. Each row is a sequence of time steps for the retransmitted field whose order is indicated by m on the left. The number under each panel indicates the time step in units of μs.

The eigenvalues of the scattering operators A and B for this cylinder are shown in Fig. 2. The eigenvalues of A are, as noted, the spatial Fourier coefficients of the field scattered by the cylinder. They tend toward zero when the order m becomes large. The eigenvalues of B are the eigenvalues of A spread to higher orders of m by a convolution with the Fourier components of the term $|\sin(\theta-\alpha)|$.

Retransmitted fields weighted by $g(f)$ in the frequency domain for several orders or values of m are shown in Fig. 3. The fields propagate inward from the far field toward the center and then back out to the far field. An interesting phenomenon is that the high-order fields do not appear to propagate to the center, but only to a ring before propagating outward. The radius of the ring is proportional to the order m and is also determined by the temporal frequency content of the fields. This effect is the result of mth-order Bessel functions of the first kind being proportional to their argument raised to mth power for small values of the argument. The effect in the frequency domain was observed previously by Chen.[35] He called the phenomenon the skin effect, so that term is used here also.

The progression of the frequency-domain reconstruction

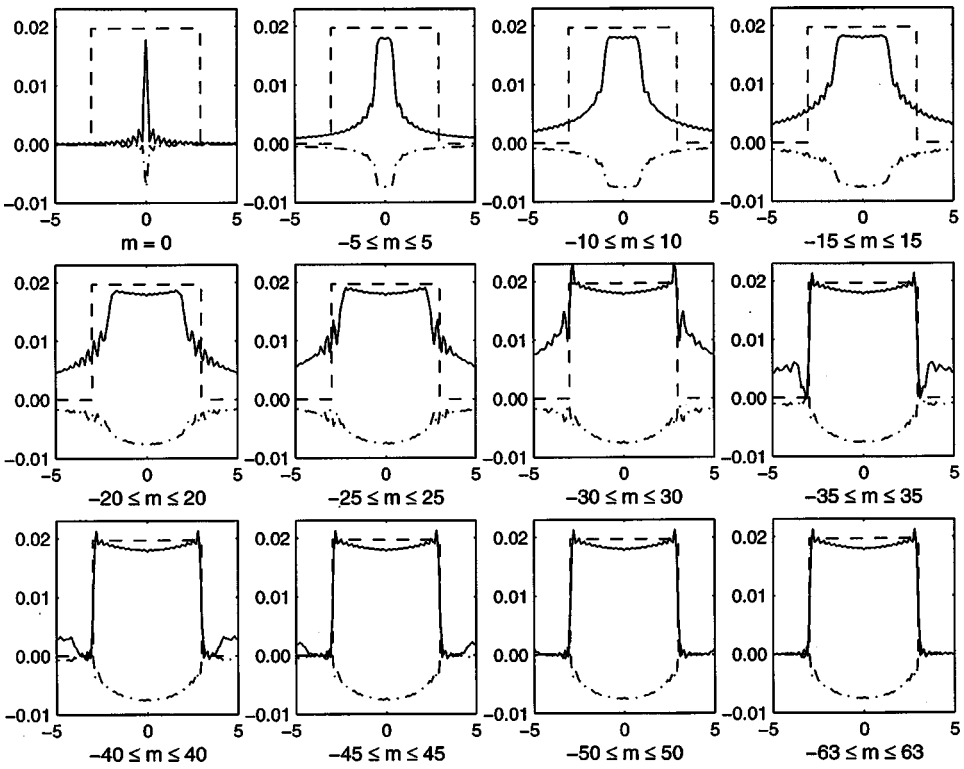


FIG. 4. Progression of the frequency-domain reconstruction of a homogeneous cylinder at 2.5 MHz. In each panel, the solid line is a cross section through the center of the real part of the reconstructed potential and the dash-dot line is the imaginary part. The dashed line indicates the true potential. The horizontal coordinate is space and ranges from -5 to 5 mm. The vertical coordinate is the scattering potential. The expression under each panel gives the lower and upper limits of orders used to construct the potential. The last panel is the final reconstruction from 128 views.

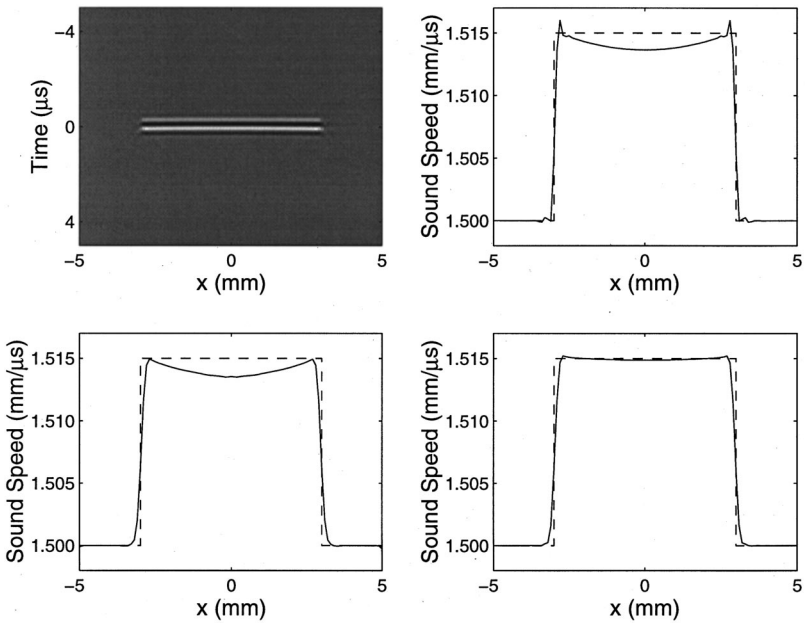


FIG. 5. Time-domain reconstructions of a cylinder. The upper-left panel is the time-dependent scattering potential with the amplitude shown on a linear gray scale. Reconstructions obtained using direct demodulation are shown in the upper-right panel for a k-space window determined by an incident pulse with a Gaussian temporal spectrum and in the lower-left panel for a k-space window with a Blackman radial shape. The reconstruction in the lower-right panel was obtained using the same Blackman k-space window and adaptive demodulation instead of direct demodulation. In the Cartesian plots, the solid line is the reconstruction, and the dashed line is the true value. In all the plots, the horizontal coordinate is distance along a line through the center of the cylinder.

as higher orders of m are included is illustrated in Fig. 4. The lower-order fields construct the potential near its center, while the potential at large radii is obtained when higher-order terms are added. The Born approximation also produces in the scattering potential a nonzero imaginary part that does not appear in the exact solution. This effect is due to the linearization in the Born approximation.[19]

Corresponding time-domain reconstructions are shown in Fig. 5. The time-dependent scattering potential is a version of the scattering potential temporally modulated by the incident waveform. The scattering potential obtained by demodulation using direct cross correlation is comparable to that of a single frequency reconstruction. However, the time-domain reconstruction is smoother due to the k-space window that results from the Gaussian spectrum of the pulse $g(t)$. The k-space window of the reconstruction was altered to the shape of a Blackman window in the postprocessing to eliminate the overshoot and undershoot near the edges of the cylinder. The disadvantage of this technique is a decrease in the slope of the object boundary in the reconstruction. In the adaptive result, the reconstructed potential inside the cylinder is much closer to the true value because the position of the reference in the cross correlation is compensated for the offset from the different speed of sound in the scattering object. This extends the valid range of the Born approximation.

Additional computations of scattering were made for a nonradial object consisting of a cylinder with three parallel filaments and random scatterers inside. The exact solution for multiple cylinders was employed to calculate the scattered wave. The parameters of the cylinder and background are those listed in Table I. The filaments were modeled as fluid cylinders with radius of 0.04 mm and sound speed of 3.0 mm/μs. The random scatterers were also modeled as cylinders each with a radius of 0.01 mm and sound speed of 2.25 mm/μs. The reconstruction of the scattering potential is shown in Fig. 6. The reconstructed speed of sound in the filaments is less than the true value because of inherent spatial averaging. However, if point resolution is defined as the -6-dB diameter of the point-spread function or the reconstruction of a point scatterer, the point resolution of this reconstruction based on the width of the filaments in the image is 0.33 mm or about one-half of the nominal wavelength at the center frequency of 2.5 MHz.

B. Measured scattering data

The total field, incident field, and scattered field for a representative view of the phantom with glass beads and three wires are shown in Fig. 7. The amplitude of the incident field is much higher than that of the scattered field because the scattering was weak. This makes the subtraction

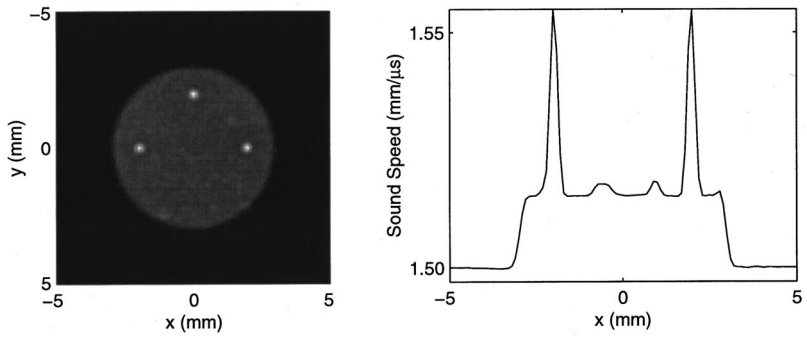


FIG. 6. Cylinder with random scatterers and three filaments reconstructed from calculated scattering. The left panel shows the scattering potential plotted on a linear scale. The right panel is a cross section of the potential at $y=0$.

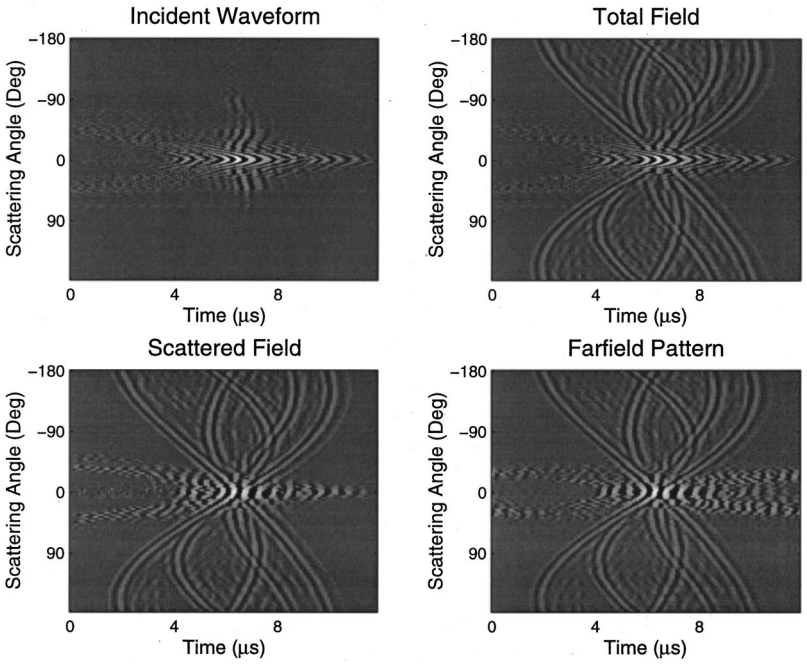


FIG. 7. Representative measured fields. The upper panels show amplitude on a 80-dB bipolar log scale while the lower panels show amplitude on a 70-dB bipolar log scale. In each panel, the backward direction corresponds to 180° while the forward direction corresponds 0°.

inaccurate, particularly in the forward direction. Also because the scattering is weak, the scattered field is mostly confined to an area within the reflections from the boundary of the cylinder. The three sinusoidally shaped wavefronts inside the reflections from the cylinder boundary are from the three filaments. The small signals within the area of reflections from the cylinder boundary are from the glass beads. The far-field pattern of the scattered field in this case is not much different from the scattered field on the measurement ring because the scattering object is small relative to the size of the measurement ring.

Reconstructions of the three scattering objects are shown in Fig. 8. The scattering objects distinguish themselves significantly from the background with sharp edges and the reconstruction of the homogeneous agar cylinder is relatively free of speckle. The random fluctuations in the middle and lower panels are from the glass beads. In the lower panels, the point contrast in three places is very high because the reflections from the fibers are strong. The -6-dB point resolution is 0.30 mm, half of the wavelength at the center frequency. The position of the images in the reconstructed field is the postition of the scattering objects relative to the center of the ring when the measurements were made.

IV. DISCUSSION

The scattering potential is assumed to be independent of frequency, and absorption is neglected in this paper. Although absorption may not be negligible in diagnostic imaging applications, the frequency-dependent dispersion that accompanies this absorption is negligible within the bandwidth of current imaging instruments so the assumption of a frequency-independent potential is reasonable from a practical viewpoint. The inclusion of density variations complicates the analysis but can be accomplished by a change of variable that converts the linearized wave equation with variable density to a wave equation of the form considered here. In any case, the reconstructions obtained from experimental data show the time-domain eigenfunction imaging method in this paper can be used to image objects like those studied.

Unlike the frequency-domain solution, the time-domain reconstruction decribed in Sec. I only yields pure real scattering potential because the derivation is based on the assumption that the scattering potential q is real. However, the Born approximation may produce an imaginary part in the reconstruction. For the case of complex scattering potential, the potential at a negative frequency is the complex conjugate of the potential at the corresponding positive frequency. Then, the demodulation in Eq. (37) can be expressed as

$$q_C(\mathbf{x}) = \int_{-\infty}^{+\infty} q_{TB}(\mathbf{x},t)\{g(t+\Delta t) + j\mathcal{H}[g(t+\Delta t)]\}dt \Big/ \int_{-\infty}^{+\infty} g^2(t)dt, \qquad (57)$$

where $\mathcal{H}$ denotes the Hilbert transform. Equation (57) yields the same real part of the scattering potential as Eq. (37) and also yields an imaginary part. For direct demodulation ($\Delta t = 0$), the imaginary part is similar to that of a frequency-domain solution. The imaginary part of the adaptive demodulation is, however, reduced by the offset Δt to a negligible value for the scattering objects in this study.

The so-called B-matrix approach has both advantages and limitations. The main advantage is the already noted computational efficiency that results from the off-diagonal Q_{lm} coefficients being zero in the expansion of the scattering potential. Another advantage is that the two-dimensional integration defined in Eq. (12) is not necessary. A theoretical limitation is that the existence of an orthonormal basis is assumed rather than proven to exist. Another limitation is the need to interpolate additional elements to keep reconstruction errors from aliasing small. The process of interpolation offsets savings in computation from the off-diagonal Q_{lm} terms being zero. On balance, however, the B-matrix ap-

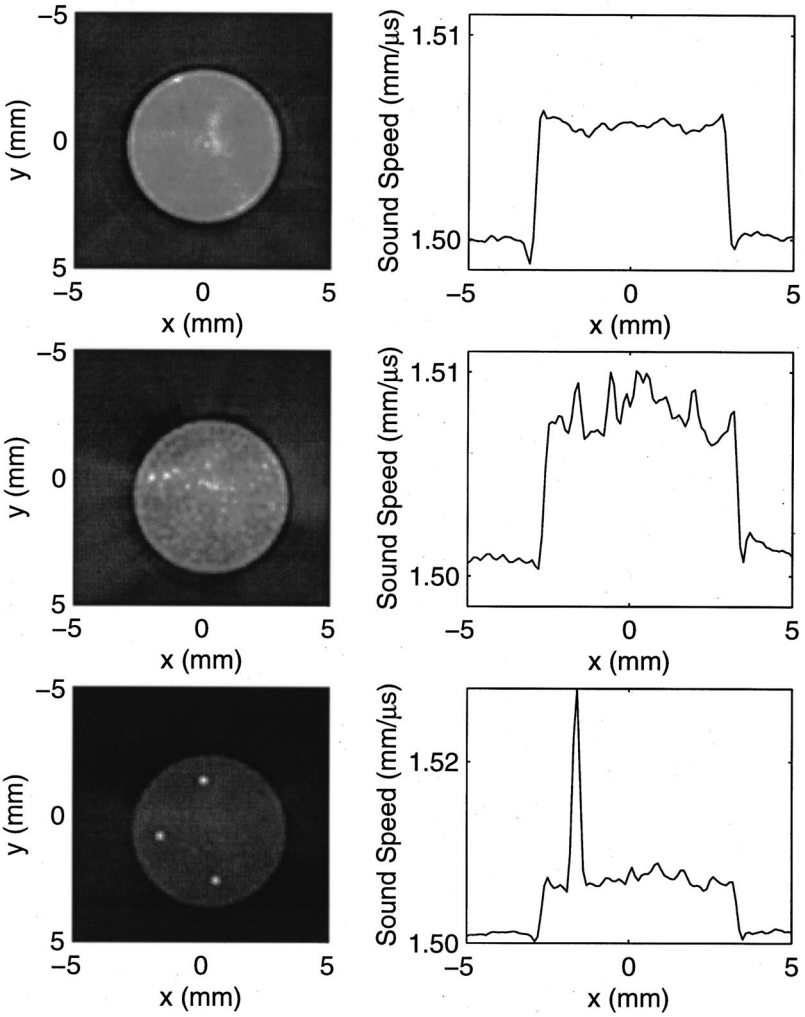


FIG. 8. Reconstructions of three scattering objects from measured data. The upper panels are for a 6-mm-diameter pure agar phantom. The middle panels are for a 6-mm-diameter agar phantom in which glass beads are uniformly distributed. The lower panels are for a 6-mm-diameter agar phantom with glass beads and three filaments. The left panels are plotted on a linear scale. The right panels show representative cross sections of the sound speed.

proach appears preferable over the A-matrix approach. All the results presented in this paper were obtained using the B-matrix approach.

The aliasing of eigenvalues caused by the term $|\sin(\theta-\alpha)|$ in the solution of the linearized reconstruction can be significant, as shown in Fig. 2, if interpolation is not used even when the sampling conditions are met. The proposed interpolation method for reduction of aliasing, however, increases the computation time considerably because the time needed is approximately proportional to the square of the number of rows or columns in the B matrix. A tradeoff, therefore, must be made between accuracy and computation time. In the studies reported here, interpolation that is equivalent to quadrupling the number of views produced satisfactory results. For radial objects, computation can be reduced by omission of terms associated with eigenvalues beyond the number that satisfies the sampling conditions. In this case, although interpolation is still needed to obtain the eigenvalues to a certain accuracy, the contribution of the terms with order higher than the number required by the sampling conditions is outside the scattering region where the medium is assumed to be homogeneous and where the scattering potential is consequently known.

Reconstruction from calculated data provides useful insight about various phenomena that influence the reconstruction because the calculations allow a wide range of parameters with more control of conditions than in an experimental setting. For single and multiple cylinders, the availability of exact solutions permits investigation of aliasing in the linear solution as well as the range over which the linearization itself yields a scattering potential that is close to the true value. The step discontinuity at the edge of a single cylinder permits study of the k-space window effect on rapid transitions. Additionally, the ability to obtain the scattered field directly avoids errors in the subtraction of the incident field from the total field.

The behavior of the retransmitted fields shown in Fig. 3 is reflected in the evolution of the image in Fig. 4 as higher orders of retransmitted fields are included in the reconstruction. Information about the scattering potential is contained in the scattered field only for radii greater than a value determined by the highest temporal frequency and the order of the retransmitted field. A consequence of this so-called skin effect is that fields only up to a sufficiently high order are needed to reconstruct an object up to its boundary and this provides an alternative way to determine the number of views given by Eq. (49) that is obtained from sampling relations.

The time-domain approach presented in this paper is based on the temporal Fourier transform of fixed frequency reconstructions. The latter are low-pass filtered versions of the true potential. The cutoff of this filter is dependent on

temporal frequency and results in an object that is time dependent. The demodulation procedure given here successfully decouples this time dependence as well as that resulting from the shape of the incident pulse.

A pulse waveform with a smooth temporal spectrum in the time-domain eigenfunction method results in a correspondingly smooth k-space window that produces less ringing than the discontinuous k-space window of the single-frequency eigenfunction method. This difference is evident in a comparison of the upper-right panel of Fig. 5 that shows a time-domain reconstruction using a k-space window determined by a Gaussian incident pulse and the lower-right panel of Fig. 4 that shows a single-frequency reconstruction. The k-space window in the time-domain method can also be modified, as already noted, by inclusion of a spatial-frequency weight during reconstruction. The effect of a radial Blackman k-space window, for example, is evident in a comparison of the lower-left and upper-right panels in Fig. 5 in which the Blackman window and pulse spectrum approaches, respectively, were used.

The difference between using $\gamma(\mathbf{k},t)$ rather than the incident pulse waveform $g(t)$ in the demodulation is only the k-space window associated with each. The result of using one reference can, therefore, theoretically be converted to the result of using the other reference. The incident waveform $g(t)$ was chosen for the demodulation reference in the time-domain reconstructions reported here for four reasons. First, the incident pulse waveform yields a smoother k-space window that produces less ringing around discontinuities in the scattering potential. Second, noise associated with high spatial freqencies is accentuated when $\gamma(\mathbf{k},t)$ is used as a demodulation reference. Third, the specification of a cutoff frequency is not necessary because the scattering potential as a function of $\mathbf{k}$ naturally decays to zero when the magnitude of $\mathbf{k}$ becomes large. Fourth, less computation is required when $g(t)$ is used because the demodulation is implemented directly in the space domain without the need to transfer back and forth from k space using a two-dimensional Fourier transform.

An adaptively positioned reference for demodulation improves the reconstruction by substantially reducing the droop or sag that appears in the central region of either high-contrast or large objects reconstructed using the Born approximation. This extends the range of parameters over which a Born inversion yields a scattering potential close to the true value. The improvement obtained using an adaptive reference is evident in a comparison of the adaptive demodulation result with the nonadaptive result in the lower right and left panels, respectively, of Fig. 5. Other reconstructions made from calculated data show that the adaptive method in this paper works well over a range of scattering object parameters, for example, for 6-mm-diameter objects with sound-speed contrast up to 4% and for objects with a sound-speed contrast of 1% and a diameter up to about 20 mm.

The use of a probe beam to determine sound-speed changes caused by temperature variations during the acquisition of experimental data significantly reduces errors when measurements are made over an extended period and the sound-speed contrast is small as in the cases reported here. Although the temperature is controlled in the ring-transducer system to about $\pm 0.2\,^{\circ}\mathrm{C}$ around $30\,^{\circ}\mathrm{C}$, this fluctuation in the temperature causes a sound-speed change of 0.0009 mm/μs[36] that produces a 0.06-μs change in the arrival time of the measured incident wavefront. Such a change is not negligible in image reconstruction, particularly for a small or low-contrast object. For example, a 1% contrast in a 3-mm-radius cylinder causes an arrival time shift of less than 0.04 μs in the incident wavefront in the forward direction. Residual arrival-time fluctuations due to temperature changes in the described experiments are estimated to be less than 0.001 μs after correction for the measured sound-speed variations.

The reconstructions of the various scattering objects presented in this paper deserve further comment. First, the homogeneous cylinder reconstructed in Fig. 5 and the cylinder that contains three strongly scattering filaments and randomly distributed weakly scattering small cylinders reconstructed in Fig. 6, each from essentially exact calculated data, are realistic examples that show the time-domain reconstruction method developed in this paper for use with experimental data performs well in an ideal situation. Second, the representative measured data plotted in Fig. 7 using bipolar log scales up to 80 dB indicate that a wide dynamic range of experimental data has been successfully collected. Third, the images and profiles in Fig. 8 demonstrate the ability to reconstruct sound speed with good fidelity from measured data because independent travel-time measurements of average sound-speed contrast in the cylinders indicate the error in the profiles shown is less than 10%. (Independent measurement of the density indicates the assumption that the cylinder density is equal to the background density was accurate to within 10%.) Fourth, the fluctuations in the interior sound speed in the cylinders that contain glass beads are a composite effect because the glass beads are too small and too closely spaced to resolve.

Other sources of error in reconstructions from experimental data are nonuniformities in the spatial amplitude of the incident plane wave in any one view and also from view to view, limited precision in the subtraction of the incident field from the total field in the forward direction, additive noise, and crosstalk between ring-transducer elements. Incident field nonidealities were kept small in the measurements reported here by a ring-transducer calibration process that reduced arrival-time fluctuations from noncircularity of the ring to less than 2 ns rms and also reduced the rms variation in amplitude sensitivity to 0.1 dB. Errors in the subtraction of the incident field to obtain the scattered field were controlled by instrument calibration as well as by the use of adaptive gain and a spatially limited incident wave, but in the forward direction subtraction of two large quantities inherently limited resolution for the weak scattering objects studied. The signal-to-noise ratio in the measurements typically was about 40 dB because adaptive variation of transmit signal amplitude and receiver gain generally kept the received signal in the center of the A/D converter range and the noise figure of the receiver is 4.3 dB. Acoustic crosstalk between receiver elements was generally less than 36 dB and, therefore, is not considered to be an important source of error except near the forward direction, where the adaptive gain

process boosts gain for received weak signals on elements near elements receiving the strong incident wave signal. Nevertheless, the images in the present study show these errors do not appear to degrade the reconstructions significantly.

V. CONCLUSION

A quantitative imaging method that uses eigenfunctions of the scattering operator has been extended to the time domain. The time-domain solution has a smoother k-space window than the single-frequency solution. The k-space window in the time-domain solution was improved to eliminate ringing in the point-spread function of the reconstruction process. Calculations for radial scattering objects showed that energy in the retransmitted fields advances to a ring, the diameter of which depends on the highest frequency in the incident pulse, and increases as the order of the retransmitted field increases. Calculations also showed that the frequency-domain reconstruction of a radial scattering potential progresses from the center outward as the order of the retransmitted field increases. An adaptive demodulation that compensates for the effect of the sound-speed variations in the inhomogeneity was demonstrated to reduce the reconstruction error that results from the linearization of the Born approximation and to extend the range over which the approximation yields a scattering potential close to the true value. The time-domain eigenfunction method of reconstruction was applied to image 6-mm-diameter scattering objects using measured data obtained with a unique ring-transducer system and a probe beam to track the background sound-speed changes due to temperature. For the reconstruction, the measured field was extrapolated to the far field using a temporal–spatial filter and an incident pulse for demodulation was obtained by averaging the incident fields that were backpropagated to the center of the measurement ring. The results show the ability of the method to image scattering objects of varying complexity with excellent fidelity in which the point resolution is about one-half of the wavelength at center frequency of the pulse.

ACKNOWLEDGMENTS

The authors thank Dr. Lawrence J. Busse for help in the development of data acquisition software, Dr. Daniel B. Phillips for help with the measurements using the ring-transducer apparatus, Dr. Carsten G. Draeger for his implementation of an exact solution for scattering from multiple nonconcentric cylinders, and Dr. Makoto Tabei for providing time-delay estimation code. This research was funded by NIH Grants Nos. DK 45533, HL 50855, and CA 74050, DARPA Grant No. N00014-96-0749, and the University of Rochester Diagnostic Ultrasound Research Laboratory Industrial Associates.

[1] G. H. Glover and J. L. Sharp, ''Reconstruction of Ultrasound Propagation Speed Distribution in Soft Tissue: Time-of-Flight Tomography,'' IEEE Trans. Sonics Ultrason. **24**(4), 229–234 (1977).
[2] J. F. Greenleaf, S. A. Johnson, and A. H. Lent, ''Measurement of Spatial Distribution of Refractive Index in Tissues by Ultrasonic Computer Assisted Tomography,'' Ultrasound Med. Biol. **3**(4), 327–339 (1978).
[3] R. K. Mueller, M. Kaveh, and G. Wade, ''Reconstructive Tomography and Applications to Ultrasonics,'' Proc. IEEE **67**(4), 567–587 (1979).
[4] A. C. Kak, ''Computerized Tomography with X-Ray Emission and Ultrasound Sources,'' Proc. IEEE **67**(9), 1245–1272 (1979).
[5] F. Stenger and S. A. Johnson, ''Ultrasonic Transmission Tomography Based on the Inversion of the Helmholtz Equation for Plane and Spherical Wave Insonification,'' Appl. Math. Notes **4**(3–4), 102–127 (1979).
[6] S. J. Norton and M. Linzer, ''Ultrasonic Reflectivity Imaging in Three Dimensions: Exact Inverse Scattering Solutions for Plane, Cylindrical, and Spherical Apertures,'' IEEE Trans. Biomed. Eng. **28**(2), 202–220 (1981).
[7] A. J. Devaney, ''A Filtered Backpropagation Algorithm for Diffraction Tomography,'' Ultrason. Imaging **4**(4), 336–350 (1982).
[8] S. X. Pan and A. C. Kak, ''A Computational Study of Reconstruction Algorithms for Diffraction Tomography: Interpolation versus Filtered Backpropagation,'' IEEE Trans. Acoust., Speech, Signal Process. **31**(5), 1262–1275 (1983).
[9] J. F. Greenleaf, ''Computerized Tomography with Ultrasound,'' Proc. IEEE **71**(3), 330–337 (1983).
[10] S. J. Norton, ''Generation of Separate Density and Compressibility Images in Tissues,'' Ultrason. Imaging **5**(3), 240–252 (1983).
[11] A. Witten, J. Tuggle, and R. C. Waag, ''A Practical Approach to Ultrasonic Imaging using Diffraction Tomography,'' J. Acoust. Soc. Am. **83**, 1645–1652 (1988).
[12] A. Witten, J. Tuggle, and R. C. Waag, ''Ultrasonic Imaging with a Fixed Instrument Configuration,'' Appl. Phys. Lett. **53**(1), 16–18 (1988).
[13] R. J. Wombell and M. A. Fiddy, ''Acoustical Imaging Beyond Born and Rytov,'' *Acoustical Imaging*, edited by L. W. Kessler (Plenum, New York, 1988), Vol. 16, pp. 373–381.
[14] T. J. Cavicchi and W. D. O'Brien, ''Numerical Study of Higher-Order Diffraction Tomography via the Sinc Basis Moment Method,'' Ultrason. Imaging **11**(1), 42–74 (1989).
[15] D. T. Borup, S. A. Johnson, W. W. Kim, and M. J. Berggren, ''Nonperturbative Diffraction Tomography via Gauss–Newton Iteration Applied to the Scattering Integral Equation,'' Ultrason. Imaging **14**(1), 69–85 (1992).
[16] J. Y. Zhang, ''Advanced Ultrasound Tomography Based on Generalized Born Approximation,'' Doctoral thesis, University of Utah (1994).
[17] H. Gan, R. Ludwig, and P. L. Levin, ''Nonlinear Diffractive Inverse Scattering for Multiple Scattering in Inhomogeneous Acoustic Background Media,'' J. Acoust. Soc. Am. **97**, 764–776 (1995).
[18] T. D. Mast, R. C. Waag, and A. I. Nachman, ''Eigenfunction and Eigenvalue Analysis of Scattering Operators,'' J. Acoust. Soc. Am. **96**, 3336(A) (1994).
[19] T. D. Mast, A. I. Nachman, and R. C. Waag, ''Focusing and Imaging using Eigenfunctions of the Scattering Operator,'' J. Acoust. Soc. Am. **102**, 715–725 (1997).
[20] R. C. Waag, D.-L. Liu, T. D. Mast, A. I. Nachman, P. Jaeger, and T. Kojima, ''An Ultrasonic Ring Transducer System for Studies of Scattering and Imaging,'' J. Acoust. Soc. Am. **100**, 2795(A) (1996).
[21] R. C. Waag, D.-L. Liu, T. D. Mast, A. I. Nachman, P. Jaeger, and T. Kojima, ''An Ultrasonic Ring Transducer System for Studies of Scattering and Imaging,'' in *41st Annual Convention, American Institute of Ultrasound in Medicine*, 23–26 March 1997, San Diego, CA [J. Ultrasound Med. **16**, No. 3 (Suppl.) S61 (1997)].
[22] R. C. Waag, ''An Ultrasonic Ring Transducer System for Studies of Scattering and Imaging,'' in *Report DURL 97-1* (Diagnostic Ultrasound Research Laboratory, University of Rochester, Rochester, NY, March 1997).
[23] F. Lin, ''Ultrasonic High Resolution and Quantitative Imaging with a Ring Transducer System,'' *Report DURL 99-1* (Diagnostic Ultrasound Laboratory, University of Rochester, Rochester, NY, February 1999).
[24] T. D. Mast, ''Wideband Quantitative Ultrasonic Imaging by Time-Domain Diffraction Tomography,'' J. Acoust. Soc. Am. **106**, 3061–3071 (1999).
[25] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), Chap. 8.
[26] T. D. Mast, A. I. Nachman, D.-L. Liu, and R. C. Waag, ''Quantitative Imaging with Eigenfunctions of the Scattering Operator,'' in *1997 IEEE International Ultrasonics Symposium Proceedings*, Toronto, Canada, IEEE Cat. No. 97CH36118-2, pp. 1507–1510.
[27] R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra* (Dover, New York, 1958), Part II, B.5.
[28] C. G. Draeger, ''Calculation of Scattering by Multiple Nonconcentric Cylinders,'' *Report DURL 99-2* (Diagnostic Ultrasound Research Laboratory, University of Rochester, Rochester, NY, April 1999).
[29] L. P. Souriau, T. D. Mast, D.-L. Liu, A. I. Nachman, and R. C. Waag, ''A *K*-Space Method for Large-Scale Models of Wave Propagation in Tissue,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control (submitted).

[30] G. Wojcik, B. Fornberg, R. C. Waag, L. Carcione, J. Mould, L. Nikodym, and T. Driscoll, ''Pseudospectral Methods for Large-Scale Bioacoustic Models,'' in *1997 IEEE International Ultrasonics Symposium Proceedings*, Toronto, Canada, IEEE Cat. No. 97CH36118-2, pp. 1501–1506.
[31] M. M. Burlew, E. L. Madsen, J. A. Zagzebski, R. A. Banjavic, and S. W. Sum, ''A New Ultrasound Tissue-Equivalent Material,'' Radiology **134**(2), 517–520 (1980).
[32] T. T. Jansson, T. D. Mast, and R. C. Waag, ''Measurements of Differential Scattering Cross Section using a Ring Transducer,'' J. Acoust. Soc. Am. **103**, 3169–3179 (1998).
[33] D.-L. Liu and R. C. Waag, ''Propagation and Backpropagation for Ultrasonic Wavefront Design,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**(1), 1–13 (1997).
[34] S. Wang, ''Finite-Difference Time-Domain Approach to Underwater Acoustic Scattering Problems,'' J. Acoust. Soc. Am. **99**, 1924–1931 (1996).
[35] Y. Chen, ''Inverse Scattering via Skin Effect,'' Inverse Probl. **13**(3), 647–667 (1997).
[36] N. Bilaniuk and G. S. K. Wong, ''Speed of Sound in Pure Water as a Function of Temperature,'' J. Acoust. Soc. Am. **93**, 1609–1612 (1993).

Scattering of antiplane shear waves by a circular cylinder in a traction-free plate

Xiao-Min Wang, C. F. Ying, and Ming-Xuan Li
Institute of Acoustics, Academia Sinica, P.O. Box 2712, Beijing 100080, People's Republic of China

(Received 28 October 1999; revised 8 May 2000; accepted 19 May 2000)

Following a well-established formula used by many researchers, the scattering of an anti-plane shear wave by an infinite elastic cylinder of arbitrary relative radius centered in a traction-free two-dimensional isotropic plate has been examined. The plate is divided into three regions by introducing two imaginary planes located symmetrically away from the surface of the cylinder and perpendicular to surfaces of the plate. The wave field is expanded into cylinder wave modes in the central bounded region containing the cylinder, while the fields in the other two outer regions are expanded into plate wave modes. A system of equations determining the expansion coefficients is obtained according to the traction-free boundary conditions on the plate walls and the stress and displacement continuity conditions across the imaginary planes. By taking an appropriate finite number of terms of the infinite expansion series and a few selected points on the two properly chosen virtual planes and the surfaces of the plate through convergence and precision tests, a matrix equation to numerically evaluate the expansion coefficients is found. The method of how to choose the locations of the imaginary planes and the terms of the expansion series as well as the points on each respective boundary is given in Sec. III in detail. Curves of the reflection and transmission coefficients against the relative radius of the cylinder in welded and slip or cracked interfacial conditions are shown. Analysis on the contrast variations of the reflection and transmission coefficients for a cylinder in bonded and debonded interfacial situations is made. The relative errors estimated by the deviation of the numerical results from the principle of the conservation of energy are found to be less than 2%. [S0001-4966(00)00809-2]

PACS numbers: 43.20.Fn [DEC]

INTRODUCTION

Much work has been done on the study of wave scattering from cylindrical objects embedded in an unbounded elastic medium since the 1950s.[1–14] A comprehensive list of early publications can be found in Ref. 14. The increasing prominence of fibrous composite materials has excited a new interest in these acoustic studies. In general, the fiber-reinforced materials are composed of a matrix and cylindrical fibers. As the interface transfers load from the fiber to the matrix, the overall mechanical performance of the fibrous composite materials is determined by the quality of the interface bonding of the fibers with the matrix. Ultrasonic scattering from such interfaces carries important information about the bonding state of the fibers, thus having potential for fiber-matrix interface characterization.

Consequently, more recent studies address multi-layered cylinder scattering.[11–18] A review of these studies is available in Ref. 17. In practice, a fibrous composite plate is fabricated by a stack of thin laminates reinforced by an array of fibers. Hence, a single solid layer containing a periodic array of elastic cylinders is the fundamental structure. The propagation and diffraction of elastic waves in such media deserves to be studied. Recently, Lu[19] analyzed the effect of periodically spaced cylinders on the dispersion relation of the guided antiplane shear wave in a layer, using the boundary element method based on the assumption that the layer thickness is much larger than the size of the reinforcement.

In the available literature, however, the scattering of guided waves by a single cylinder in a traction-free plate has not, to the authors' knowledge, been reported despite the extensive investigation of the scattering of elastic guided waves by the cracks and weldment in a plate.[20–28]

In the present paper, we examine the scattering of an antiplane shear wave in a traction-free two-dimensional isotropic plate by a single elastic cylinder of arbitrary relative radius embedded in the plate. An analytic solution is obtained and is approximately plotted. Our motivation to attack this problem is mainly to prepare for development along two directions. One is to look for a similar solution for the case of incidence in the plate of a Lamb wave of which we are not aware of any previous attempt. This search turns out to be more complicated than expected, but is near completion and will be summarized in a separate paper. The second direction of development will be to extend the problem to the scattering by an array of cylinders.

The theory developed here is based on the method of expanding the field into a sum of wave modes and on the method of mode matching. This mode matching technique was applied previously by Nilsen[29] and Lakhtakia *et al.*[30,31] to electromagnetic wave scattering and elastic bulk wave propagation problems, respectively. For numerical computation, truncation of the field expansion is applied, for which a convergence criterion is adopted for the choice of the number of retained terms, and an expression is formulated to estimate the precision of the numerical results.

THEORY

The plate considered is assumed to be two dimensional, of uniform thickness $2h$ with traction-free surfaces, and of density ρ and shear modulus μ. A Cartesian coordinate sys-

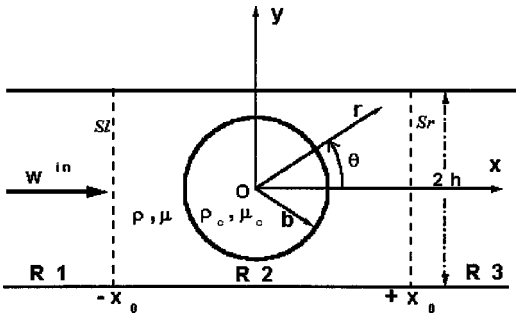


FIG. 1. Plate waveguide with an elastic cylindrical post.

tem is shown in Fig. 1, the origin of which is in the middle plane of the plate. An elastic circular cylinder of radius b is centered in the plate with its axis coinciding with the z-axis of the coordinate system. The material of the cylinder is taken to be of density ρ_c and shear modulus μ_c.

The waves are assumed to be time harmonic, but the complex time factor $e^{-j\omega t}$ (where ω is the angular frequency, t is the time, and $j=\sqrt{-1}$) will be omitted in all the following expressions. Let a single mode *SH*-type plate wave of order m traveling in the direction of positive x-axis and polarized along the z-axis impinge on the interface of the cylinder. For simplicity, the wave mode will be assumed to be symmetric. Antisymmetric mode incidence can easily be similarly formulated.

An incident shear wave of mode m can be expressed as

$$w^{in}=A_m\cos(\beta_m y)\exp(jk_m x), \tag{1}$$

where A_m is the amplitude of the given incident wave, $k_m=\sqrt{k^2-\beta_m^2}$ is the wave number of the incident mode, $\beta_m=m\pi/h$, and $k=\omega\sqrt{\rho/\mu}$ is the wave number for the shear wave in the infinite plate medium. The elastic wave displacement of all the scattering modes must be parallel to the z-axis since the cross section of the cylinder holds the same shape in the direction z. Thus the z components of the incident, the scattered, and the total particle displacements alone are nonvanishing.

On dividing the plate up into three regions R1, R2, and R3 by two imaginary planes *Sl* and *Sr* at $x=-x_0$ and $x=+x_0 (x_0>b)$, the respective fields in the two outer regions R1 and R3 can be expanded into the sum of all propagating and evanescent plate wave modes

$$w_1=w^{in}+\sum_{n=0}^{\infty} B_n\cos(\beta_n y)\exp(-jk_n x), \quad (x,y)\in R1, \tag{2}$$

$$w_3=\sum_{n=0}^{\infty} C_n\cos(\beta_n y)\exp(jk_n x), \quad (x,y)\in R3, \tag{3}$$

where B_n and C_n are the amplitudes of the reflected and transmitted waves. The fields in the center region R2 and the region inside the cylinder can be expanded into the sum of cylindrical wave modes:

$$w_2=\sum_{l=0}^{\infty}[D_l J_l(kr)+E_l Y_l(kr)]\cos(l\theta), \quad r>b, \tag{4}$$

$$w_c=\sum_{l=0}^{\infty} F_l J_l(k_c r)\cos(l\theta), \quad r<b, \tag{5}$$

where D_l, E_l, and F_l are complex expansion coefficients and $J_l(.)$ and $Y_l(.)$ are the Bessel and Neuman functions of the lth order. In the arguments (kr) and $(k_c r)$ of the cylinder functions, r is the distance from the cylinder axis to the field point, and k_c is the complex wave number in the medium of the cylinder. The reason to select cosine functions only in Eqs. (4) and (5) is that the wave fields in region R2 and the region inside the cylinder are totally symmetrical[29] about the xOz-plane.

We now come to the various boundary conditions. The boundary conditions in the present problem can be divided into two groups: the group concerned with the plane interfaces shown in Fig. 1 and the group concerned with the curved interface between the cylinder and the matrix.

Boundary conditions of the first group are given by

$$\frac{\partial w_2}{\partial y}=0, \quad y=\pm h, \tag{6}$$

$$w_1=w_2, \quad x=-x_0, \tag{7}$$

$$\frac{\partial w_1}{\partial x}=\frac{\partial w_2}{\partial x}, \quad x=-x_0, \tag{8}$$

$$w_2=w_3, \quad x=+x_0, \tag{9}$$

$$\frac{\partial w_2}{\partial x}=\frac{\partial w_3}{\partial x}, \quad x=+x_0. \tag{10}$$

The boundary condition of the second group may vary with the actual situation and so we shall consider several possible cases.

Welded interface

This is the usual boundary condition. Here,

$$w_2^b=w_c^b, \quad \sigma_2^b=\sigma_c^b, \tag{11}$$

where $w_2^b=w_2|_{r=b}$, etc.; $\sigma_2=\sigma_{rz,2}=\mu(\partial w_2/\partial r)$.

From Eqs. (4) and (5) we get

$$\sigma_2=\mu k\sum_{l=0}^{\infty}[D_l J_l'(kr)+E_l Y_l'(kr)]\cos(l\theta), \tag{12}$$

$$\sigma_c=\mu_c k_c\sum_{l=0}^{\infty} F_l J_l'(k_c r)\cos(l\theta). \tag{13}$$

The welded boundary conditions Eqs. (11) then lead to

$$\sum_{l=0}^{\infty}[D_l J_l(kb)+E_l Y_l(kb)]\cos(l\theta)=\sum_{l=0}^{\infty} F_l J_l(k_c b)\cos(l\theta), \tag{14}$$

$$\sum_{l=0}^{\infty}\mu k[D_l J_l'(kb)+E_l Y_l'(kb)]\cos(l\theta) =\sum_{l=0}^{\infty}\mu_c k_c F_l J_l'(k_c b)\cos(l\theta). \tag{15}$$

So

$$E_l = -D_l \frac{\mu k J_l(k_c b) J_l'(kb) - \mu_c k_c J_l'(k_c b) J_l(kb)}{\mu k J_l(k_c b) Y_l'(kb) - \mu_c k_c J_l'(k_c b) Y_l(kb)}, \tag{16}$$

and consequently

$$w_2 = \sum_{l=0}^{\infty} D_l \Bigg[J_l(kr) - Y_l(kr) \times \frac{\mu k J_l(k_c b) J_l'(kb) - \mu_c k_c J_l'(k_c b) J_l(kb)}{\mu k J_l(k_c b) Y_l'(kb) - \mu_c k_c J_l'(k_c b) Y_l(kb)} \Bigg] \cos(l\theta). \tag{17}$$

Similarly, using Eqs. (14) and (15), F_l can be expressed in terms of D_l and hence w_c also can be.

Slip boundary condition (=cracked interface=cavity)

Generally, the slip boundary condition implies continuity of the normal components of stress and displacement across the interface, vanishing of the shear component of stress. For the antiplane shear wave, particularly, there are no normal components of stress and displacement across the interface, and the slip boundary condition implies vanishing of the shear component of stress which gives

$$\sigma_2^b = \mu \left. \frac{\partial w_2}{\partial r} \right|_{r=b} = 0, \quad \sigma_c^b = \mu_c \left. \frac{\partial w_c}{\partial r} \right|_{r=b} = 0, \tag{18}$$

the second equation of which leads to F_l in $w_c = 0$. Consequently, for this case $w_c = 0$. This result is expected since now the incident wave cannot cross the interface and hence will not enter the cylinder. The situation is equivalent to the interface cracked over the complete circumference of the cylinder, and therefore is also equivalent to a cavity.

Presently,

$$E_l = -D_l \frac{J_l'(kb)}{Y_l'(kb)} \tag{19}$$

and

$$w_2 = \sum_{l=0}^{\infty} D_l \left[J_l(kr) - Y_l(kr) \frac{J_l'(kb)}{Y_l'(kb)} \right] \cos(l\theta). \tag{20}$$

Partly cracked interface

In addition to the case of completely cracked interface, we have also considered the case of the partly cracked interface (cracked over an angular width θ_c, $0 < \theta_c < 2\pi$). Owing to limited space, we shall not include the detailed discussion here. The basic principle to deal with this novel problem is to expand the crack-opening-displacement (COD) jump into a series of Chebyshev polynomials which have the correct square root behavior at the crack edge.[11] On the traction-free boundary condition within the angular width θ_c of the partial crack and the orthogonality of Chebyshev polynomials, a equation with respect to the expansion coefficients can be found. These coefficients are closely related with the expansion coefficients of the cylindrical wave modes by the boundary conditions on the surface of the cylinder, and are also related with the expansion coefficients of guided wave modes by the boundary conditions across the virtual planes. Thus a uniform system of equations can be formed to obtain all of those expansion coefficients, and the wave fields in each region can be determined.

The boundary conditions between the cylinder and the matrix are discussed further in the Appendix in terms of a thin interfacial layer instead of an interface. Consistent results are obtained for the simple limiting cases.

The second group of boundary conditions having delineated E_l and F_l in terms of D_l; the only expansion coefficients remain to be determined are B_n, C_n and D_l. These coefficients can be evaluated by using the first group of boundary conditions, Eqs. (6)–(10). Since the fields are all symmetrical about the xOz plane, only half of the plate corresponding to, say, $y > 0$, has to be taken into account. Equation (6) then will be evaluated only for $y = +h$. The set of linear equations for determining B_n, C_n, and D_l can now be shown to be, corresponding, respectively, to Eqs. (9), (10), (6), (7), and (8);

$$\sum_{n=0}^{\infty} C_n \cos(\beta_n y_a) \exp(j k_n x_0) - \sum_{l=0}^{\infty} D_l Z_l(kr_a) \cos(l\theta_a) = 0, \tag{21}$$

$$\sum_{n=0}^{\infty} C_n k_n \cos(\beta_n y_a) \exp(j k_n x_0) + j \sum_{l=0}^{\infty} D_l \Phi_l(k, r_a, \theta_a) = 0, \quad \left(y_a = x_0 \tan\theta_a, \ r_a = \frac{x_0}{\cos\theta_a}, \ 0 < \theta_a \leqslant \tan^{-1}\left(\frac{h}{x_0}\right) \right), \tag{22}$$

$$\sum_{l=0}^{\infty} D_l \Psi_l(k, r_b, \theta_b) = 0, \quad \left(r_b = \frac{h}{\sin\theta_b}, \tan^{-1}\left(\frac{h}{x_0}\right) \leqslant \theta_b \leqslant \pi - \tan^{-1}\left(\frac{h}{x_0}\right) \right), \tag{23}$$

$$\sum_{n=0}^{\infty} B_n \cos(\beta_n y_c) \exp(j k_n x_0) - \sum_{l=0}^{\infty} D_l Z_l(kr_c) \cos(l\theta_c) = -A_m \cos(\beta_m y_c) \exp(-j k_m x_0), \tag{24}$$

$$\sum_{n=0}^{\infty} B_n k_n \cos(\beta_n y_c) \exp(j k_n x_0) - j \sum_{l=0}^{\infty} D_l \Phi_l(k, r_c, \theta_c) = A_m \alpha_m \cos(\beta_m y_c) \exp(-j k_m x_0), \quad \left(y_c = -x_0 \tan\theta_c, \quad r_c = \frac{-x_0}{\cos\theta_c}, \quad \pi - \tan^{-1}\left(\frac{h}{x_0}\right) < \theta_c \leqslant \pi \right). \tag{25}$$

In Eqs. (21)–(25), the function $Z_l(\xi)$ is defined to be

$$Z_l(\xi) = J_l(\xi) + \alpha_l Y_l(\xi), \tag{26}$$

its derivative $Z_l'(\xi)$ being

$$Z_l'(\xi)=J_l'(\xi)+\alpha_l Y_l'(\xi), \tag{27}$$

where α_l is a constant related to the frequency of the incident wave, the size and the material of the cylinder, and the material of the plate:

$$\alpha_l=-\frac{\mu k J_l(k_c b)J_l'(kb)-\mu_c k_c J_l'(k_c b)J_l(kb)}{\mu k J_l(k_c b)Y_l'(kb)-\mu_c k_c J_l'(k_c b)Y_l(kb)}$$

$$\text{(for welded interface)}, \tag{28}$$

or

$$\alpha_l=-\frac{J_l'(kb)}{Y_l'(kb)} \quad \text{(for slip interface)}. \tag{29}$$

Now

$$w_2=\sum_{l=0}^{\infty} D_l Z_l(kr)\cos(l\theta), \tag{30}$$

$$\sigma_2=\mu k\sum_{l=0}^{\infty} D_l Z_l'(kr)\cos(l\theta). \tag{31}$$

The functions $\Phi_l(k,r,\theta)$ and $\Psi_l(k,r,\theta)$ in Eqs. (22), (23), and (25) are defined, respectively, as

$$\Phi_l(k,r,\theta)=kZ_l'(kr)\cos(l\theta)\cos\theta + Z_l(kr)\frac{l}{r}\sin(l\theta)\sin\theta, \tag{32}$$

$$\Psi_l(k,r,\theta)=kZ_l'(kr)\cos(l\theta)\sin\theta - Z_l(kr)\frac{l}{r}\sin(l\theta)\cos\theta. \tag{33}$$

We have now apparently solved our problem by finding out the wave functions in all regions of the plate: R1, R2, R3, and the region within the cylinder. To repeat, w_1, w_2, and w_3 are given, respectively, by Eq. (2), Eq. (30), and Eq. (3), in terms of, respectively, B_n, D_l, and C_n, together with the known A_m, while w_c is given by Eq. (5) in terms of F_l which can be determined from D_l by the use of Eqs. (14) and (15). The unknown coefficients B_n, C_n, and D_l are given by Eqs. (21)–(25) for two possible kinds of interfacing between the cylinder and the matrix.

From these expansion coefficients, two field quantities of practical interest, because they are more easily measurable, may be defined and deduced. They are:

$R_n^m=$ reflection coefficient of the nth scattered mode for the mth incident mode:

$$=B_n/A_m, \tag{34}$$

$T_n^m=$ transmission coefficient of the same

$$=C_n/A_m, \tag{35}$$

both holding in the region $|x|\geqslant x_0$. The two field quantities are numerically illustrated in a later section.

APPROXIMATE EVALUATION OF THE EXPANSION COEFFICIENTS

The expansion coefficients B_n, C_n, and D_l are defined by Eqs. (21)–(25). Yet their practical evaluation is difficult because the plate wave modes and the cylinder wave modes are infinite in number, that is, n, $l\in[0,+\infty)$. In addition, the locations of the virtual planes at $x=\pm x_0$ in Fig. 1 are still undetermined up to now. For the infinite expansion series, some approximation is required and one way is to truncate them somewhere such that $n\in[0,N]$, $l\in[0,L]$, N and L being some finite positive integers. As for the distance x_0 from the center of the cylinder to the virtual planes, a reasonable initial value satisfying the condition $x_0>b$ is needed to be set. Then, a system of linear equations to obtain the expansion coefficients of the truncated series has to be formed in some way when these three crucial parameters, N, L, and x_0, are determined. The question is then raised as to how to choose values of N, L, and x_0 such that relatively small errors in the evaluation of the expansion coefficients will be assured.

It turns out that before the demand on precision, another demand on convergence has to be satisfied in choosing N, L, and x_0. During our numerical computation we find that certain choices of these three parameters will induce divergence and hence evidently have to be avoided. The details of how to choose N, L, and x_0, the way to form the system of linear equations, and our approach to check the accuracy of our numerical results once N, L, and x_0 have been chosen are as follows.

The values of the three parameters N, L, and x_0 are mutually related. Physically, the scattering near the surface of the scatter may include more evanescent modes with sufficient amplitude to cause near-field superposition effects on the wave fields on the imaginary boundaries when the virtual planes are close to the surface of the cylinder. On the contrary, if the boundaries at $x=\pm x_0$ are located away from the surface of the cylinder, the evanescent higher modes due to the scattering of the post have little effect on the wave fields along the virtual planes. In this case, the influence of the evanescent higher modes is involved in region 2 pertinent with the cylinder wave mode expansion. Mathematically, this physical interpretation is given by the values of the terms $\exp(-k_n^e x_0)$ and $Y_l(k\sqrt{x_0^2+y^2})$ in the plate and cylinder wave mode expansions, respectively. Here, k_n^e is the wave number of the evanescent plate wave mode of order n, k is the bulk shear wave number of the plate, and y is the position of a geometrical point on the virtual planes.

It seems that the larger the value of x_0, the smaller the values of N and L needed in the numerical evaluation, and thus the more stable the evaluation procedure. One may intend to choose a value of x_0 large enough to account for the effects of the far-field and the propagating modes only on the wave fields along the imaginary boundaries. However, the finite thickness of the plate prevents such an intention, because the traction-free boundary conditions on the surfaces of the plate implied in Eq. (23) has to be satisfied with the same value of L determined for the sake of the imaginary boundaries. As a result, the thickness of the plate will certainly influence the selection of the values of N and x_0 for

these two parameters are mutually related with L.

Practically, one of the possible schemes to make the optimized selection of these three parameters N, L, and x_0 is to set a reasonable initial value for $x_0 > b$ approximately and then to search for the appropriate values of N and L so mutually matching with each other that all the boundary conditions implied in Eqs. (21)–(25) can be satisfied in a certain precision. This is the mode-matching technique in nature. Once the values of N and L are found, modifications for the initial approximate value of x_0 has to be made to improve the precision of the results in the case of the precision being unsatisfactory, especially when the diameter of the cylinder is close to the thickness of the plate.

On account of the finite thickness of the plate and the symmetry of the geometrical configuration, since the locations of the virtual planes may be selected arbitrarily as long as condition $x_0 > b$ is satisfied, the reasonable initial value for x_0 is approximately, but not necessarily, equal to half the thickness of the plate. To emphasize, this initial value of x_0 has to be modified on the convergence and precision tests.

To convince us of the analysis on the choice of the initial value for x_0, convergence tests are made for a guided SH-wave mode of order $m=1$ incident on a hard cylindrical post. To outline the basic principle, only the convergence test graphs for the hard cylinder of diameter being half the thickness of the plate are shown in Figs. 2(a) and 2(b) to indicate the variations of the amplitudes of the transmission coefficient $|T_1^1|$ and the relative errors ϵ vs x_0/h for $kh=1.5\pi$ and $kh=2.5\pi$, respectively. From Fig. 2(a), one sees that the transmission coefficients $|T_1^1|$ converge steadily to 0.59 in the range of $0.8 \leqslant x_0/h \leqslant 1.4$ for $kh=1.5\pi$ and 0.33 in the range of $0.89 \leqslant x_0/h \leqslant 1.33$ for $kh=2.5\pi$, respectively. In Fig. 2(b), the respective maximum relative error ε is 1.3% for $kh=1.5\pi$ when $0.8 \leqslant x_0/h \leqslant 1.4$ and 4.1% for $kh=2.5\pi$ when $0.89 \leqslant x_0/h \leqslant 1.33$. At $x_0/h=1$, the relative errors are 0.1% for $kh=1.5\pi$ and 0.15% for $kh=2.5\pi$, respectively. Figures 2(a) and 2(b) indicate that h may be chosen as one of the reasonable initial values for x_0, for a hard cylinder of radius being $0.5h$, especially when $kh=2.5\pi$.

Convergence tests for hard and soft cylinders and a cylindrical cavity of any radius less than $0.9h$ convince us that the choice of h as the initial value of x_0 leads us to obtain the converged solutions with the maximum relative errors ranging from 0.2% to 10.9% in different situations. However, for a large cylinder or a cylindrical cavity in radius, this initial value of $x_0=h$ has to be modified for different situations depending on the precision graphs similar to Fig. 2(b) to obtain the results with satisfactory precisions.

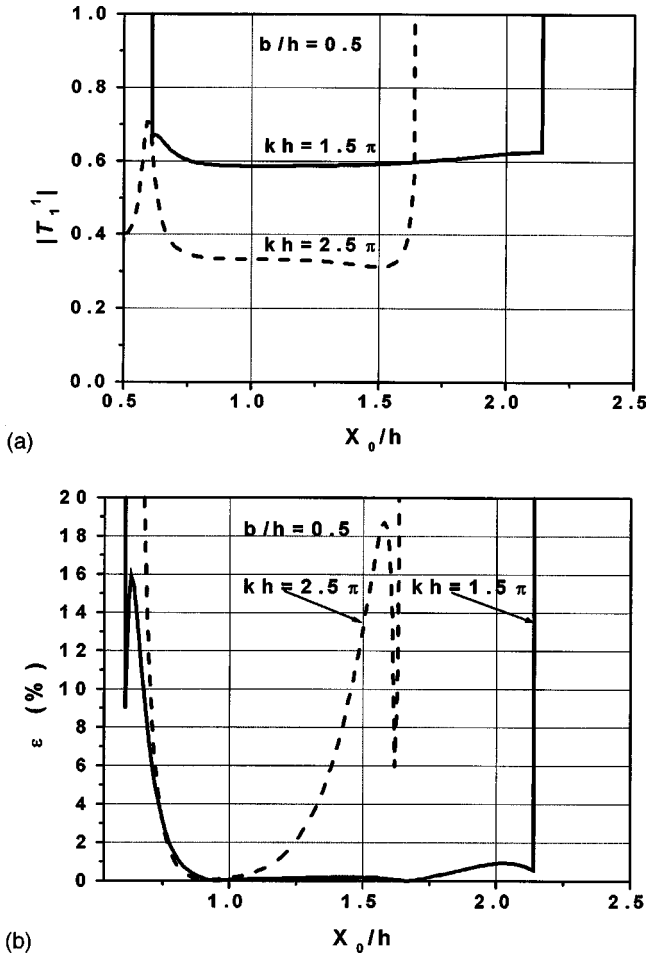


FIG. 2. Convergence test for the SH guided mode ($kh=1.5\pi, 2.5\pi$) of order 1 incidence to a hard cylinder: (a) $|T_1^1|$ vs x_0/h; (b) ϵ vs x_0/h.

To indicate how the modifications for the initial value of x_0 improve the precisions of the numerical results in different situations, in Table I we show $\varepsilon_{\text{initial}}$ denoting the maximum value of the relative errors with $x_0/h \equiv 1$ for $0.01 \leqslant b/h \leqslant 0.9$ and $\varepsilon_{\text{improved}}$ representing the maximum value of the relative errors with modified x_0/h when b/h is larger than a certain value (in the bracket). From Table I one sees clearly that the precisions are surely improved by modifications for the initial value of $x_0/h=1$. We should mention here, however, a particular situation of the soft cylinder for $kh=2.5\pi$ in Table I. We found that, although the precision

TABLE I. The maximum values of the relative errors with $x_0/h=1$ fixed and modified.

		Hard cylinder	Soft cylinder	Cavity
$kh=1.5\pi$	$\varepsilon_{\text{initial}}$	7.3%	0.2%	2.6%
$N=3$	$\varepsilon_{\text{improved}}$	1.1%	0.2%	0.6%
$L=16$	x_0/h	1.2	1.0	1.2
$M_p=5$		(when $b/h \geqslant 0.71$)		(when $b/h \geqslant 0.6$)
$kh=2.5\pi$	$\varepsilon_{\text{initial}}$	1.0%	10.9%	3.1%
$N=4$	$\epsilon_{\text{improved}}$	0.8%	3.2%	0.5%
$L=19$	x_0/h	1.1	1.25	1.2
$M_p=6$		(when $b/h \geqslant 0.85$)	(when $b/h \geqslant 0.83$)	(when $b/h \geqslant 0.65$)

is satisfactorily improved with the maximum relative error dropping to 3.2% from 10.9% by the modified $x_0/h=1.25$ for $b/h \geqslant 0.83$, a much higher precision with the maximum relative error being 1.9% can be achieved by choosing $N=5$, $L=22$ and $M_p=7$ with $x_0/h \equiv 1$ for $0.01 \leqslant b/h \leqslant 0.9$. It seems that the values of N and L have a more powerful influence on the precision than the modification of the initial value of x_0.

To determine a suitable value of N in the plate wave mode expansion, all the propagating and nonpropagating modes and an appropriate number of evanescent modes have to be included.[32] Thus $N=N_p+N_1+N_E$; here, N_p is the number of the propagating modes, N_1 denotes the nonpropagating mode, and N_E represents the finite number of evanescent modes considered. Larger N_E likely means better numerical precision, but it turns out that too large N_E will cause divergence in solving the matrix equation because we find that the evanescent modes of higher orders may cause singularity of the matrix. In our calculation, usually a pair of evanescent modes is included. For example, as $kh=1.5\pi$, $n=0$ and $n=1$ modes are propagating waves, while $n=2$ and $n=3$ are evanescent modes and no nonpropagating mode exists. In this case, we take $N=3$. For another example, as $kh=2.5\pi$, $n=0,1$ and 2 are propagating modes, $n=3$ and $n=4$ are evanescent modes. Here, we take $N=4$.

To determine the suitable number of L in the cylindrical mode expansion, it is necessary to make an auxiliary small program in order to examine the variations of the cylindrical functions with an increase of order l for different argument values. In this way, the number $l=L$ can be found through the test of convergence with the selected values of N, L, x_0, and Mp (see next paragraph). This procedure has the characteristic of the trial and error.

For chosen values of N, L, and x_0, there are $2N+L+3$ unknown coefficients B_n, C_n and D_l, $n \in [0,N]$, $l \in [0,L]$ to be determined. A system of linear equations needed to obtain these unknown coefficients may be formed by inner-product method or point-mate method. On account of the fact that Bessel and Neuman functions involved in cylinder wave mode expansion are high oscillatory functions, we prefer the point-mate technique. Based on the fact that any arbitrary point on the top surface of the plate has to satisfy Eq. (23), any arbitrary point on the upper half of the right virtual plane has to satisfy Eqs. (21)–(22), and any arbitrary point on the upper half of the left virtual plane has to satisfy Eqs. (24)–(25); a sufficient number of equations, $2N+L+3$ in number, needed for solving the problem can be found from Eqs. (21)–(25) by taking advantage of the arbitrariness of θ_a, θ_b, and θ_c. For example, if one takes a single point on each partition of the plate (*Sl*, *Sr* or the top of the plate section), five equations can be selected from Eqs. (21)–(25) by giving θ_a, θ_b, and θ_c a specific value, respectively; if two points are taken on each partition, ten equations can be selected, and so on. For the problem at hand, the number of the points needed on each partition is $\mathrm{Mp}=(2N+L+3)/5$. Let the chosen geometric points spread with equal angular spacing on each respective partition; then the corresponding values of θ_a, θ_b, and θ_c can subsequently be calculated. Substituting these values of N, L, Mp, x_0, θ_a, θ_b, and θ_c, and the values of the material parameters, the dimensionless quantity kh and the ratio b/h into Eqs. (21)–(25), a matrix equation necessary to solve the problem can be constructed.

The method to select a few of specific points, Mp in number, on each involved boundaries includes two steps when x_0/h, N, L, and hence Mp are known. First, determine the respective angular width $\vartheta_{\mathrm{LEFT}}$, $\vartheta_{\mathrm{RIGHT}}$, and ϑ_{TOP} of each pertinent side of the rectangle in Fig. 1, i.e., the angular width of the upper half of the left and right virtual planes and the angular width of the section of length $2x_0$ on top of the plate. From Fig. 1, one obtains $\vartheta_{\mathrm{LEFT}}=\vartheta_{\mathrm{RIGHT}}=\mathrm{ctg}^{-1}(x_0/h)$, $\vartheta_{\mathrm{TOP}}=(\pi-2\vartheta_{\mathrm{RIGHT}})$. Second, divide each angular width into $(\mathrm{Mp}-1)$ equal elements with one single element being $\Delta\vartheta_{\mathrm{LEFT}}=\Delta\vartheta_{\mathrm{RIGHT}}=[\mathrm{ctg}^{-1}(x_0/h)]/(\mathrm{Mp}-1)$ for $\vartheta_{\mathrm{LEFT}}$ and $\vartheta_{\mathrm{RIGHT}}$ and $\Delta\vartheta_{\mathrm{TOP}}=[\pi/(\mathrm{Mp}-1)-2\Delta\vartheta_{\mathrm{RIGHT}}]$ for ϑ_{TOP}. The points of intersection of the geometrical rays of polar angles $\theta_a=(I-1)\cdot\Delta\vartheta_{\mathrm{RIGHT}}$ on the right imaginary plane, $\theta_b=\vartheta_{\mathrm{RIGHT}}+(I-1)\cdot\Delta\vartheta_{\mathrm{TOP}}$ on the top surface of the plate and $\theta_c=(\pi-\vartheta_{\mathrm{LEFT}})+(I-1)\cdot\Delta\vartheta_{\mathrm{LEFT}}$ on the left virtual plane are the selected ones on each respective partition, here, $I=1,2,\ldots,\mathrm{Mp}$. All the selected points, Mp in number, at each side of the rectangle spread with equal angular spacing. They are dispersed in the whole range of each partition. Their coordinates (r_a,θ_a), (r_b,θ_b), and (r_c,θ_c) in the polar coordinate system can easily be obtained by the formulas of trigonometry. One may choose these points in a different way as long as the relative errors calculated by Eq. (39) in the final results are satisfactory.

To have some knowledge of how the selection of points affects the final results and what a satisfactory way of selection to obtain higher precision results is, tests have been made to compare the final results by two different selection ways. One is described in the above paragraph when $x_0/h \equiv 1$ with no modifications for x_0/h, the other is to let the points spread in the angularly central range of each pertinent side of the rectangle. Figure 3(a) shows, as an example, the curves of the reflection coefficients $|R_1^1|$ and the transmission coefficients $|T_1^1|$ against b/h for $kh=1.5\pi$ calculated in these two extremely different point selections. In Fig. 3(a), the dash-dot lines denote the results obtained by the first way, i.e., all the points on each boundary are dispersed in the whole part ($0 \leqslant \theta_a \leqslant \pi/4$, $\pi/4 \leqslant \theta_b \leqslant 3\pi/4$, and $3\pi/4 \leqslant \theta_c \leqslant \pi$) of each side, while the dash lines stand for the results obtained by the second selection way ($\pi/16 \leqslant \theta_a \leqslant 3\pi/16$, $3\pi/8 \leqslant \theta_b \leqslant 5\pi/8$, and $13\pi/16 \leqslant \theta_c \leqslant 15\pi/16$). Some differences between the final results appear when b/h is larger than 0.7 approximately. Which way is better? To answer this question, we also show, in Fig. 3(a), the results (solid lines) having higher precision known from Table I by the first point selection way but with modifications for x_0/h. One sees that the results (dash-dot lines) obtained by the first point selection way is closer to the higher precision results (solid lines). Figure 3(b) shows the relative error curves for the situations corresponding to Fig. 3(a). Figure 3(b) convinces us of that the way to select the points in the whole range of each partition is better than the way of part range selection.

To check the accuracy of the numerical technique, the

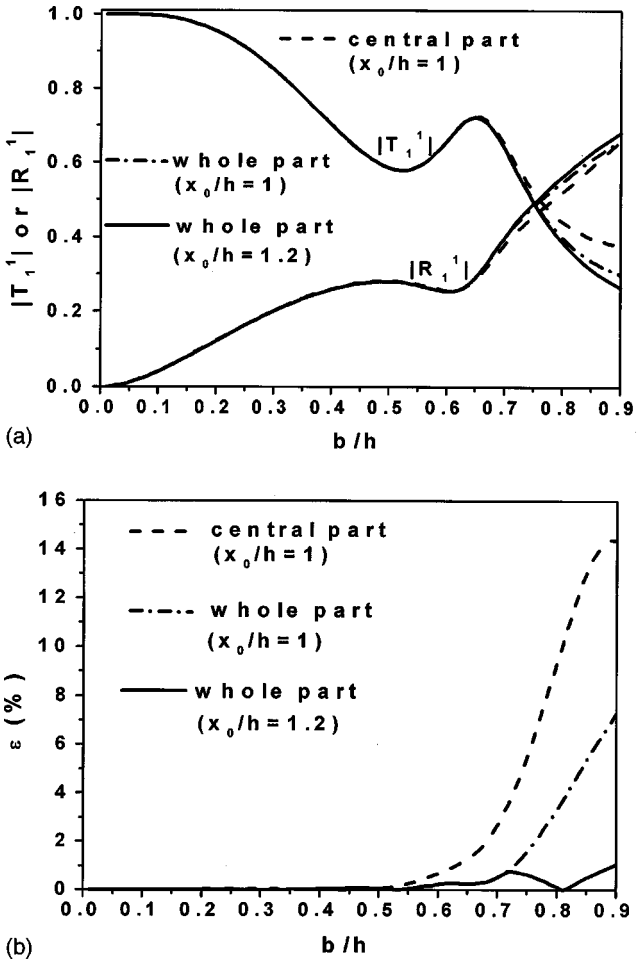


FIG. 3. Point selection test for the SH guided mode ($kh=1.5\pi$) of order 1 incidence to a hard cylinder: (a) $|T_1^1|$ or $|R_1^1|$ vs b/h; (b) ϵ vs b/h.

principle of conservation of energy is used. This states that[24,27]

$$P^{in}=\sum_{n=0}^{N_P}(P_n^r+P_n^t), \tag{36}$$

where P^{in} is the energy flux transported by the incident mode through the imaginary planes, P_n^r and P_n^t represent the energy fluxes transported by the reflected and transmitted nth mode through the imaginary planes, and N_p, as before, is the number of propagating modes. The transported energy fluxes in Eq. (36) are obtained from the equation[24]

$$P_n=-\frac{1}{2}\int_0^h \mathrm{Re}(\overline{\sigma_{xz,n}}\cdot\dot{w}_n)_{x=\pm x_0}\,dy=\frac{h}{4}\delta_n\mu\omega k_n|A_n|^2, \tag{37}$$

where $\delta_0=2$, $\delta_n=1$, and A_n represents the amplitude of the mode transporting the energy flux P_n, $\sigma_{xz,n}$, and w_n are the stress and displacement associated with the mode under consideration. The overbar and the overdot in Eq. (37) denote the complex conjugate and the time derivative, respectively. Now, the error involved in the numerical calculation can be estimated by the deviation from Eq. (36). Thus[27] the relative error of the reflected and transmitted energies from the input energy flux P^{in} is

$$\epsilon=1-\frac{\Sigma_{n=0}^{N_p}(P_n^r+P_n^t)}{P^{in}}. \tag{38}$$

On substitution of Eq. (37) into Eq. (38) and taking its absolute value, the relative error can be expressed as

$$\epsilon=\left|1-\frac{\Sigma_{n=0}^{N_p}\delta_n k_n(|R_n^m|^2+|T_n^m|^2)}{\delta_m k_m}\right|, \tag{39}$$

which is used to evaluate the consistency of the numerically computed results.

TABLE II. The material constants used for numerical calculations.

	Density (g/cm^3)	Shear modulus (GPa)
Matrix	2.7	26.0
Hard cylinder	1.9	147.0
Soft cylinder	2.7	13.1

TABLE III. Terms in the series summations taken in numerical calculations.

$2h/\lambda$	N	L	Mp
1.5	3	16	5
2.0	4	19	6
2.5	4	19	6
3.0	5	22	7

NUMERICAL IMPLEMENTATION

In this section, some numerical results are presented showing the coefficients of reflection and transmission as the functions of the ratio of the radius of the cylinder to the half thickness of the plate b/h for given value of kh/π. The relative errors are estimated for each case.

The material constants used for numerical calculations are shown in Table II. Both the hard cylinder (with higher velocity) and the soft cylinder (with lower velocity) compared to the matrix (aluminum) in the welded interface boundary condition and the slip or completely cracked interface boundary condition (cavity) are considered. We set the half thickness of the plate $h=1.0$ mm, and let the radius b of the cylinder vary from 0.01 to 0.9 mm. The incoming antiplane shear wave is a single propagating mode of order m

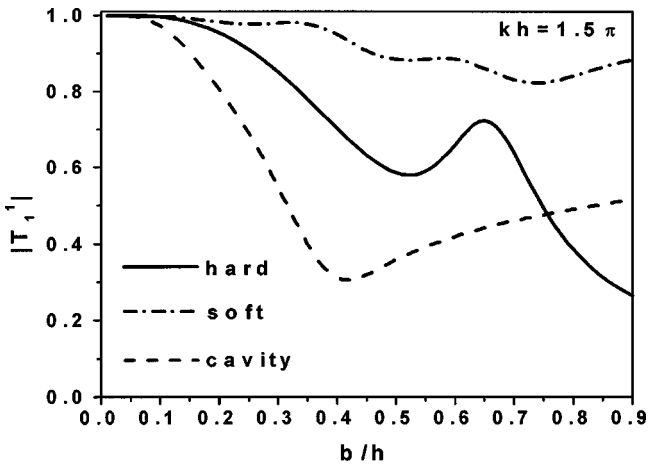


FIG. 4. Transmission coefficient $|T_1^1|$ vs normalized cylinder radius b/h for $kh=1.5\pi$.

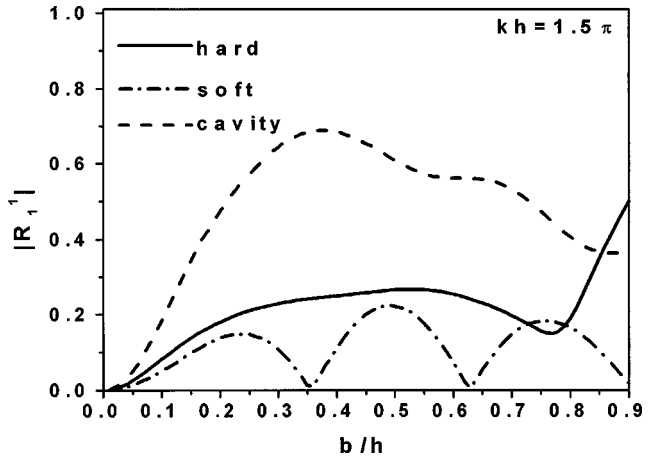


FIG. 5. Reflection coefficient $|R_1^1|$ vs normalized cylinder radius b/h for $kh=1.5\pi$.

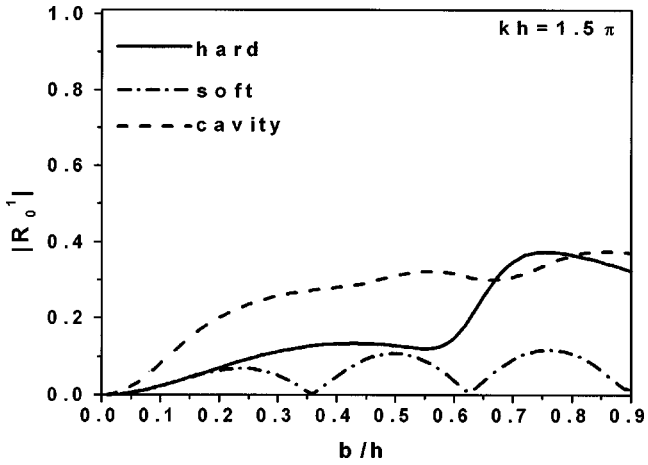


FIG. 7. Reflection coefficients $|R_0^1|$ vs normalized cylinder radius b/h for $kh=1.5\pi$.

$=1$ with unit amplitude. In this case, the condition $2h/\lambda > 1$ has to be satisfied, in which λ is the wavelength of the incident mode.

Table III gives the number of plate wave modes N and that of the cylindrical wave modes L, as well as the number of the geometric points Mp on each partition of the plate used in our calculations for some different values of $2h/\lambda$.

From Table III one sees that when the wavelength is fixed, the thicker the plate, the more modes of both the plate and cylindrical waves have to be taken into account and the more field points need to be used.

For simplicity, we only show the numerically calculated reflection and transmission coefficients of the prevalent propagating modes at the values of $kh=1.5\pi$ and 2.5π versus the normalized radius b/h for the welded and slip (cracked) boundary conditions in Figs. 4–10. For $kh=1.5\pi$, curves of the reflection and transmission coefficients are given in Figs. 4–7, while for $kh=2.5\pi$, only the transmission coefficient curves are shown in Figs. 8–10. The relative errors induced from truncation of the infinite series in these calculations and computed according to Eq. (39) are given in Figs. 11 and 12. In all of these figures the solid thick lines and the dash-dot lines represent the hard cylinder and the soft cylinder in boundary condition of welded interface, respectively, while the dash lines represent the boundary condition of slip or cracked interface (cavity).

The number of the prevalent scattered propagating modes depends on the value of kh, as may be seen from the expression for the mode wave number $k_n = k\sqrt{1-(n\pi/kh)^2}$; the larger the kh, the more the modes. Thus in Figs. 4–7, there are only two scattered propagating modes, while in Figs. 8–10, there are three.

From the figures it may be seen that when the radius of the cylinder is very small, $(b/h \leqslant 0.1)$, very little energy of the incident mode is scattered such that $T_1^1 \approx 1$ and R_1^1, T_n^1, $R_n^1 \approx 0$ $(n \neq 1)$, as expected. As the diameter of the cylinder increases, T_1^1 decreases but may rise again due to the reflection and complicated mode conversion from the scattering of the incident mode by the cylinder.

From Figs. 4 and 5 one sees that the incident wave can almost always penetrate the soft cylinder at the dimensionless frequency $kh=1.5\pi$ such that $|T_1^1| \geqslant 0.8$, $|R_1^1| \leqslant 0.2$ approximately in the range of $0.01 \leqslant b/h \leqslant 0.9$ (see the dash-dot lines). However, with the increases of the radius of the cylinder, the hard cylinder and the cavity generate strong reflections to the incident wave such that $|T_1^1|$ descends and $|R_1^1|$

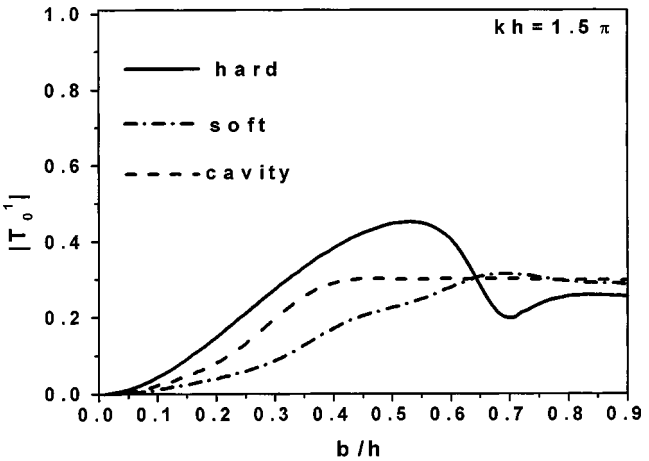


FIG. 6. Transmission coefficient $|T_0^1|$ vs normalized cylinder radius b/h for $kh=1.5\pi$.

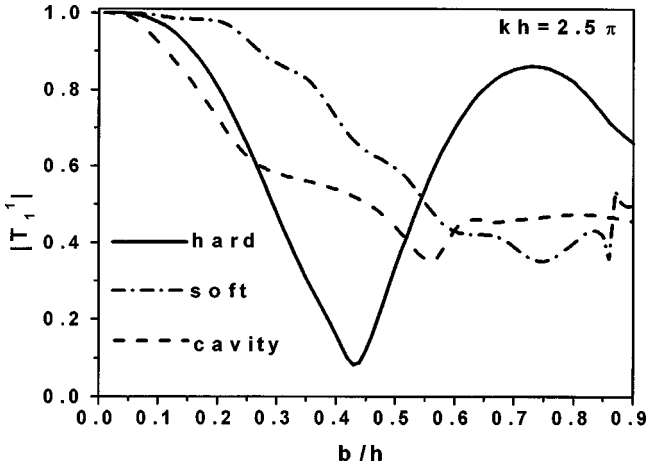


FIG. 8. Transmission coefficients $|T_1^1|$ vs normalized cylinder radius b/h for $kh=2.5\pi$.

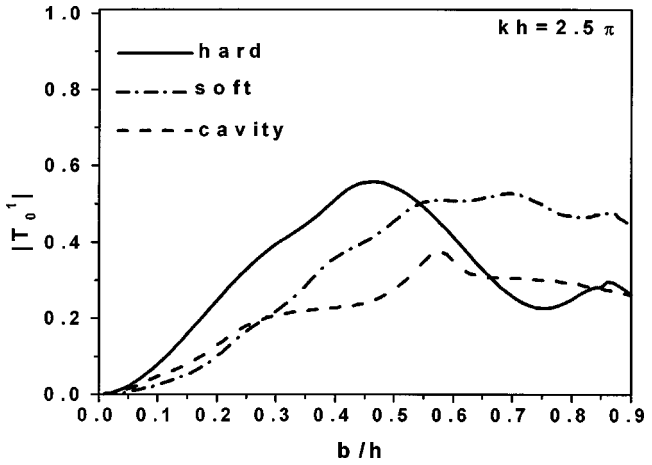


FIG. 9. Transmission coefficients $|T_0^1|$ vs normalized cylinder radius b/h for $kh=2.5\pi$.

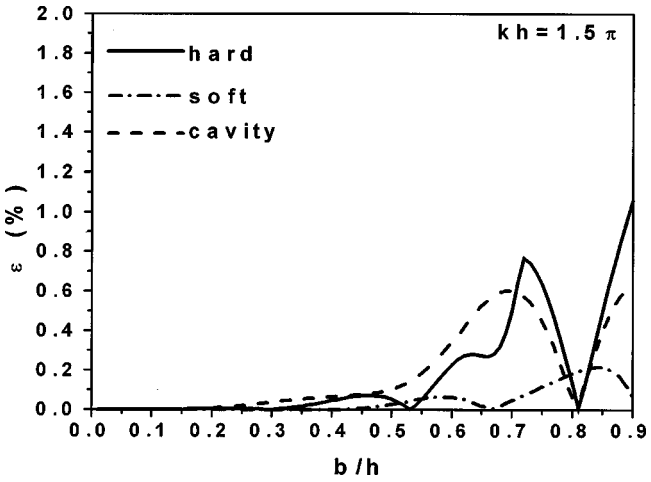


FIG. 11. Relative errors in numerical calculation for $kh=1.5\pi$.

ascends rapidly when $0.1\leqslant b/h<0.4$. In the range of $0.1\leqslant b/h<0.75$, the cavity generates stronger reflection to the incident wave than the hard cylinder. Complicated mode conversion curves for the transmitted and reflected modes due to the scattering of the cylinder are shown in Figs. 6 and 7, respectively. These two figures, with Figs. 4 and 5 together, may be reasonably interpreted by the principle of energy transform and conservation in the course of complex wave mode interactions.

For testing for disbonds, we focus on Fig. 4 once again. When the dimensionless frequency $kh=1.5\pi$, for the soft cylindrical post, the transmission coefficients $|T_1^1|$ are sufficiently different in the welded and slip or cracked (cavity) interface conditions. The transmission coefficient $|T_1^1|$ for the perfectly bonded soft cylinder is always much larger than that for the one imperfectly bonded, unless the diameter of the soft cylinder is very small. However, for a hard cylinder, the similar phenomenon happens only when $b/h<0.7$.

From Fig. 8, one sees that, unlike Fig. 4, the transmission coefficient $|T_1^1|$ for the perfect bonded soft cylinder descends fast down to less than 0.4 at $b/h\approx 0.75$. This indicates that the soft cylinder may generate stronger reflections to the incident wave with the increases of the dimensionless frequency kh. At the value of $kh=2.5\pi$, the differences of $|T_1^1|$ for the soft cylinder between the welded and the cracked (cavity) boundary conditions can still be seen in the range of $0.1<b/h<0.55$. For the perfectly bonded hard cylinder, $|T_1^1|$ descends steeply down to near zero at $b/h=0.43$ from which it rises rapidly up to 0.85 at $b/h\approx 0.7$ and then decreases again. The differences of $|T_1^1|$ for the hard cylinder in the welded and the cracked (cavity) boundary conditions are very prominent at $b/h=0.43$ or in the range of $0.6\leqslant b/h\leqslant 0.9$.

Interpretations for mode conversion curves of transmitted modes $|T_0^1|$ and $|T_2^1|$ shown in Figs. 9 and 10 may be made in a similar way to draw possible information for the evaluation of the bonding state of the cylinder. One may find the contrast variations of the coefficients $|T_0^1|$ or $|T_2^1|$ in bond and debond situations.

The accuracy of our numerical results is sufficient. From Figs. 11 and 12, it may be seen that the maximum relative error is less than 1.9% in the range of $0.01<b/h<0.9$.

CONCLUSIONS

Analytic solutions are obtained for the scattering of an antiplane SH-wave propagating in a traction-free two-

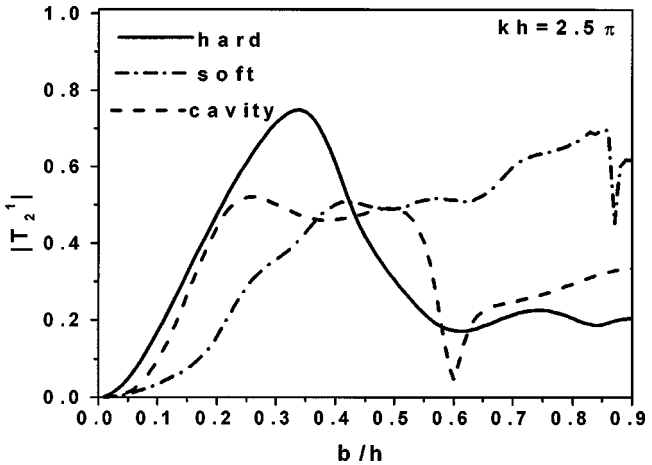


FIG. 10. Transmission coefficients $|T_2^1|$ vs normalized cylinder radius b/h for $kh=2.5\pi$.

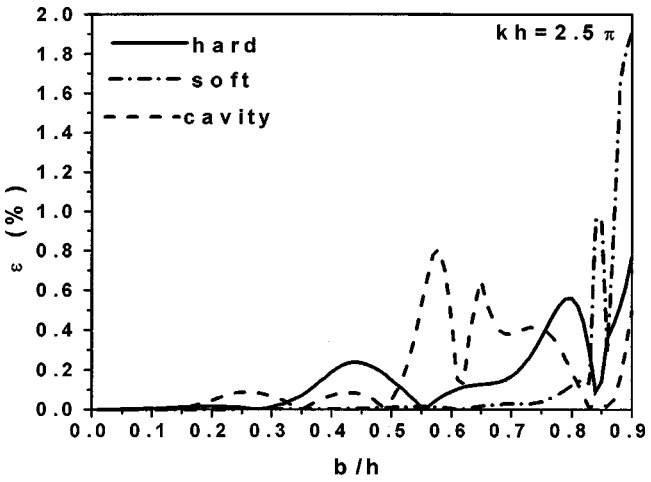


FIG. 12. Relative errors in numerical calculation for $kh=2.5\pi$.

dimensional plate by an infinite cylinder embeded therein, bonded or completely disbonded, by the use of the method of wave expansion and the method of mode matching. The expressions for the reflection and transmission coefficients are numerically evaluated by truncating the infinite expansion series. The numerical results are found to be sufficiently accurate with the maximum value of the relative errors being smaller than 2%.

According to the theoretical results, detection of disbonds between the cylinder and matrix is possible by examining the amplitude of a suitable scattered mode, chosen from the approximately computed curves, some illustrations of which are given in the present paper.

The approach of analysis used here has been applied also to the case of an incident Lamb wave. The results will be reported elsewhere.

ACKNOWLEDGMENTS

The work reported here is supported by a grant from the National Natural Science Foundation of China (No. 19774065). The first author is grateful to Dr. Liu, Y.B. for his effort in initiating this project. Thanks to the reviewer of our first manuscript for his/her helpful suggestions to improve our numerical precision.

APPENDIX: BOUNDARY CONDITION BETWEEN THE CYLINDER AND THE MATRIX IN TERMS OF AN INTERFACIAL LAYER

Assume that between the cylinder and the matrix, the interface is a thin solid layer instead of the two media in physical contact. Let the interfacial layer be described by the interphase ''*f*'' as shown in Fig. A1, with density ρ_f, shear modulus μ_f and thickness Δh. Its two surfaces contact, respectively, the cylinder at $r=b$ and the matrix at $r=a$, where $a=b+\Delta h$. Assuming welded contacts at $r=a$ and at $r=b$, we have

$$w_f^b=w_c^b, \quad \sigma_f^b=\sigma_c^b, \tag{A1}$$

$$w_f^a=w_2^a, \quad \sigma_f^a=\sigma_2^a, \tag{A2}$$

where $w_f^b=w_f|_{r=b}$, etc. The wave displacements w_2 of the matrix and w_c of the cylinder are given in the context by, respectively, Eq. (4) and Eq. (5):

$$w_2=\sum_{l=0}^{\infty}[D_lJ_l(kr)+E_lY_l(kr)]\cos l\theta \quad (r>a), \tag{A3}$$

$$w_c=\sum_{l=0}^{\infty}F_lJ_l(k_cr)\cos l\theta \quad (r<b). \tag{A4}$$

The wave displacement of the interfacial layer can likewise be written as

$$w_f=\sum_{l=0}^{\infty}[P_lJ_l(k_fr)+Q_lY_l(k_fr)]\cos l\theta \quad (b<r<a), \tag{A5}$$

where P_l and Q_l are the yet undetermined coefficients. The stress is related to the displacement by the following relation for any of the above w's and σ's:

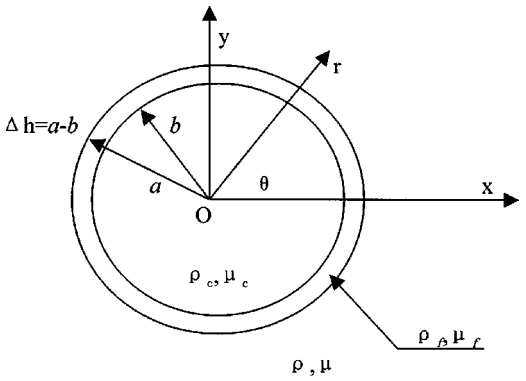


FIG. A1. Cross-section of the cylinder (''c'') and its interfacial layer (''f'') in the matrix.

$$\sigma=\sigma_{rz}=\mu\frac{\partial w}{\partial r}. \tag{A6}$$

Consider the interfacial layer itself. From Eq. (A5), we have, for the lth mode,

$$\begin{bmatrix} w_{fl}^a \\ \sigma_{fl}^a \end{bmatrix}=\begin{bmatrix} J_l(k_fa) & Y_l(k_fa) \\ \mu_fk_fJ_l'(k_fa) & \mu_fk_fY_l'(k_fa) \end{bmatrix} \times\begin{bmatrix} J_l(k_fb) & Y_l(k_fb) \\ \mu_fk_fJ_l'(k_fb) & \mu_fk_fY_l'(k_fb) \end{bmatrix}^{-1}\begin{bmatrix} w_{fl}^b \\ \sigma_{fl}^b \end{bmatrix}. \tag{A7}$$

With $k_fa=k_f(b+\Delta h)=k_fb+k_f\Delta h$, if the interfacial layer is so thin that $k_f\Delta h\ll 1$, then on expanding $J_1(k_fa)$ or $Y_l(k_fa)$ in $(k_f\Delta h)$ and on retaining only the first order terms, there results

$$J_l(k_fa)\approx J_l(k_fb)+(k_f\Delta h)J_l'(k_fb), \tag{A8}$$

$$Y_l(k_fa)\approx Y_l(k_fb)+(k_f\Delta h)Y_l'(k_fb). \tag{A9}$$

Inserting Eqs. (A8) and (A9) into Eq. (A7), one obtains

$$\begin{bmatrix} w_{fl}^a \\ \sigma_{fl}^a \end{bmatrix}=\begin{bmatrix} 1 & \dfrac{1}{R_p} \\ -\omega^2m_p & 1-\dfrac{\Delta h}{b} \end{bmatrix}\begin{bmatrix} w_{fl}^b \\ \sigma_{fl}^b \end{bmatrix}. \tag{A10}$$

Assume, in addition, $\Delta h\ll b$, the above relation becomes

$$\begin{bmatrix} w_{fl}^a \\ \sigma_{fl}^a \end{bmatrix}=\begin{bmatrix} 1 & \dfrac{1}{R_p} \\ -\omega^2m_{pl} & 1 \end{bmatrix}\begin{bmatrix} w_{fl}^b \\ \sigma_{fl}^b \end{bmatrix}. \tag{A11}$$

In this equation,

$$R_p=\mu_f/\Delta h, \quad m_{pl}=\rho_f\Delta h(1-l^2/k_f^2b^2), \tag{A12}$$

where $k_f^2=\omega^2\rho_f/\mu_f$ is the squared wave number of the shear waves in the interfacial layer. Equation (A11) describes the interfacial layer ''*f*'' as a thin plate, with the transverse stiffness R_p and the ''plate mass'' m_{pl}.[33] In other words, for the layer ''*f*,''

$$w_{fl}^a-w_{fl}^b\approx\sigma_{fl}^b/R_p, \quad \sigma_{fl}^a-\sigma_{fl}^b\approx-\omega^2m_{pl}w_{fl}^b. \tag{A13}$$

Return now to the boundary conditions between the cylinder and the matrix and consider the following two limiting

cases:

(1) $\Delta h \to 0$

This implies that the cylinder and the matrix are in direct contact. Presently, $\sigma_{fl}^b / R_p \to 0$, $m_{pl} \to 0$. From Eq. (A13) one sees that the relationship of the displacements and stresses between the exterior surface at $r=a$ and the interior surface at $r=b$ is independent to the wave mode l. Since the total field is the linear sum of all modes, from Eq. (A13) one has

$$w_f^a \approx w_f^b, \quad \sigma_f^a \approx \sigma_f^b, \tag{A14}$$

and it follows by the substitution of Eqs. (A1) and (A2),

$$w_2^b \approx w_c^b, \quad \sigma_2^b \approx \sigma_c^b, \tag{A15}$$

which is the ''welded'' interface boundary condition mentioned in the context in Eq. (11).

(2) $\mu_f \to 0$

This implies that the interfacial layer is filled with a nonviscous fluid. Such a layer does not support any shear wave and hence the incident wave can not cross it to enter the cylinder. For this case, $R_\rho \to 0$, $m_{pl} = \rho_f \Delta h$, and again one finds from Eq. (A13) that the relationship of the displacements and stresses between the exterior surface at $r=a$ and the interior surface at $r=b$ is independent to the wave mode l. Thus as before, from Eq. (A13), one obtains

$$\sigma_f^a = \sigma_f^b = 0, \tag{A16}$$

and it follows from Eqs. (A1) and (A2)

$$\sigma_2^b = 0, \quad \sigma_c^b = 0. \tag{A17}$$

This is the interface boundary condition we call in the context the ''slip'' or the ''cracked'' one in Eq. (18).

[1] J. J. Faran, Jr., ''Sound scattering by solid cylinders and spheres,'' J. Acoust. Soc. Am. **23**, 405–418 (1951).
[2] C. F. Ying and R. Truell, ''Scattering of plane longitudinal wave by a spherical obstacle in an isotropically elastic solid,'' J. Appl. Phys. **27**, 1086–1097 (1956).
[3] R. M. White, ''Elastic wave scattering at a cylindrical discontinuity in a solid,'' J. Acoust. Soc. Am. **30**, 771–785 (1958).
[4] Y. H. Pao and C. C. Mow, *Diffraction of Elastic Waves and Dynamic Stress Concentrations* (Crane Russak, New York, 1973).
[5] J. Miklowitz, *Theory of Elastic Waves and Waveguides* (North-Holland, New York, 1978).
[6] D. J. Jain and R. P. Kanwal, ''Scattering of elastic waves by circular cylindrical flaws and inclusions,'' J. Appl. Phys. **50**, 4067–4109 (1979).
[7] J. D. Achenbach, Y. Lu, and M. Kitahara, ''3-D reflection and transmission of sound by an array of rods,'' J. Sound Vib. **125**, 463–467 (1988).
[8] Y. B. Liu and R. S. Wu, ''Scattering and attenuation of a visco-elastic cylinder: I. P-wave incidence,'' J. Acoust. Soc. Am. **95**, 2803 (1994).
[9] Y. Lu, ''Three-dimensional reflection and transmission of elastic waves by an array of cylinders,'' J. Sound Vib. **175**, 1–15 (1994).
[10] Y. B. Liu and C. F. Ying, ''Scattering and attenuation of a visco-elastic cylinder: II. SV-wave incidence,'' J. Acoust. Soc. Am. **101**, 3153 (1997).
[11] A. N. Norris and Y. Yang, ''Shear wave scattering from a debounded fiber,'' in *Elastic Waves and Ultrasonic Nondestructive Evaluation*, edited by S. K. Datta, J. D. Achenbach, and Y. S. Rajapakse (Elsevier Science, North-Holland, 1990), pp. 437–440.
[12] Y. C. Chu and S. I. Rokhlin, ''Determination of macro- and micromechanical and interfacial elastic properties of composites from ultrasonic data,'' J. Acoust. Soc. Am. **92**, 920–931 (1992).
[13] P. Beattie, R. C. Chivers, and L. W. Andson, ''Ultrasonic backscattering from solid cylindrical inclusions in solid elastic matrices: A comparison of theory and experiment,'' J. Acoust. Soc. Am. **94**, 3421–3427 (1993).
[14] A. N. Sinclair and R. C. Anddison, ''Acoustic diffraction spectrum of a SiC fiber in a solid elastic medium,'' J. Acoust. Soc. Am. **94**, 1126–1135 (1993).
[15] Y. C. Chu and S. I. Rokhlin, ''Fiber-matrix interphase characterization in composites using ultrasonic velocity data,'' J. Appl. Phys. **76**, 4121–4129 (1994).
[16] W. Huang, S. I. Rokhlin, and Y. J. Wang, ''Effect of fiber-matrix interphase on wave propagation along, and scattering from, multilayered fibers in composites: Transfer matrix approach,'' Ultrasonics **33**, 365–375 (1995).
[17] W. Huang, S. Brisuda, and S. I. Rokhlin, ''Ultrasonic wave scattering from fiber-matrix interphases,'' J. Acoust. Soc. Am. **97**, 807–817 (1995).
[18] W. Huang, Y. J. Wang, and S. I. Rokhlin, ''Oblique scattering of an elastic wave from a mutilayered cylinder in a solid: Transfer-matrix approach,'' J. Acoust. Soc. Am. **99**, 2742–2754 (1996).
[19] Y. Lu, ''Guided antiplane shear wave propagation in layers reinforced by periodically spaced cylinders,'' J. Acoust. Soc. Am. **99**, 1937–1943 (1996).
[20] S. I. Rokhlin, ''Diffraction of Lamb waves by a finite crack in an elastic layer,'' J. Acoust. Soc. Am. **67**, 1157–1165 (1980).
[21] S. I. Rokhlin, ''Analysis of Lamb waves by a finite crack using the method of generalized scattering matrices 1980,'' Ultrason. Symp. Proc. IEEE, New York, pp. 922–926 (1980).
[22] S. I. Rokhlin, ''Resonance phenomena of Lamb waves scattering by a finite crack in a solid layer,'' J. Acoust. Soc. Am. **69**, 922–928 (1981).
[23] M. Koshiba, S. Karakida, and M. Suzuki, ''Finite-element analysis of discontinuity problem of SH modes in an elastic plate waveguide,'' Electron. Lett. **17**, 480–482 (1981).
[24] Z. Abduljabbar, S. K. Datta, and A. H. Shah, ''Diffraction of horizontally polarized shear waves by normal cracks in a plate,'' J. Appl. Phys. **54**, 461–472 (1983).
[25] M. Koshiba, H. Morita, and M. Suzuki, ''Finite-element analysis of Lamb wave scattering in an elastic plate waveguide,'' IEEE Trans. Sonics Ultrason. **SU-31**, 18–25 (1984).
[26] S. I. Rokhlin, ''Lamb wave interaction with lap-shear adhesive joints: Theory and experiment,'' J. Acoust. Soc. Am. **89**, 2758–2765 (1991).
[27] Y. N. Al-Nassar, S. K. Datta, and A. H. Shah, ''Scattering of Lamb waves by a normal rectangular strip of weldment,'' Ultrasonics **29**, 125–132 (1991).
[28] D. N. Alleyne and P. Cawley, ''The interaction of Lamb waves with defects,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 381–397 (1992).
[29] E. D. Nielsen, ''Scattering by a cylindrical post of complex permittivity in a waveguide,'' IEEE Trans. Microwave Theory Tech. **MTT-17**, 148–153 (1969).
[30] A. Lakhtakia, V. V. Varadan, and V. K. Varadan, ''Reflection characteristics of an elastic slab containing a periodic array of cylinders: SH wave analysis,'' J. Acoust. Soc. Am. **80**, 311–316 (1986).
[31] A. Lakhtakia, V. V. Varadan, and V. K. Varadan, ''Reflection characteristics of an elastic slab containing a periodic array of cylinders: P and SV wave analysis,'' J. Acoust. Soc. Am. **83**, 1267–1275 (1988).
[32] B. A. Auld, *Acoustic Fields and Waves in Solids, Vol. II* (Wiley, New York, 1973).
[33] S. I. Rokhlin and Y. J. Wang, ''Analysis of boundary conditions for elastic wave interaction with an interface between two solids,'' J. Acoust. Soc. Am. **89**, 503–515 (1991).

On approximating guided waves in plates with thin anisotropic coatings by means of effective boundary conditions

A. Jonas Niklasson, Subhendu K. Datta, and Martin L. Dunn
University of Colorado, Department of Mechanical Engineering, Campus Box 427, Boulder, Colorado 80309-0427

(Received 15 May 1999; revised 1 May 2000; accepted 5 May 2000)

In this paper, effective boundary conditions for elastic wave propagation in plates with thin coatings are derived. These effective boundary conditions are used to obtain an approximate dispersion relation for guided waves in an isotropic plate with thin anisotropic coating layers. The accuracy of the effective boundary conditions is investigated numerically by comparison with exact solutions for two different material systems. The systems considered consist of a metallic core with thin superconducting coatings. It is shown that for wavelengths long compared to the coating thickness there is excellent agreement between the approximate and exact solutions for both systems. Furthermore, numerical results presented might be used to characterize coating properties by ultrasonic techniques. © *2000 Acoustical Society of America.* [S0001-4966(00)03908-4]

PACS numbers: 43.20.Mv, 43.35.Cg [DEC]

INTRODUCTION

Ultrasonic guided wave propagation in thin coating layers on a thick elastic substrate has been investigated by many researchers for applications in seismology, electronic devices, and thermal barrier coatings, to name a few. In most of these studies the substrate is assumed to be much thicker than the wavelengths of interest so that it can be effectively modeled as a semi-infinite medium. References to early studies can be found in the review article by Farnell and Adler.[1] Recent studies by Bouden and Datta,[2] and Sotiropoulos,[3] of guided waves in an elastic layer over an anisotropic halfspace have revealed that guided waves in the layer may not exist for all frequencies. References to other related works can also be found in Refs. 2 and 3.

Of particular relevance to the present work are studies of guided waves in anisotropic layered or laminated plates. There is considerable recent interest in these problems for ultrasonic characterization of composite plates. In order to develop/improve such experimental methods, mathematical modeling and numerical simulation are of great importance. This is especially true since the anisotropy has been found to alter the characteristics of guided elastic waves significantly (Auld[4] and Nayfeh[5]). Li and Thompson,[6] and Solie and Auld,[7] have presented detailed studies of guided waves in anisotropic plates. These investigations have been restricted, however, to homogeneous plates. Karunasena *et al.*[8] consider guided waves in multilayered cross-ply composite plates. For an extensive review of the use of guided waves in plates for material characterization, we refer to the review article by Chimenti[9] and the references therein. The readers may also wish to refer to Nayfeh[5] and Datta,[10] where the theory of guided waves in layered anisotropic plates has been discussed in detail.

In this paper, our attention is focused on guided waves in a plate with thin superconducting coating layers. The motivation here is to develop a fundamental understanding of ultrasonic guided waves in tapes that are fabricated for commercial high-current applications. Fabrication processes are being developed for different classes of high-T_c materials systems. The tapes are composites consisting of a brittle superconducting phase and a ductile metal phase. They are: (1) in the form of a superconducting oxide layer (such as BSCCO) sandwiched in a high-conductivity metal like silver, (2) dip-coated silver tape in which the oxide layers are outside the core conducting layer, (3) silver-sheathed dip-coated tapes, and (4) deposition of biaxially aligned oxide (YBCO) coating on nickel tapes. Mechanical processes such as rolling are used to get the crystallographic texture most favorable to high current capacity of the tapes. Such mechanical processes coupled with thermal cycling cause microcracking of the brittle oxide layer(s) which limits the current carrying capacity. The degree of current carrying capacity reduction is a strong function of the crystallographic texture of the oxide layer and the nature of microcracking, which is also a strong function of the texture. These effects influence the mechanical response such as ultrasonic velocities and attenuation of guided waves propagating along the tapes. Exploitation of the connection between the electrical and mechanical response may prove to be an efficient and accurate means of nondestructive material property characterization during and after processing of the tapes.

Investigation of *in situ* mechanical behavior and properties of thin superconducting layers and tapes has been very limited. However, it is well known that the properties are highly dependent on the manufacturing processes, external and internal stress fields, interface properties and porosity, to name a few. In order to accurately model these different effects, it is necessary to understand the basic problem of guided wave propagation in three-layered tapes. The tapes may consist of either an elastic isotropic metallic core coated with thin anisotropic oxide layers or a thin oxide layer sandwiched between two metallic layers. In this paper, a model study of the first case is presented. It is assumed that the thickness of the oxide layer is small compared to all wavelengths in the coatings or layers. One way of approximating the coatings or layers is to use plate theory. For a comparison

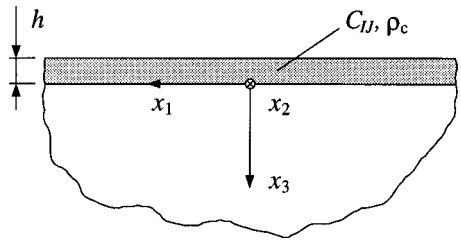


FIG. 1. An anisotropic layer on a half-space.

between the use of plate theory and exact theory for wave propagation in composite plates, see Lih and Mal.[11] A more rational approach to approximating thin coatings or layers is to expand the fields in the (small) thickness and obtain approximations by truncation of these infinite series. This approach has been taken by Bövik[12–14] to approximate thin elastic interface layers and thin coatings. In these cases the materials are assumed to be elastically isotropic, except for a two-dimensional anti-plane case. Its and Lee[15] use series expansions to study the propagation of surface waves across a transversely isotropic interface layer in a way which is similar to the approach by Bövik. In this paper, we extend the approach by Bövik to anisotropic coatings for the three-dimensional case. The approach taken by Rokhlin and co-workers[16–19] is also similar to the one by Bövik. Rokhlin and his group, however, do not derive approximate interface conditions by expanding the field in the layer. Instead, they derive the transfer matrix of the layer and expand its components in the thickness of the layer.

The plan of the present paper is as follows. The effective boundary conditions (BCs) for a thin anisotropic coating perfectly bonded to an elastic (possibly anisotropic) half-space are derived in the next section as series expansions in the thickness of the coating. The propagation of guided waves in a plate made of an isotropic core and two identical anisotropic coatings is approximated by means of a truncation of the series for the BCs. In the numerical examples, dispersion curves are presented for two different superconducting tapes. Figures are shown that illustrate some of the features of guided waves in the tapes. Comparisons are also made between exact and approximate dispersion curves.

I. THE EFFECTIVE BOUNDARY CONDITIONS

In this section, we derive effective BCs for wave propagation in a thin anisotropic coating on an elastic half-space. First, the problem is stated mathematically in Sec. I A and the BCs are derived as series expansions in the thickness of the coating in Sec. I B

A. Problem statement

Consider a thin anisotropic coating of thickness h perfectly bonded to an elastic half-space (see Fig. 1). Let the displacement and stress be denoted by u_j and σ_{jm}, respectively, in the half-space ($x_3>0$) and by U_j and Σ_{jm}, respectively, in the coating ($-h<x_3<0$). In the coating we assume that the usual elastodynamic equations of motion hold

$$\partial_m \Sigma_{mj} = \rho_c \partial_t^2 U_j, \tag{1}$$

where ρ_c is the density of the coating and the summation convention is used. The notation ∂_m means $\partial/\partial x_m$ and ∂_t means $\partial/\partial t$, where t is the time. For simplicity, we assume that the x_1x_2 plane is a plane of elastic symmetry of the coating (i.e., monoclinic material symmetry). Hooke's law may then be written as, using Voigt's abbreviated notation (see Auld[4]),

$$\begin{pmatrix} \Sigma_{11} \\ \Sigma_{22} \\ \Sigma_{33} \\ \Sigma_{23} \\ \Sigma_{13} \\ \Sigma_{12} \end{pmatrix} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & 0 & 0 & C_{16} \\ C_{12} & C_{22} & C_{23} & 0 & 0 & C_{26} \\ C_{13} & C_{23} & C_{33} & 0 & 0 & C_{36} \\ 0 & 0 & 0 & C_{44} & C_{45} & 0 \\ 0 & 0 & 0 & C_{45} & C_{55} & 0 \\ C_{16} & C_{26} & C_{36} & 0 & 0 & C_{66} \end{pmatrix} \times \begin{pmatrix} \partial_1 U_1 \\ \partial_2 U_2 \\ \partial_3 U_3 \\ \partial_3 U_2 + \partial_2 U_3 \\ \partial_3 U_1 + \partial_1 U_3 \\ \partial_2 U_1 + \partial_1 U_2 \end{pmatrix}. \tag{2}$$

Symmetry of the stress tensors has been used in writing Eq. (2) and in the following. Since the bond between the coating and the half-space is assumed to be perfect, the following conditions must hold:

$$u_j = U_j, \quad x_3 = 0, \tag{3a}$$

$$\sigma_{j3} = \Sigma_{j3}, \quad x_3 = 0. \tag{3b}$$

On the boundary $x_3 = -h$, an applied traction T_j is assumed to exist, i.e.,

$$\Sigma_{j3} = T_j, \quad x_3 = -h. \tag{4}$$

The goal is now to replace the effect of the coating on the half-space by an effective BC. This is done in the next section as a series expansion in the thickness of the coating h.

B. The derivation of the boundary conditions

Introduce the notation $(f)_a$ for $f(x_1,x_2,a,t)$ and expand the traction Σ_{j3} in a Taylor series

$$(\Sigma_{j3})_{-h} = (\Sigma_{j3})_0 - h(\partial_3 \Sigma_{j3})_0 + O(h^2). \tag{5}$$

From Eqs. (3b) and (4) we immediately have $(\Sigma_{j3})_{-h} = T_j$ and $(\Sigma_{j3})_0 = (\sigma_{j3})_0$. The equations of motion (1) may be written as

$$\partial_3 \Sigma_{j3} = \rho_c \partial_t^2 U_j - \partial_1 \Sigma_{j1} - \partial_2 \Sigma_{j2}. \tag{6}$$

From Eq. (3a), we have $(U_j)_0 = (u_j)_0$ and when $j=3$, Eq. (6) at $x_3 = 0$ becomes

$$(\partial_3 \Sigma_{33})_0 = \rho_c (\partial_t^2 u_3)_0 - (\partial_1 \sigma_{13})_0 - (\partial_2 \sigma_{23})_0, \tag{7}$$

if the conditions at the interface (3) are used. In order to express the remaining two components of Eq. (6) at $x_3 = 0$ in u_j and/or σ_{j3}, we consider Hooke's law (2). By differentiating the expressions for Σ_{11}, Σ_{12}, and Σ_{22} with respect to x_1 and x_2, we get

$$\partial_1\Sigma_{11}=C_{11}\partial_1^2U_1+C_{12}\partial_1\partial_2U_2+C_{13}\partial_1\partial_3U_3 + C_{16}(\partial_1\partial_2U_1+\partial_1^2U_2), \tag{8a}$$

$$\partial_1\Sigma_{12}=C_{16}\partial_1^2U_1+C_{26}\partial_1\partial_2U_2+C_{36}\partial_1\partial_3U_3 + C_{66}(\partial_1\partial_2U_1+\partial_1^2U_2), \tag{8b}$$

$$\partial_2\Sigma_{12}=C_{16}\partial_1\partial_2U_1+C_{26}\partial_2^2U_2+C_{36}\partial_2\partial_3U_3 + C_{66}(\partial_2^2U_1+\partial_1\partial_2U_2), \tag{8c}$$

$$\partial_2\Sigma_{22}=C_{12}\partial_1\partial_2U_1+C_{22}\partial_2^2U_2+C_{23}\partial_2\partial_3U_3 + C_{26}(\partial_2^2U_1+\partial_1\partial_2U_2). \tag{8d}$$

From Hooke's law, we further have

$$\partial_3U_3=(\Sigma_{33}-C_{13}\partial_1U_1-C_{23}\partial_2U_2-C_{36}\partial_2U_1-C_{36}\partial_1U_2)/C_{33}. \tag{9}$$

If Eqs. (3) and (9) are used in Eq. (8) at $x_3=0$ and this is inserted into Eq. (6) for $j=1$ and 2, all of the components of $(\partial_3\Sigma_{j3})_0$ are finally obtained.

With Eq. (6) expressed in u_j and σ_{j3}, the expansion (5) may be written as

$$\boldsymbol{\tau}+h(A_\sigma\boldsymbol{\tau}+A_u\mathbf{u})=\mathbf{T}+O(h^2), \quad x_3=0, \tag{10}$$

where $\boldsymbol{\tau}=(\sigma_{13},\sigma_{23},\sigma_{33})^T$, $\mathbf{u}=(u_1,u_2,u_3)^T$, $\mathbf{T}=(T_1,T_2,T_3)^T$ and the superscript T means transpose. The nonzero elements of the matrix A_σ are

$$(A_\sigma)_{13}=\frac{C_{13}}{C_{33}}\partial_1+\frac{C_{36}}{C_{33}}\partial_2, \tag{11a}$$

$$(A_\sigma)_{23}=\frac{C_{36}}{C_{33}}\partial_1+\frac{C_{23}}{C_{33}}\partial_2, \tag{11b}$$

$$(A_\sigma)_{31}=\partial_1, \tag{11c}$$

$$(A_\sigma)_{32}=\partial_2, \tag{11d}$$

and the nonzero elements of the symmetric matrix A_u are

$$(A_u)_{11}=-\rho_c\partial_t^2+\left(C_{11}-\frac{C_{13}^2}{C_{33}}\right)\partial_1^2 + 2\left(C_{16}-\frac{C_{13}C_{36}}{C_{33}}\right)\partial_1\partial_2+\left(C_{66}-\frac{C_{36}^2}{C_{33}}\right)\partial_2^2, \tag{12a}$$

$$(A_u)_{12}=\left(C_{16}-\frac{C_{13}C_{36}}{C_{33}}\right)\partial_1^2 + \left(C_{12}+C_{66}-\frac{C_{13}C_{23}}{C_{33}}-\frac{C_{36}^2}{C_{33}}\right)\partial_1\partial_2 + \left(C_{26}-\frac{C_{23}C_{36}}{C_{33}}\right)\partial_2^2, \tag{12b}$$

$$(A_u)_{22}=-\rho_c\partial_t^2+\left(C_{66}-\frac{C_{36}^2}{C_{33}}\right)\partial_1^2 + 2\left(C_{26}-\frac{C_{23}C_{36}}{C_{33}}\right)\partial_1\partial_2+\left(C_{22}-\frac{C_{23}^2}{C_{33}}\right)\partial_2^2, \tag{12c}$$

$$(A_u)_{33}=-\rho_c\partial_t^2. \tag{12d}$$

If the thickness h is small compared to all wavelengths in the coating, it should be possible to obtain an accurate approximation of the effect of the wave propagation in the coating on the half-space by using the approximate BC,

$$\boldsymbol{\tau}+h(A_\sigma\boldsymbol{\tau}+A_u\mathbf{u})=\mathbf{T}, \quad x_3=0, \tag{13}$$

on the half-space. The accuracy of these approximate effective BCs are investigated numerically in Sec. III for guided waves in superconducting tapes. An approximate dispersion relation for guided waves in a coated plate is derived in the next section. In addition, for the special case when the in-plane and anti-plane waves decouple, BCs which are exact up to the order h^2 are derived in Appendix A. These, more refined, BCs are used in Appendix B to derive a more accurate approximation of the dispersion relation for the plate. This approximation is also considered in the numerical examples in Sec. III.

If we compare the approximate BCs, Eq. (13), to the ones derived by Bövik,[14] we find that they are identical for isotropic coatings. We also find that they give the same set of BCs as the ones derived by Rokhlin and Huang[19] if we assume that our fields are in the form $u_j(x_1,x_2,x_3,t)=v_j(x_3)e^{i(kx_1-\omega t)}$ and $\sigma_{j3}(x_1,x_2,x_3,t)=s_{j3}(x_3)e^{i(kx_1-\omega t)}$ and rewrite the BCs in Ref. 19 for a coating layer instead of an interface layer. It should be noted that the method developed in Ref. 19 is significantly different from the method used here. Here, the field quantities in the thin anisotropic layer are expanded in powers of the thickness h in order to derive rational boundary conditions [Eq. (13)]. As mentioned in the Introduction, the authors of Ref. 19 use transfer matrices to describe transmission and reflection by an interface layer.

II. GUIDED WAVES IN A PLATE WITH THIN ANISOTROPIC COATINGS

In this section, we use the effective BCs derived in the previous section to obtain an approximate dispersion relation for a layered anisotropic plate. The plate, shown in Fig. 2,

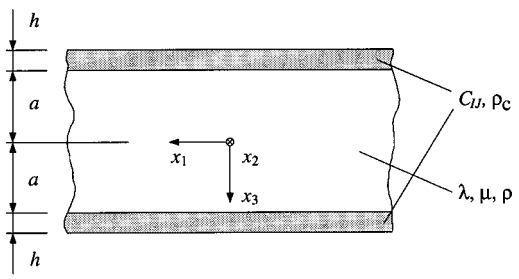


FIG. 2. The anisotropic plate.

consists of a thick isotropic core and two identical anisotropic thin coating layers.

In order to derive the approximate dispersion relation, we replace the thin coatings by the truncated series expansion (13). The equations we need to solve are then given by

$$\partial_m \sigma_{mj} = \rho \partial_t^2 u_j, \qquad -a < x_3 < a, \tag{14a}$$

$$\boldsymbol{\tau} \mp h(A_\sigma \boldsymbol{\tau} + A_u \mathbf{u}) = 0, \qquad x_3 = \pm a, \tag{14b}$$

where

$$\sigma_{jm} = \lambda \delta_{jm} \partial_n u_n + \mu(\partial_m u_j + \partial_j u_m), \tag{14c}$$

with the solution in the form $u_j(x_1, x_2, x_3, t) = v_j(x_3) e^{i(kx_1 - \omega t)}$.

The general solution to Eq. (14a) may be written in terms of potentials as (see Achenbach[20])

$$\mathbf{u} = \nabla \Phi + \nabla \times \boldsymbol{\Psi}, \quad \nabla \cdot \boldsymbol{\Psi} = 0, \tag{15a}$$

$$\Phi = (A_0 \sin px_3 + B_0 \cos px_3) e^{i(kx_1 - \omega t)}, \tag{15b}$$

$$p = \sqrt{k_p^2 - k^2}, \quad k_p = \omega / c_p, \quad c_p^2 = (\lambda + 2\mu)/\rho,$$

$$\Psi_j = (A_j \sin qx_3 + B_j \cos qx_3) e^{i(kx_1 - \omega t)}, \tag{15c}$$

$$q = \sqrt{k_s^2 - k^2}, \quad k_s = \omega / c_s, \quad c_s^2 = \mu/\rho,$$

where $j = 1,2,3$. The dispersion relation is obtained by applying (14b) to (15). Due to the symmetry of the problem we split the solution into a symmetric and an antisymmetric part. For the symmetric modes we obtain the relation

$$\begin{vmatrix} E_{11} h \cos qa + F_{11} \sin qa & E_{12} h \cos pa + F_{12} \sin pa & E_{13} h \cos qa \\ E_{21} h \cos qa & E_{22} h \cos pa & E_{23} h \cos qa + F_{23} \sin qa \\ E_{31} \cos qa + F_{31} h \sin qa & E_{32} \cos pa + F_{32} h \sin pa & 0 \end{vmatrix} = 0, \tag{16a}$$

and for the antisymmetric modes we obtain the relation

$$\begin{vmatrix} F_{11} \cos qa - E_{11} h \sin qa & -F_{12} \cos pa + E_{12} h \sin pa & E_{13} h \sin qa \\ -E_{21} h \sin qa & E_{22} h \sin pa & -F_{23} \cos qa + E_{23} h \sin qa \\ -F_{31} h \cos qa + E_{31} \sin qa & F_{32} h \cos pa - E_{32} \sin pa & 0 \end{vmatrix} = 0, \tag{16b}$$

where

$$E_{11} = \left(\rho_c \omega^2 - \left(C_{11} - \frac{C_{13}(C_{13} + 2\mu)}{C_{33}} \right) k^2 \right) q, \tag{17a}$$

$$E_{12} = -i \left(\rho_c \omega^2 - \left(C_{11} + \frac{C_{13}(\lambda - C_{13})}{C_{33}} \right) k^2 - (\lambda + 2\mu) \frac{C_{13}}{C_{33}} p^2 \right) k, \tag{17b}$$

$$E_{13} = -i \left(C_{16} - \frac{C_{13} C_{36}}{C_{33}} \right) k_s^2 k, \tag{17c}$$

$$E_{21} = -\left(C_{16} - \frac{C_{36}(C_{13} + 2\mu)}{C_{33}} \right) k^3 q, \tag{17d}$$

$$E_{22} = i \left((\lambda + 2\mu) \frac{C_{36}}{C_{33}} p^2 + \left(C_{16} + \frac{C_{36}(\lambda - C_{13})}{C_{33}} \right) k^2 \right) k^2, \tag{17e}$$

$$E_{23} = i \left(\rho_c \omega^2 - \left(C_{66} - \frac{C_{36}^2}{C_{33}} \right) k^2 \right) k_s^2, \tag{17f}$$

$$E_{31} = 2i\mu kq, \tag{17g}$$

$$E_{32} = -(\lambda k^2 + (\lambda + 2\mu) p^2), \tag{17h}$$

$$F_{11} = \mu(q^2 - k^2), \tag{17i}$$

$$F_{12} = -2i\mu kp, \tag{17j}$$

$$F_{23} = i\mu k_s^2 q, \tag{17k}$$

$$F_{31} = i(\mu(k^2 - q^2) - \rho_c \omega^2) k, \tag{17l}$$

$$F_{32} = (\rho_c \omega^2 - 2\mu k^2) p. \tag{17m}$$

It is seen that the use of the effective boundary conditions leads to 3-by-3 determinantal equations for the symmetric and antisymmetric cases instead of 9-by-9 determinantal equations obtained by using the global matrix method (see Sec. III). This leads to a substantial decrease in the computational effort. Another important advantage is that the thickness of the layer h appears explicitly as a simple factor in the elements of the determinants (16a) and (16b).

III. NUMERICAL EXAMPLES

In this section we present some numerical examples for guided waves in the coated plate presented in Sec. II. In the

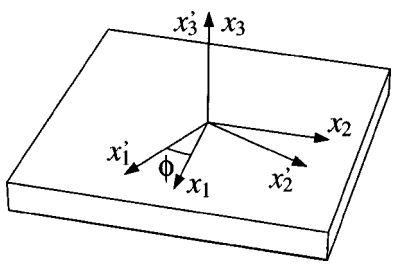


FIG. 3. The orientation of the crystal axes system.

TABLE I. Material properties of YBCO, Bi-2212, Ni, and Ag (C'_{JM} in GPa and ρ in kg/m^3).

Material	C'_{11}	C'_{22}	C'_{33}	C'_{44}	C'_{55}	C'_{66}	C'_{12}	C'_{13}	C'_{23}	ρ
YBCO[a]	268	231	186	37	49	95	132	95	71	6333
Bi-2212[b]	75.8	125.2	125.2	23.15	15.8	15.8	56	56	78.9	6510
Ni	299.0	299.0	299.0	84.7	84.7	84.7	129.5	129.5	129.5	8910
Ag	141.5	141.5	141.5	30.24	30.24	30.24	81.01	81.01	81.01	10500

[a]Reference 21.
[b]Reference 22.

examples, we present the phase velocity $c=\omega/k$ as a function of the frequency $f=\omega/2\pi$, where the pairs (k, ω) are solutions to the different dispersion relations derived in this paper.

We consider two different superconducting tapes. Both tapes have the same dimensions with the thickness of the core being $2a=100\,\mu$m and the thickness of each coating being $h=5\,\mu$m. In the first example, we consider $YBa_2Cu_3O_{7-\delta}$ (YBCO) coatings on a nickel (Ni) core and in the second, we consider $Bi_2Sr_2CaCu_2O_{8+\delta}$ (Bi-2212 or BSCCO) coatings on a silver (Ag) core. The material properties of all materials are given in Table I. Note that only the nonzero elastic constants are included. As is seen from the table, YBCO is orthotropic, Bi-2212 is transversely isotropic, and Ni and Ag are isotropic. The prime on the elastic constants means that they are given in the materials crystal axes system, denoted by (x'_1,x'_2,x'_3). The orientation of the crystal axes system relative to the (x_1,x_2,x_3) system is given by the angle ϕ as shown in Fig. 3. The transformation from the crystal axes system to this latter system is given by Auld.[4]

In the figures, we compare approximations of the dispersion relation to the exact one. We also compare exact dispersion curves for different orientations of the anisotropic coatings. The exact dispersion relation is obtained via the global matrix method applied to a layered plate consisting of two coatings and one core with perfect bonds and traction-free boundaries. Briefly, in the global matrix method, a global matrix equation is derived by first calculating general solutions in each layer and then using the boundary and/or the interface conditions for each layer. This leads to $6N$ homogeneous equations in $6N$ unknown integration constants for an N-layered plate. For the use of the global matrix method on layered plates, we refer to the article by Ju and Datta[23] and the references stated therein. Note that we obtain two 9-by-9 determinantal equations in this case since symmetric and antisymmetric modes decouple.

A. The YBCO/Ni/YBCO tape

The first tape consists of an isotropic core of nickel (Ni) and two identical thin coatings made of the orthotropic superconductor $YBa_2Cu_3O_{7-\delta}$ (YBCO). In all figures, we consider the frequency range 0–100 MHz. In the plots, the shortest wavelengths, corresponding to quasi-S waves in the coatings, are approximately five times the thickness of the coatings. This means that the coatings may be considered to be relatively thin.

In Fig. 4, the dispersion curves for a pure nickel plate of thickness $2a$ (i.e., the case when $h=0$) are shown together with the dispersion curves for the coated tape for $\phi=45°$. Here, only the exact dispersion curves are considered. This figure is included in order to show the influence of the thin coatings on the dispersion curves for the Ni plate. As is seen from the figure, the difference is significant, except for the lowest three modes. Note that the lowest three modes are (for the Ni plate): flexural (A_0), horizontally polarized shear (SH_0), and longitudinal (S_0). For example, at $c=4$ km/s, the decrease in frequency f is approximately 6.3% for S_0, 6.8% for SH_1, 6.5% for A_1, and 7.2% for SH_2.

In Fig. 5, the velocity c of the lowest (flexural) mode for the layered YBCO/Ni/YBCO tape is shown as a function of the angle ϕ for three different frequencies. It is seen from this figure that the anisotropy of the coatings introduces a variation in the velocity.

In Fig. 6, we compare the exact dispersion curves to the ones obtained when the coatings are approximated by the effective boundary conditions (16). Here, the angle between

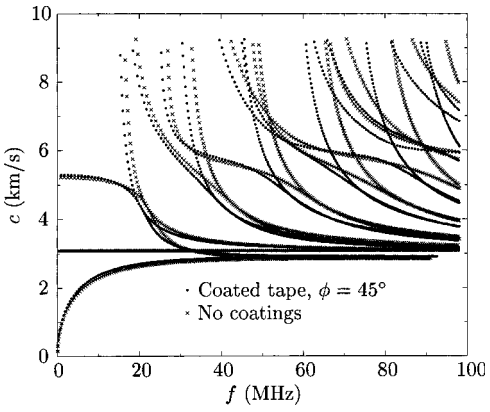


FIG. 4. The exact solutions for the coated YBCO/Ni/YBCO tape ($\phi=45°$) and for the uncoated Ni tape.

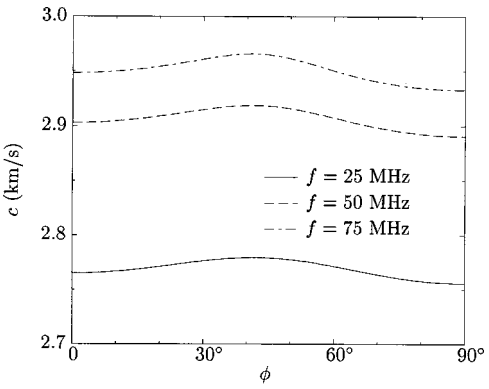


FIG. 5. The velocity from the exact solutions for the coated YBCO/Ni/YBCO tape as a function of the direction.

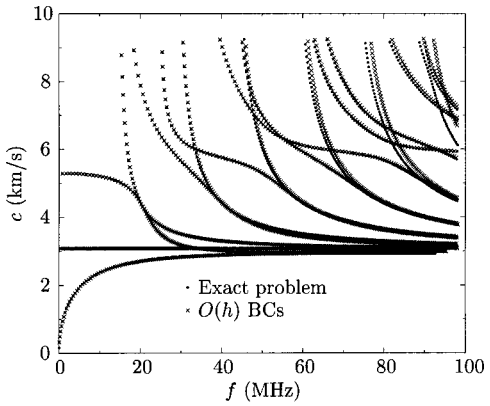


FIG. 6. The solution to the exact problem versus the solution to the approximate problem obtained from the $O(h)$ BCs for $\phi=45°$ for the YBCO/Ni/YBCO tape.

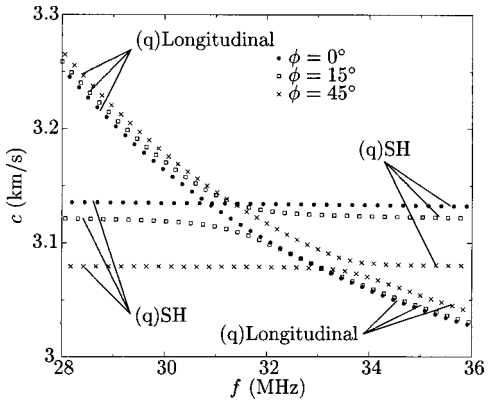


FIG. 8. A magnification of the exact dispersion curves for the YBCO/Ni/YBCO tape for different values of ϕ.

the crystal axes system and the coordinate system is $\phi=45°$, making it a three-dimensional (3D) problem. From the figure, we see that the agreement is excellent when the frequency is below 60 MHz. The wavelength of the slowest quasi-S wave in the coatings at this frequency is about 40 μm. It is interesting to note that as the frequency is increased above 60 MHz the largest discrepancies occur for the qSH modes near cut off. It is seen that the $O(h)$ approximation predicts higher frequencies at a fixed phase velocity.

In Fig. 7, we consider the case when the crystal axes system and the coordinate system coincide. The orientation makes this a two-dimensional (2D) case where the in-plane and anti-plane waves decouple (see Appendix A). This problem has been discussed in detail by Pan and Datta.[24] The conclusion in this case is the same as for Fig. 6. The main reason for including this case is for comparison with the $O(h^2)$ BCs (see Appendix A) below.

Figures 6 and 7 look very much alike, but there are some major differences between the two cases. When the coating material is oriented in a direction other than $\phi=n\pi/2$, $n=0,1,...$, all waves will couple due to the anisotropy of the coatings. We therefore show magnifications of the exact dispersion curves for different values of the angle ϕ in Figs. 8 and 9. Even though the coatings are thin and moderately anisotropic, the effect of the anisotropy is clearly seen in both figures. In the case when $\phi=0°$, the curves will cross since the anti-plane waves decouple from the in-plane ones. For the other cases shown, the curves cannot cross since all waves couple (note that the symmetric and antisymmetric modes still decouple). As is seen from the figures, the anisotropy alters the minimum distance between the curves and the location of that point quite a bit. This information might be useful for ultrasonic characterization of the tapes. Especially Fig. 8 is interesting since it involves the lowest modes. We do not show the approximate dispersion curves in the figures, but it should be mentioned that the coupling is preserved by the approximate BCs.

The dispersion curves for the more refined $O(h^2)$ BCs [Eqs. (A8) and (A13)] are compared to the exact case in Fig. 10 for the 2D case when $\phi=0°$. By comparing this figure to Fig. 7, which is the $O(h)$ case, we see that the accuracy of the approximation is improved considerably for frequencies above 60 MHz. Note that the $O(h^2)$ approximation predicts lower frequencies at a fixed phase velocity.

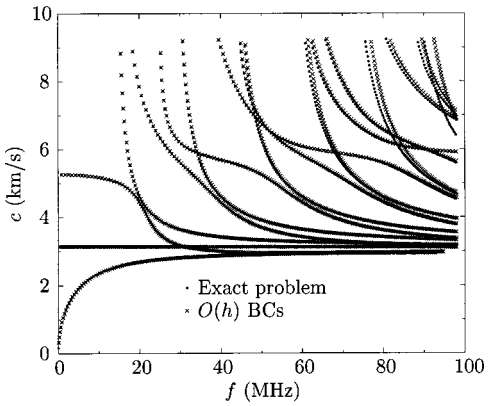


FIG. 7. The solution to the exact problem versus the solution to the approximate problem obtained from the $O(h)$ BCs for $\phi=0°$ for the YBCO/Ni/YBCO tape.

B. The Bi-2212/Ag/Bi-2212 tape

The second tape consists of an isotropic core of silver (Ag) and two identical thin coatings made of the superconductor $Bi_2Sr_2CaCu_2O_{8+\delta}$ (Bi-2212), which is transversely isotropic. In this case, we consider the frequency range 0–50 MHz. In the plots, the shortest wavelengths, corresponding

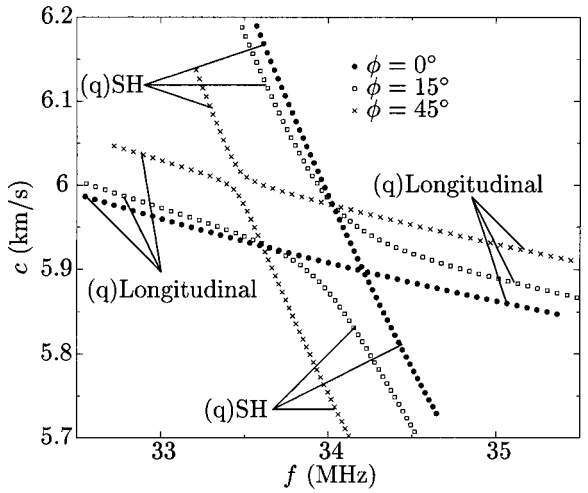


FIG. 9. A magnification of the exact dispersion curves for the YBCO/Ni/YBCO tape for different values of ϕ.

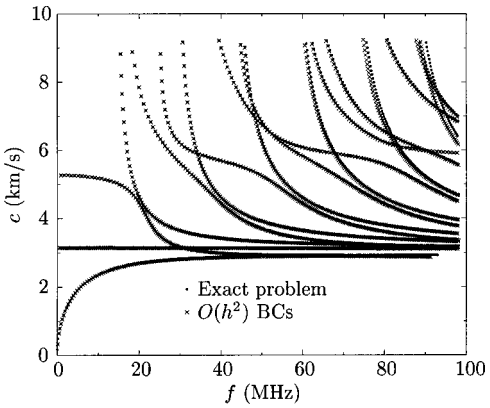


FIG. 10. The solution to the exact problem versus the solution to the approximate problem obtained from the $O(h^2)$ BCs for $\phi=0°$ for the YBCO/Ni/YBCO tape.

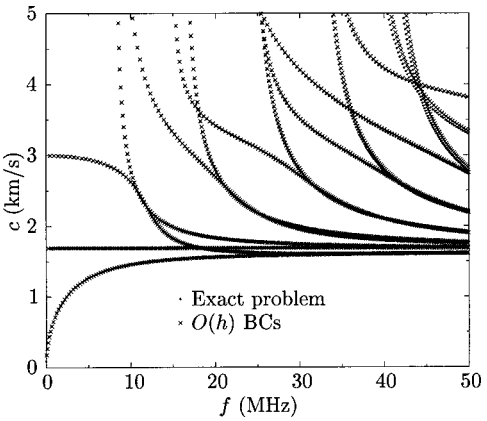


FIG. 12. The solution to the exact problem versus the solution to the approximate problem obtained from the $O(h)$ BCs for $\phi=0°$ for the Bi-2212/Ag/Bi-2212 tape.

to quasi-S waves in the coatings, are approximately six times the thickness of the coatings. This means that the coatings may be considered to be relatively thin.

In Fig. 11, the dispersion curves for a pure silver plate of thickness $2a$ (i.e., the case when $h=0$) are shown together with the dispersion curves for the coated Bi-2212/Ag/Bi-2212 tape for $\phi=45°$. Here, only the exact dispersion curves are considered. Again, this figure is included in order to show that there is a measurable influence of the thin coatings on the dispersion curves for the isotropic plate.

In Figs. 12 and 13, we compare the exact dispersion curves to those obtained using the $O(h)$ boundary conditions for the cases $\phi=0$ and $\phi=45°$, respectively. As in the case with the YBCO/Ni/YBCO tape, we find excellent agreement except for the highest modes.

Finally, we show a magnification of the exact dispersion curves for different values of ϕ in Fig. 14. The influence of the anisotropy is clearly seen here as well. This figure, as Fig. 8, only involves the lowest modes and this makes it especially interesting.

IV. CONCLUDING REMARKS

In summary, we derive effective BCs for a thin anisotropic coating on an anisotropic half-space. The BCs are obtained by expanding the traction in the coating in a series through the thickness of the coating and by using the boundary and interface conditions together with the equations of motion in the coating. We then obtain effective approximate BCs by truncating the series expansion. Here we only consider truncation after the linear term in the 3D case and after the quadratic term in the 2D case. The greatest benefit to this approach, compared to using, for example, plate theory, is that it is rational and it is (in principle) straightforward to obtain higher order approximations.

The effective BCs are used to obtain approximate dispersion relations for guided waves in a plate consisting of a thick isotropic core sandwiched between two identical thin anisotropic coatings. Here, we see several benefits in using the effective BCs. For example, we only need to concern ourselves with the wave numbers and polarization vectors of the isotropic core. This simplifies things greatly since in the isotropic case these quantities may be written down explicitly but in the anisotropic case (at least in general) this is not possible. Another benefit is that for the case with the tape we only have to deal with a single-layer plate instead of a three-layer plate. All this simplifies the analytical calculations a great deal and we obtain fairly simple analytical expressions for the approximate dispersion relations. Also the determinantal equations resulting in the dispersion relations are sig-

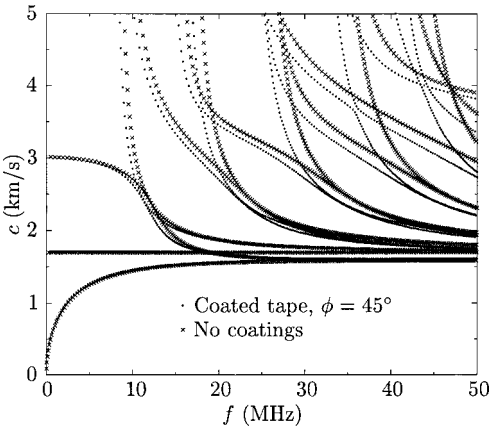


FIG. 11. The exact solutions for the coated Bi-2212/Ag/Bi-2212 tape ($\phi=45°$) and for the uncoated Ag tape.

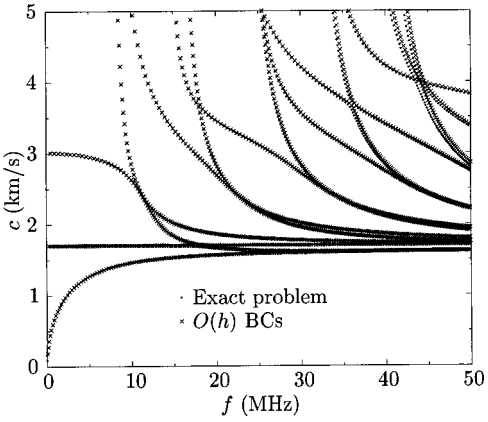


FIG. 13. The solution to the exact problem versus the solution to the approximate problem obtained from the $O(h)$ BCs for $\phi=45°$ for the Bi-2212/Ag/Bi-2212 tape.

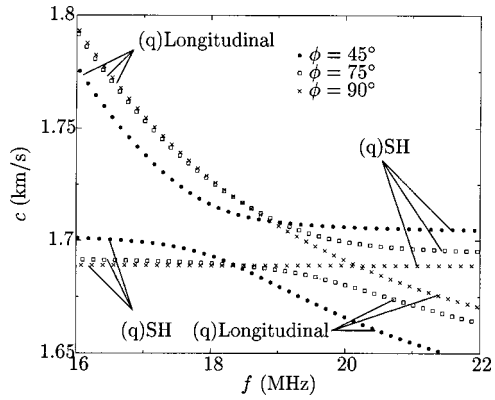


FIG. 14. A magnification of the exact dispersion curves for the Bi-2212/Ag/Bi-2212 tape for different values of ϕ.

nificantly smaller than the exact ones obtained via the global matrix method, resulting in a substantial decrease in the computational effort.

In the numerical examples, two superconducting anisotropic tapes are considered. The core of the first tape is made of (isotropic) nickel and the thin coatings are made of YBCO, which is superconducting and elastically orthotropic. Comparisons between the exact dispersion curves and the ones obtained when the $O(h)$ and the $O(h^2)$ BCs are used are made. In all cases, the agreement is good but, as would be expected, the $O(h^2)$ BCs gives the best approximation. Since the $O(h)$ BCs gives an agreement that is close to the one obtained from the $O(h^2)$ BCs in the frequency interval considered, we believe that in most cases it is sufficient to use these, taking into account that they are considerably simpler (especially if the 3D case is considered). In the second example, the core of the tape is made of (isotropic) silver and the thin coatings are made of Bi-2212, which is superconducting and elastically transversely isotropic. In this case we only compare the exact dispersion curves to the ones obtained from using the $O(h)$ BCs. As in the first case, we find a very good agreement. For both systems, we also compare the exact dispersion curves to the curves obtained for a plate consisting of only the isotropic core. This is done to investigate the influence of the thin coatings. Magnifications of the exact dispersion curves for different orientations of the anisotropic coating materials are shown for both systems. The figures clearly show the coupling of the waves due to the anisotropy of the coatings.

When tapes such as the ones considered in this study are manufactured, large residual stresses are introduced in the coatings due to the mismatch of the thermal expansion coefficients. In the future, we hope to address this issue. It should be possible to include the effects into the effective BCs.

ACKNOWLEDGMENTS

The work of the first author (A.J.N.) has been financially supported by the Swedish Foundation for International Cooperation in Research and Higher Education (STINT) and this is gratefully acknowledged. The work of S.K.D. and M.L.D. was supported by a grant from the Division of Engineering, Office of Basic Energy Sciences, DOE (Contract No. DE-FG03-97ER14738).

APPENDIX A: $O(h^2)$ BOUNDARY CONDITIONS FOR A 2D CASE

Here, we derive the effective BCs for a 2D case. In this case we compute both the h and h^2 coefficients in the series expansion explicitly. The dispersion relations for guided waves in the coated tape (see Sec. II for the tape) are derived in Appendix B.

Assume that the field is propagating in the x_1x_3 plane and that this plane is a plane of elastic symmetry (x_1x_2 is a plane of elastic symmetry as before). This latter assumption results in that the elastic constants $C_{16}=C_{26}=C_{36}=C_{45}=0$. We may then consider in-plane and anti-plane waves, independent of x_2, separately.

The series expansion of the traction for the in-plane part reads

$$(\Sigma_{\alpha 3})_{-h}=(\Sigma_{\alpha 3})_0-h(\partial_3\Sigma_{\alpha 3})_0+h^2(\partial_3^2\Sigma_{\alpha 3})_0/2+O(h^3), \tag{A1}$$

where $\alpha=1,3$. The relevant part of Hooke's law in this case is

$$\begin{pmatrix}\Sigma_{11}\\ \Sigma_{33}\\ \Sigma_{13}\end{pmatrix}=\begin{pmatrix}C_{11} & C_{13} & 0\\ C_{13} & C_{33} & 0\\ 0 & 0 & C_{55}\end{pmatrix}\begin{pmatrix}\partial_1U_1\\ \partial_3U_3\\ \partial_3U_1+\partial_1U_3\end{pmatrix}, \tag{A2}$$

which gives us

$$\begin{aligned}\partial_3U_1&=\Sigma_{13}/C_{55}-\partial_1U_3,\\ \partial_3U_3&=(\Sigma_{33}-C_{13}\partial_1U_1)/C_{33}.\end{aligned} \tag{A3}$$

We now proceed exactly as in Sec. I B to obtain $(\partial_3\Sigma_{\alpha 3})_0$. In order to obtain expressions for $(\partial_3^2\Sigma_{\alpha 3})_0$ we differentiate the equations of motion to obtain

$$\begin{aligned}\partial_3^2\Sigma_{13}&=\rho_c\partial_t^2\partial_3U_1-\partial_1\partial_3\Sigma_{11},\\ \partial_3^2\Sigma_{33}&=\rho_c\partial_t^2\partial_3U_3-\partial_1\partial_3\Sigma_{13}.\end{aligned} \tag{A4}$$

Since differentiation of Hooke's law, Eq. (A2), gives us

$$\begin{aligned}\partial_1\partial_3\Sigma_{11}&=C_{11}\partial_1^2\partial_3U_1+C_{13}\partial_1\partial_3^2U_3,\\ \partial_1\partial_3\Sigma_{13}&=C_{55}\partial_1\partial_3^2U_1+C_{55}\partial_1^2\partial_3U_3,\end{aligned} \tag{A5}$$

and

$$\begin{aligned}\partial_3^2U_1&=\frac{1}{C_{55}}\partial_3\Sigma_{13}-\partial_1\partial_3U_3\\ &=\frac{1}{C_{55}}(\rho_c\partial_t^2U_1-\partial_1\Sigma_{11})-\frac{1}{C_{33}}(\partial_1\Sigma_{33}-C_{13}\partial_1^2U_1),\end{aligned} \tag{A6a}$$

$$\begin{aligned}\partial_3^2U_3&=\frac{1}{C_{33}}\partial_3\Sigma_{33}-\frac{C_{13}}{C_{33}}\partial_1\partial_3U_1\\ &=\frac{1}{C_{33}}(\rho_c\partial_t^2U_3-\partial_1\Sigma_{13})\\ &\quad-\frac{C_{13}}{C_{33}C_{55}}(\partial_1\Sigma_{13}-C_{55}\partial_1^2U_3),\end{aligned} \tag{A6b}$$

we obtain

$$\partial_1\partial_3\Sigma_{11}=\frac{C_{13}}{C_{33}}\rho_c\partial_t^2\partial_1 U_3 + \frac{(C_{11}C_{33}-C_{13}(C_{13}+C_{55}))}{C_{33}C_{55}}\partial_1^2\Sigma_{13} - \frac{(C_{11}C_{33}-C_{13}^2)}{C_{33}}\partial_1^3 U_3, \tag{A7a}$$

$$\partial_1\partial_3\Sigma_{13}=\rho_c\partial_t^2\partial_1 U_1-\frac{C_{13}}{C_{33}}\partial_1^2\Sigma_{33} - \frac{(C_{11}C_{33}-C_{13}^2)}{C_{33}}\partial_1^3 U_1, \tag{A7b}$$

where we have used Hooke's law to express $\partial_1\Sigma_{11}$ in $\partial_1^2 U_1$ and $\partial_1\Sigma_{33}$. Insertion into the series expansion, Eq. (A1), finally gives us the effective BC,

$$\boldsymbol{\tau}+h(A_\sigma\boldsymbol{\tau}+A_u\mathbf{u})+h^2(B_\sigma\boldsymbol{\tau}+B_u\mathbf{u})/2=\mathbf{T}+O(h^3), \quad x_3=0, \tag{A8}$$

where $\boldsymbol{\tau}=(\sigma_{13},\sigma_{33})^T$, $\mathbf{u}=(u_1,u_3)^T$, and $\mathbf{T}=(T_1,T_3)^T$ in this in-plane case. The nonzero elements of the matrices A_σ, A_u, B_σ, and B_u are

$$(A_\sigma)_{12}=\frac{C_{13}}{C_{33}}\partial_1, \tag{A9a}$$

$$(A_\sigma)_{21}=\partial_1, \tag{A9b}$$

$$(A_u)_{11}=-\rho_c\partial_t^2+\left(C_{11}-\frac{C_{13}^2}{C_{33}}\right)\partial_1^2, \tag{A9c}$$

$$(A_u)_{22}=-\rho_c\partial_t^2, \tag{A9d}$$

$$(B_\sigma)_{11}=\frac{1}{C_{55}}\left(\rho_c\partial_t^2-\frac{(C_{11}C_{33}-C_{13}(C_{13}+C_{55}))}{C_{33}}\partial_1^2\right), \tag{A9e}$$

$$(B_\sigma)_{22}=\frac{1}{C_{33}}(\rho_c\partial_t^2+C_{13}\partial_1^2), \tag{A9f}$$

$$(B_u)_{12}=(B_u)_{21}=\frac{1}{C_{33}}(-\rho_c(C_{13}+C_{33})\partial_t^2+(C_{11}C_{33}-C_{13}^2)\partial_1^2)\partial_1. \tag{A9g}$$

We next consider the anti-plane part, i.e., we use the expansion

$$(\Sigma_{23})_{-h}=(\Sigma_{23})_0-h(\partial_3\Sigma_{23})_0+h^2(\partial_3^2\Sigma_{23})_0/2+O(h^3), \tag{A10}$$

to derive the effective BCs. Hooke's law and the equations of motion, in this case, give us

$$\Sigma_{12}=C_{66}\partial_1 U_2, \quad \Sigma_{23}=C_{44}\partial_3 U_2, \quad \partial_1\Sigma_{12}+\partial_3\Sigma_{23}=\rho_c\partial_t^2 U_2. \tag{A11}$$

From these relations, we immediately obtain

$$\partial_3\Sigma_{23}=\rho_c\partial_t^2 U_2-C_{66}\partial_1^2 U_2, \quad \partial_3^2\Sigma_{23}=(\rho_c\partial_t^2\Sigma_{23}-C_{66}\partial_1^2\Sigma_{23})/C_{44}, \tag{A12}$$

and if we use the boundary and interface conditions, we finally obtain

$$\sigma_{23}+(C_{66}\partial_1^2-\rho_c\partial_t^2)(hu_2-h^2\sigma_{23}/2C_{44})=T_2+O(h^3), \quad x_3=0, \tag{A13}$$

from Eq. (A10).

APPENDIX B: GUIDED 2D WAVES IN A PLATE WITH THIN ANISOTROPIC COATINGS

For the in-plane waves, we may write the field in the core by means of potentials as (see Sec. II)

$$u_1=\partial_1\varphi-\partial_3\psi, \quad u_3=\partial_3\varphi+\partial_1\psi, \tag{B1a}$$

$$\varphi=(A_0\sin px_3+B_0\cos px_3)e^{i(kx_1-\omega t)}, \quad \psi=(A_1\sin qx_3+B_1\cos qx_3)e^{i(kx_1-\psi t)}, \tag{B1b}$$

with p, q, k_p, and k_s as before. Applying the boundary conditions Eq. (A8) with a zero on the right-hand side at $x_3=\pm a$, gives us the following dispersion relations (again we split it into a symmetric and an antisymmetric part):

$$\begin{vmatrix}(\alpha_{11}+h^2\gamma_{11})\sin pa+h\beta_{11}\cos pa & (\alpha_{12}+h^2\gamma_{12})\sin qa+h\beta_{12}\cos qa\\ (\alpha_{21}+h^2\gamma_{21})\cos pa+h\beta_{21}\sin pa & (\alpha_{22}+h^2\gamma_{22})\cos qa+h\beta_{22}\sin qa\end{vmatrix}=0, \tag{B2a}$$

for the symmetric part and

$$\begin{vmatrix}(\alpha_{11}+h^2\gamma_{11})\cos pa-h\beta_{11}\sin pa & -(\alpha_{12}+h^2\gamma_{12})\cos qa+h\beta_{12}\sin qa\\ -(\alpha_{21}+h^2\gamma_{21})\sin pa+h\beta_{21}\cos pa & (\alpha_{22}+h^2\gamma_{22})\sin qa-h\beta_{22}\cos qa\end{vmatrix}=0, \tag{B2b}$$

for the antisymmetric part. The coefficients above are given by

$$\alpha_{11}=2\mu kp, \tag{B3a}$$

$$\beta_{11}=(\rho_c\omega^2-(C_{11}+C_{13}(\lambda-C_{13})/C_{33})k^2-C_{13}(\lambda+2\mu)p^2/C_{33})k, \tag{B3b}$$

$$\gamma_{11}=(C_{13}^2(C_{55}-2\mu)k^2-C_{33}(C_{55}-2\mu)(C_{11}k^2-\rho_c\omega^2)-C_{13}C_{55}(2\mu k^2-\rho_c\omega^2))kp/(2C_{33}C_{55}), \tag{B3c}$$

$$\alpha_{12}=\mu(q^2-k^2), \tag{B3d}$$

$$\beta_{12}=(\rho_c\omega^2-(C_{11}-C_{13}(C_{13}+2\mu)/C_{33})k^2)q, \tag{B3e}$$

$$\gamma_{12}=(C_{13}^2((\mu-C_{55})k^2-\mu q^2)k^2 + C_{33}((C_{55}-\mu)k^2+\mu q^2)(C_{11}k^2-\rho_c\omega^2) - C_{13}C_{55}(\rho_c\omega^2+\mu(q^2-k^2))k^2)/(2C_{33}C_{55}), \tag{B3f}$$

$$\alpha_{21}=-\lambda k^2-(\lambda+2\mu)p^2, \tag{B3g}$$

$$\beta_{21}=(\rho_c\omega^2-2\mu k^2)p, \tag{B3h}$$

$$\gamma_{21}=((\lambda+2\mu)(\rho_c\omega^2+C_{13}k^2)p^2 - ((C_{13}+C_{33}-\lambda)\rho_c\omega^2 - (C_{11}C_{33}+C_{13}(\lambda-C_{13}))k^2)k^2)/(2C_{33}), \tag{B3i}$$

$$\alpha_{22}=2\mu kq, \tag{B3j}$$

$$\beta_{22}=(\mu(k^2-q^2)-\rho_c\omega^2)k, \tag{B3k}$$

$$\gamma_{22}=((C_{11}C_{33}-C_{13}(C_{13}+2\mu))k^2 - (C_{13}+C_{33}+2\mu)\rho_c\omega^2)kq/(2C_{33}). \tag{B3l}$$

The solution for the anti-plane wave may be written as

$$u_2=(A_2\cos qx_3+B_2\sin qx_3)e^{i(kx_1-\omega t)}, \tag{B4}$$

with q as before. Applying the BC Eq. (A13) with a zero on the right-hand side at $x_3=\pm a$, gives us the following dispersion relations (again we split it into a symmetric and an antisymmetric part):

$$\mu q(1+h^2(C_{66}k^2-\rho_c\omega^2)/(2C_{44}))\sin qa + h(\rho_c\omega^2-C_{66}k^2)\cos qa=0, \tag{B5a}$$

for the symmetric part and

$$\mu q(1+h^2(C_{66}k^2-\rho_c\omega^2)/(2C_{44}))\cos qa - h(\rho_c\omega^2-C_{66}k^2)\sin qa=0, \tag{B5b}$$

for the antisymmetric part.

[1] G. W. Farnell and E. L. Adler, ''Elastic wave propagation in thin layers,'' in *Physical Acoustics*, edited by W. P. Mason and R. N. Thurston (Academic, New York, 1972), Vol. 9, Chap. 2, pp. 35–127.

[2] M. Bouden and S. K. Datta, ''Rayleigh and Love waves in cladded anisotropic medium,'' ASME Trans. J. Appl. Mech. **57**, 398–403 (1990).

[3] D. A. Sotiropoulos, ''The effect of anisotropy on guided elastic waves in a layered half-space,'' Mech. Mater. **31**, 215–223 (1999).

[4] B. A. Auld, *Acoustic Fields and Waves in Solids* (Krieger, Malabar, FL, 1990), Vols. 1 and 2.

[5] A. H. Nayfeh, *Wave Propagation in Layered Anisotropic Media with Applications to Composites* (Elsevier, Amsterdam, 1995).

[6] Y. Li and R. B. Thompson, ''Influence of anisotropy on the dispersion characteristics of guided ultrasonic plate modes,'' J. Acoust. Soc. Am. **87**, 1911–1931 (1990).

[7] L. P. Solie and B. A. Auld, ''Elastic waves in free anisotropic plates,'' J. Acoust. Soc. Am. **54**, 50–65 (1973).

[8] W. Karunasena, A. H. Shah, and S. K. Datta, ''Wave propagation in a multilayered laminated cross-ply composite plate,'' ASME Trans. J. Appl. Mech. **58**, 1028–1032 (1991).

[9] D. E. Chimenti, ''Guided waves in plates and their use in materials characterization,'' Appl. Mech. Rev. **50**, 247–284 (1997).

[10] S. K. Datta, ''Wave propagation in composite plates and shells,'' in *Comprehensive Composite Materials*, edited by T.-W. Chou (Elsevier, Amsterdam, 2000, to be published), Vol. 1, Chap. 18.

[11] S.-S. Lih and A. K. Mal, ''On the accuracy of approximate plate theories for wave fields calculations in composite plates,'' Wave Motion **21**, 17–34 (1995).

[12] P. Bövik, ''Thin Layers in Elastic Wave Propagation—Effective Modelling and Diffraction,'' Ph.D. thesis, Chalmers University of Technology, Göteborg, Sweden, 1992.

[13] P. Bövik, ''On the modelling of thin interface layers in elastic and acoustic scattering problems,'' Q. J. Mech. Appl. Math. **47**, 17–42 (1994).

[14] P. Bövik, ''A comparison between the Tiersten model and $O(h)$ boundary conditions for elastic surface waves guided by thin layers,'' ASME Trans. J. Appl. Mech. **63**, 162–167 (1996).

[15] E. N. Its and J. S. Lee, ''Propagation of surface waves across an anisotropic layer,'' ASME Trans. J. Appl. Mech. **61**, 596–604 (1994).

[16] W. Huang, S. I. Rokhlin, and Y. J. Wang, ''Analysis of different boundary condition models for study of wave scattering from fiber-matrix interphases,'' J. Acoust. Soc. Am. **101**, 2031–2042 (1997).

[17] S. I. Rokhlin and W. Huang, ''Ultrasonic wave interaction with a thin anisotropic layer between two anisotropic solids: Exact and asymptotic-boundary-condition methods,'' J. Acoust. Soc. Am. **92**, 1729–1742 (1992).

[18] S. I. Rokhlin and Y. J. Wang, ''Equivalent boundary conditions for thin orthotropic layer between two solids: Reflection, refraction, and interface waves,'' J. Acoust. Soc. Am. **91**, 1875–1887 (1992).

[19] S. I. Rokhlin and W. Huang, ''Ultrasonic wave interaction with a thin anisotropic layer between two anisotropic solids. II. Second-order asymptotic boundary conditions,'' J. Acoust. Soc. Am. **94**, 3405–3420 (1993).

[20] J. D. Achenbach, *Wave Propagation in Elastic Solids* (North-Holland, Amsterdam, 1973).

[21] M. Lei, J. L. Sarrao, W. M. Visscher, T. M. Bell, J. D. Thompson, A. Migliori, U. W. Welp, and B. W. Veal, ''Elastic constants of a monocrystal of superconducting $YBa_2Cu_3O_{7-\delta}$,'' Phys. Rev. B **47**, 6154–6156 (1993).

[22] M. Boekholt, J. V. Harzer, B. Hillebrands, and G. Güntherodt, ''Determination of the sound velocities and the complete set of elastic constants for $Bi_2Sr_2CaCu_2O_{8+\delta}$ single crystal using Brillouin light scattering,'' Physica C **179**, 101–106 (1991).

[23] T. H. Ju and S. K. Datta, ''Dynamics of a laminated composite plate with interface layers,'' J. Nondestruct. Eval. **11**, 227–235 (1992).

[24] E. Pan and S. K. Datta, ''Ultrasonic waves in multilayered superconducting plates,'' J. Appl. Phys. **86**, 543–551 (1999).

Acoustic streaming in annular thermoacoustic prime-movers

Vitalyi Gusev
Laboratoire de Physique de l'Etat Condensé, UPRESA-CNRS 6087, Faculté des Sciences and Ecole Nationale Supérieure d'Ingénieurs du Mans, Univesité du Maine, 72085 Le Mans, France

Stéphane Job, Hélène Bailliet, Pierrick Lotton, and Michel Bruneau
Laboratoire d'Acoustique, UMR-CNRS 6613, IAM, Faculté des Sciences, Université du Maine, av. O. Messiaen, 72085 Le Mans Cedex 09, France

(Received 4 February 2000; accepted for publication 12 May 2000)

The theory of acoustic streaming in an annular thermoacoustic prime-mover is developed. It is predicted that above the threshold for traveling wave excitation the device considered (which does not contain any moving parts or externally imposed pressure gradients) produces circulation of fluid. The heat flux carried by this directional mass flow inside the thermoacoustic stack exceeds (or is comparable with) the heat flux associated with the acoustically induced increase of thermal diffusivity of the gas. The effects investigated are important for optimization of the performance of thermoacoustic devices. © *2000 Acoustical Society of America.* [S0001-4966(00)00309-X]

PACS numbers: 43.25.Nm [MFH]

INTRODUCTION

A thermoacoustic prime-mover is a gas-filled acoustic resonator in which a system of solid plates is installed parallel to the resonator axis (i.e., so-called stack), and which is subjected to a temperature gradient by external heating.[1] When the external thermal action on the system exceeds some critical value, gas oscillations start in the absence of any other driver (such as a loudspeaker).[1–4] Thermal energy is transformed into the energy of acoustic oscillations whose amplitude initially grows. Subsequent saturation of sound wave amplitude growth is caused by one of a number of possible nonlinear mechanisms.[2,3,5,6] Among these mechanisms is reverse influence of the acoustic waves on the temperature distribution along the stack. In accordance with Le Chatelier's principle,[7] the excitation of the acoustic waves should induce processes which lead to a reduction of the externally imposed temperature gradient. One of these processes is well understood. In the presence of a temperature gradient the acoustic oscillations will induce effective heat transport via interaction with temperature oscillations.[1] Speaking generally, we can say that in an ideal gas a nonzero enthalpy flow $\rho_m c_p \langle v_x T \rangle$ is induced[1] (where ρ_m denotes the mean density, c_p is the isobaric heat capacity, v_x denotes the axial component of the oscillating particle velocity in the acoustic wave ($\langle v_x \rangle = 0$), and T is the oscillating component of the temperature field ($\langle T \rangle = 0$), $\langle ... \rangle$ standing for time averaging over a period of the oscillation). In the case of a thermoacoustic prime-mover, this flow can cause smoothing of the externally imposed temperature gradient.[2,8,9] Another process which is expected to be induced by the oscillations and which is expected to contribute to heat transport[1] is the excitation of acoustic streaming.[10] A nonzero mass flow $M \equiv (\rho_m v_{xm} + \langle \rho v_x \rangle)$ is generated in the acoustic field as a consequence of various nonlinear processes[10,11] [here, v_{xm} denotes the mean axial component of the particle velocity, ρ is the oscillating component of the density ($\langle \rho \rangle = 0$)]. The role of the associated enthalpy flow $M c_p T_m \equiv \rho_m c_p [v_{xm} + (\langle \rho v_x \rangle / \rho_m)] T_m$ (where T_m is the mean temperature) is much less understood in thermoacoustic theory. Even an order-of-magnitude estimate of the relative importance of the two heat-transfer mechanisms described above (i.e., an estimate of the ratio $M T_m / (\rho_m \langle v_x T \rangle)$, for example) is not available in the literature.

The first experimental observations of the acoustic streaming in a standing wave thermoacoustic device were reported in Ref. 12. A photograph of acoustic streaming in a standing-wave thermoacoustic prime-mover is presented in Ref. 13. Recently, the importance of acoustic streaming in the operation of a pulse-tube refrigerator was demonstrated experimentally.[14] It should be noted that in these experimental configurations[12–14] the total time-averaged mass flow through any cross section of the resonator (and, consequently, through any cross section of thermoacoustic stack) is equal to zero. As the net mass flow along the tube must be zero, then in each cross section of the resonator there exist upward and downward streaming currents.[10,14] In general, this situation is expected to diminish the role of the acoustic streaming in the heat transport along the stack in comparison with the devices where there exists a closed-loop path for steady streaming and where, as a consequence, a nonzero net mass flow is possible. A discussion of how a nonzero net time-averaged mass flow can arise in Stirling and pulse-tube cryocoolers whenever a closed-loop path exists for steady flow can be found in Ref. 15. Recent experiments[16] confirm the existence of these circular (closed-loop) streamings. Moreover, it was demonstrated[16] that suppression of the streaming leads to a significant increase in the efficiency of the device performance. This observation confirms the extremely important role of heat transfer by acoustic streaming in pulse-tube refrigerators (with an existing closed loop for hydrodynamic streaming). It also leads to the conclusion[16] that in the traveling wave thermoacoustic devices (where there exists a closed-loop path for a unidirectional acoustic wave, and, consequently, there is a closed-loop path for a nonzero mass flow)[17] the role of the acoustic streaming could be very important as well. Recently, a traveling wave ther-

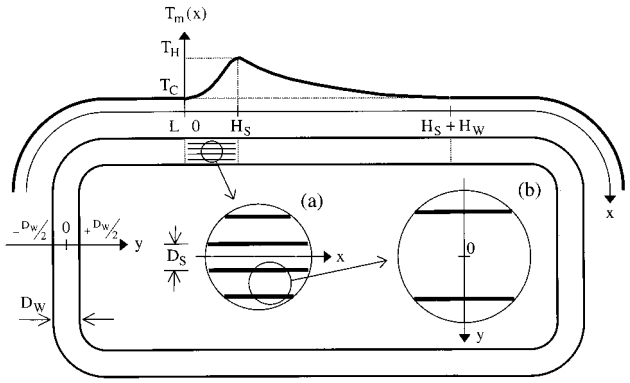


FIG. 1. Schematic presentation of an annular thermoacoustic prime-mover and the gas temperature distribution $T_m(x)$ along its axis (x axis). Here, L denotes the total length of the annular resonator ($0 \leqslant x \leqslant L$), H_S is the length of thermoacoustic stack, H_W is the length of region with decreasing temperature ($dT_m/dx < 0$), and D_W is the width of the waveguide. The inset (a) presents the structure of the stack with plate separation denoted by D_S. The inset (b) presents the coordinate axis y which is used in the theoretical analysis of the thermoacoustic phenomena inside the stack. The comparable coordinate axis for the waveguide part of the resonator is shown on the left-hand side of the figure.

moacoustic prime-mover was created for the first time in practice.[18] The first experimental investigations of a role of the acoustic streaming in operation of an annular thermoacoustic prime-mover has been reported in Ref. 19. These experimental observations[17–19] make it of interest to develop the theory for acoustic streaming in annular thermoacoustic devices.

The present publication addresses the theoretical analysis of acoustic streaming in annular thermoacoustic prime-movers. The schematic presentation of the device together with a qualitative picture of mean temperature $T_m(x)$ distribution along the x axis of the device is given in Fig. 1. It is assumed that the system is above threshold for thermoacoustic instability. The following specific features of this resonator system are taken into account in the analysis of streaming excitation: (1) the existence of spatial regions with very different resistance to hydrodynamic flow (i.e., the stack and the rest of the resonator), (2) the existence of regions with spatially inhomogeneous properties [for example, both the temperature $T_m(x)$ and the density $\rho_m \propto 1/T_m(x)$ are x-dependent inside the region $0 \leqslant x \leqslant H_S + H_W$ in Fig. 1], (3) the existence of regions where the acoustic wave is amplified and regions where it is attenuated (i.e., regions of growing and diminishing wave amplitude). Particular attention is paid to identification of the sources of acoustic streaming. It is demonstrated that volume Reynolds stresses, which can be presented as an x derivative of a potential [for example, $\langle v_x \partial v_x/\partial x \rangle = (1/2)\partial \langle v_x^2 \rangle/\partial x$] and which can play a major role in systems where the path of the acoustic wave and the path of streaming do not coincide completely,[20] are not active in streaming excitation in annular devices. These volume sources are totally compensated by pressure gradients because of the x periodicity of both the acoustic wave and the streaming. On the other hand, such volume sources of streaming as, for example, $\propto \partial \langle v_x v_y \rangle/\partial y$ [where v_y denotes oscillating part of the y component of particle velocity ($\langle v_y \rangle = 0$) and y is the lateral coordinate in Fig. 1], which do not contribute to flow excitation in a plane wave approximation[10,20,21] ($v_y \equiv 0$), can be important in traveling wave devices.

In Sec. I we formulate all the basic assumptions which are necessary to make the problem tractable without losing important physical features of the phenomena. In Sec. II a general solution for the cross-sectional average mass flow in an annular thermoacoustic device is obtained. Section III is devoted to a detailed analysis of the mass flow and the acoustic streaming velocity in the case where the interaction between the acoustic waves and the stack is quasiadiabatic. In Sec. IV the limiting regime of the quasi-isothermal interaction between the sound and the stack is evaluated. Section V presents the results of a comparison of the magnitude of the streaming-induced enthalpy flow with the enthalpy flow associated with the additional acoustically induced thermal conductivity of the gas. This is followed by conclusions.

I. THEORETICAL ASSUMPTIONS

The goal of the present analysis is to gain insight into the physics of the streaming phenomenon in thermoacoustic devices and to derive theoretical relationships which provide reliable order-of-magnitude estimates for the effects considered. To achieve this in the simplest way, a few assumptions concerning the geometry of the resonator and the stack are made. The assumptions concerning the structure of an annular thermoacoustic prime-mover are the following:

(1) The looped tube in Fig. 1 has a rectangular rather than a circular cross section, and we analyze it as a two-dimensional problem neglecting the dependence of the physical quantities on the z coordinate (which is orthogonal to the plane of Fig. 1). In other words, we consider that the annular resonator is composed of two wave-guiding surfaces parallel to the z axis. This assumption provides for a more compact presentation of some of the results in comparison with the case of a cylindrical tube. However, it should be noted that the results obtained can be used for order-of-magnitude estimates for cylindrical tubes as well.

(2) The looped waveguide has an interwall distance D_W which is significantly less than the total length L of the waveguide (W)

$$L \gg D_W. \tag{1}$$

This condition allows us to neglect the influence of channel curvature (i.e., of waveguide looping) on the propagation of sound waves and on the hydrodynamic flow. Because of this assumption the looped coordinate x in Fig. 1 can be straightened in the subsequent analysis. The looping will be taken into account in this paper by imposing periodicity conditions on the physical quantities $\psi(x=L) = \psi(x=0)$ (where ψ stands for an arbitrary physical function of the problem under consideration).

(3) The external action on the system is achieved by the stack (S) in the region $0 \leqslant x \leqslant H_S$ in Fig. 1 and two heat exchangers near its edges (to cool the left edge and to heat the right edge of the stack). The axial dimensions of

the heat exchangers are significantly less than the length, H_S, of the stack. Their influence on the acoustic wave and their hydrodynamic resistance to acoustic streaming are neglected in the analysis (and consequently they are not presented in Fig. 1). In order to achieve good coupling between thermal and acoustic waves, the stack should contain many channels.[1] Consequently, we assume that the width, D_S, of individual channels in the stack is much less than the width of the waveguide

$$D_W \gg D_S. \tag{2}$$

For simplicity we neglect possible blockage of the sound waves and of the streaming by the stack (i.e., we assume that the stack has a porosity equal to 1).

(4) The length of the waveguide is much greater than the stack length, i.e.,

$$L \gg H_S. \tag{3}$$

This corresponds to practical experimental configurations ($L/H_S > 60$ in Ref. 18). Importantly, for the fundamental acoustic resonance (where the acoustic wavelength λ is similar to the resonator length L), Eq. (3) is equivalent to the statement that thermoacoustic stack is acoustically thin. As a consequence, the inequality (3) allows us a simplified analysis of acoustic wave propagation in the stack where the medium is spatially inhomogeneous due to heating and where the traditional representation of the acoustic field as the superposition of two counterpropagating waves is generally invalid.[22]

(5) $$(L/H_S)(D_S/D_W)^2 \ll 1. \tag{4}$$

This relates assumptions (2) and (3) and indicates that the stack provides an important increase in the area of boundary layers (where the interaction between the acoustic and thermal waves mostly takes place) without occupying a significant amount of the resonator volume.

(6) The final assumption concerning the stack dimensions is

$$H_S \gg D_S, \tag{5}$$

which is fulfilled in all thermoacoustic devices.[1–4,18]

(7) The region $0 \leqslant x \leqslant H_S$ of the temperature increase is accompanied by another region with spatially inhomogeneous properties where temperature gradually falls to its ambient value T_C. The characteristic length of this region is denoted by H_W in Fig. 1. In the region $H_S \leqslant x \leqslant H_S + H_W$, the lowering of temperature is caused by heat loss from the gas to the walls of the waveguide. The length H_W can be also fixed by installation of an additional heat exchanger at ambient temperature in the position $x = H_S + H_W$. The following assumptions are made concerning the scale of H_W:

$$L \gg H_W, \tag{6}$$

$$H_W \gg D_W. \tag{7}$$

These are similar to the assumptions expressed in Eq. (3) and Eq. (5), respectively. Note that Eq. (1) follows from Eq. (6) and Eq. (7).

II. GENERAL SOLUTION FOR THE ACOUSTIC STREAMING

The equation describing the lateral variation of the axial streaming velocity v_{xm} in the boundary layer in the presence of an axial temperature gradient was derived by N. Rott[11]

$$\nu_m \frac{\partial^2}{\partial y^2} v_{xm} = \frac{1}{\rho_m} \frac{\partial}{\partial x}(p_h + \rho_m \langle v_x^2 \rangle) + \frac{\partial}{\partial y}(\langle v_x v_y \rangle) - \beta \frac{\nu_m}{T_m} \frac{\partial}{\partial y} \left\langle \left\langle T \frac{\partial}{\partial y} v_x \right\rangle \right\rangle. \tag{8}$$

Here, y is the coordinate normal to the wall (Fig. 1), ν_m is the fluid viscosity evaluated at the mean temperature ($\nu_m \propto T_m^{\beta+1}$), and p_h is the hydrodynamic pressure (which accompanies the streaming). Note that the last term on the right-hand side (rhs) of Eq. (8) describes the excitation of streaming due to the dependence of fluid viscosity on temperature.[11,14] The hydrodynamic pressure p_h does not depend on the y coordinate.

Equation (8) is based on the boundary layer equation and was supposed valid only in the boundary layer.[11] However, the streaming circulating in the annular device has, in fact, the form of a jet (bounded by the plates) and the hydrodynamic equations for a jet are known to have the same form as those for the boundary layer flow[23] (because the important scaling property $\partial/\partial x \ll \partial/\partial y$ is the same in both cases). In the prime-mover we are considering, the scaling property $\partial/\partial x \ll \partial/\partial y$ is effected by the conditions (1), (5), and (7). Consequently, Eq. (8) can be applied to the evaluation of the streaming velocity in the dominant part of the annular device.

The terms which are neglected in Eq. (8) should be taken into account and the Navier–Stokes equation applied for the lateral component v_{ym} of streaming velocity near the cross sections located at $x = 0$, $x = H_S$, and $x = H_S + H_W$, but only if a description giving a smooth transformation of velocity profile in the vicinity of the stack edges and at the boundary between the heated and cold part of the resonator is desirable. However, the most important features of acoustic streaming in annular thermoacoustic prime-movers can be described, when matching the flows at the interfaces between different parts of the device, by using the conditions for the physical quantities averaged over the cross-section of the waveguide. For the present analysis, the most useful of these conditions is continuity of the cross-sectional average mass flow $\overline{M}$

$$\overline{M} \equiv \rho_m \overline{v}_{xm} + \overline{\langle \rho v_x \rangle} \equiv \text{const}. \tag{9}$$

Here, the procedure of averaging over the cross section is defined by $\overline{(\cdots)} \equiv (1/2)\int_{-1}^{1}(\cdots)d\eta$, where the dimensionless coordinate $\eta = 2y/D$ is introduced [$D = D(x)$ is equal to the separation D_W of the walls in the waveguide or equal to the separation D_S of the plates in the stack]. Relationship (9) can be proved by integrating the time-averaged continuity equation $\partial(\rho_m v_{xm} + \langle \rho v_x \rangle)/\partial x + \partial(\rho_m v_{ym} + \langle \rho v_y \rangle)/\partial y = 0$ over the cross section of the channel. Applying the boundary condition that at the walls of each individual channel $v_{ym} = v_y = 0$, we derive $\partial(\rho_m \overline{v}_{xm} + \overline{\langle \rho v_x \rangle})/\partial x = \partial \overline{M}/\partial x = 0$, and,

finally, we get Eq. (9). It describes a rather evident physical fact that under stationary conditions the mean mass flow is the same in the stack and in the waveguide.

The steps in the evaluation of Eq. (8) are: (1) integrate twice over the η coordinate, taking into account the conditions $v_y(\eta=0)=0$, $\partial v_x/\partial\eta(\eta=0)=0$, $v_{xm}(\eta=-1)=0$, and the nondependence of T_m on η; (2) average over the cross section. The result is

$$\frac{\nu_m}{(D/2)^2}\overline{v_{xm}} = \frac{1}{\rho_m}\frac{\partial}{\partial x}\left(\overline{\int_{-1}^{\eta} d\eta' \int_0^{\eta'} d\eta''(p_h+\rho_m\langle v_x^2\rangle)}\right) + \frac{1}{(D/2)}\overline{\int_{-1}^{\eta} d\eta'\left(\langle v_x v_y\rangle - \frac{1}{(D/2)}\beta\frac{\nu_m}{T_m}\left\langle T\frac{\partial}{\partial\eta'}v_x\right\rangle\right)}. \tag{10}$$

After the substitution of $\overline{v_{xm}}$ from Eq. (10) into Eq. (9), it is appropriate to present the result in the form

$$\frac{\nu_m}{(D/2)^2}\overline{M} = \frac{\partial}{\partial x}\left(\overline{\int_{-1}^{\eta} d\eta' \int_0^{\eta'} d\eta''(p_h+\rho_m\langle v_x^2\rangle)}\right) + \frac{\rho_m}{(D/2)} \times \overline{\int_{-1}^{\eta} d\eta'\left(\langle v_x v_y\rangle - \frac{1}{(D/2)}\beta\frac{\nu_m}{T_m}\left\langle T\frac{\partial}{\partial\eta'}v_x\right\rangle\right)} + \frac{\nu_m}{(D/2)^2}\langle\overline{\rho v_x}\rangle. \tag{11}$$

Now, the integration of Eq. (11) with respect to x over the closed loop ($0\leqslant x\leqslant L$) provides an opportunity to eliminate the function

$$\overline{\int_{-1}^{\eta} d\eta' \int_0^{\eta'} d\eta''(p_h+\rho_m\langle v_x^2\rangle)}$$

which includes the unknown pressure distribution p_h. This is possible because of the spatial periodicity of the considered system (the period is equal to the total length L of the annular resonator). Taking into account that [in accordance with Eq. (9)] $\overline{M}$ on the left-hand side of Eq. (11) does not depend on the x coordinate, we obtain

$$\overline{M} = \oint s(x)dx \Big/ \oint (2/D)^2\nu_m\, dx. \tag{12}$$

Here, we call

$$s \equiv \overline{\frac{2\rho_m}{D}\int_{-1}^{\eta} d\eta'\left(\langle v_x v_y\rangle - \frac{2\beta}{D}\frac{\nu_m}{T_m}\left\langle T\frac{\partial v_x}{\partial\eta'}\right\rangle\right) + \frac{4\nu_m}{D^2}\langle\rho v_x\rangle} \tag{13}$$

the density of sources inducing the acoustic streaming. When integrated over x between two cross sections of the resonator, the introduced density of the mass flow sources provides the effective difference in Reynolds stresses between these cross sections. Consequently, the denominator in Eq. (12) can be called the hydrodynamic resistance to the mass flow. If the acoustic field is known, then Eqs. (12) and (13) provide a solution for the circulating mass flow. If the mass flow $\overline{M}$ is found, then the average streaming velocity $\overline{v_{xm}}$ can be evaluated in each particular part of the annular thermoacoustic prime-mover with the help of Eq. (9)

$$\overline{v_{xm}} = \frac{1}{\rho_m}(\overline{M} - \langle\overline{\rho v_x}\rangle). \tag{14}$$

In accordance with Eq. (14), $\overline{v_{xm}}$ (in contrast to $\overline{M}$) depends on the axial coordinate x.

The equation for the steady-state pressure oscillations p in thermoacoustic devices is well-known.[1,24] It is presented here for an ideal gas, neglecting the temperature oscillations in the walls of the tube and in the plates composing the stack

$$\frac{d^2\tilde{p}}{dx^2} + \left\{1+\frac{1}{1-f_\nu}\left[\frac{f_\nu-f_k}{1-\sigma} - T_N\frac{\partial f_\nu}{\partial T_N}\right]\right\}\frac{1}{T_N}\frac{dT_N}{dx}\frac{d\tilde{p}}{dx} + k_c^2\left\{1+\frac{1}{1-f_\nu}[f_\nu+(\gamma-1)f_k]\right\}\frac{1}{T_N}\tilde{p}=0. \tag{15}$$

In Eq. (15) and later on we use the notation $\tilde{\psi}$ for the complex amplitude of the harmonic function ψ, in particular for the pressure oscillations $p\equiv\mathrm{Re}[\tilde{p}\exp(-i\omega t)]$, where ω denotes the cyclic frequency of the oscillations and t is the time. In Eq. (15) T_N denotes the mean temperature normalized to the temperature T_c in the cold homogeneous part of the resonator ($T_N=T_m/T_c$), $k_c=\omega/a_c$ is the adiabatic acoustic wave number in the cold gas in the absence of the solid boundaries, $a_c=\sqrt{\gamma p_m/\rho_c}$ is the adiabatic sound velocity in the cold part of the resonator, γ is the ratio of specific heats, p_m is the mean pressure, and $\rho_c\equiv\rho_m(T_m=T_c)$.

The functions

$$f_\nu \equiv \frac{(1+i)\delta_\nu}{D}\tanh\left[\frac{D}{(1+i)\delta_\nu}\right], \qquad f_k \equiv \frac{(1+i)\delta_k}{D}\tanh\left[\frac{D}{(1+i)\delta_k}\right], \tag{16}$$

characterize the interaction between the acoustic wave and the solid boundaries. In Eq. (16) the width of the viscous boundary layer $\delta_\nu\equiv\sqrt{2\nu_m/\omega}$ and the width of the thermal boundary layer $\delta_k\equiv\sqrt{2k_m/\omega}$ (where k_m is the thermal diffusivity of the gas at mean temperature) depend on the mean temperature because $\nu_m\propto k_m\propto T_m^{\beta+1}$ (β is the phenomenological parameter[25,26]). Consequently, the functions f_ν and f_k depend on the x coordinate. Finally, in Eq. (15) $\sigma\equiv\nu_m/k_m$ denotes the Prandtl number.

It should be mentioned that there is an additional assumption for the validity of Eq. (15) in thermoacoustic prime-movers. In the case of the thermoacoustic prime-movers, it is implicitly assumed that the stabilization (saturation of the growth) of the wave amplitude, when the system is above the threshold of thermoacoustic instability, is not caused by the nonlinear effects in the acoustic wave propagation. The corresponding nonlinear terms (see, for example, Refs. 3, 6, 27–30) are neglected in Eq. (15). The

physical process leading to the stabilization of the acoustic wave amplitude in this case can be described as follows. (1) If the temperature distribution induced in the stack by an external heating is sufficiently steep and the amplification of the acoustic wave inside the stack exceeds its attenuation in the rest of the resonator, the amplitude $|\tilde{p}|$ of the pressure oscillations starts to grow. (2) The rise of the amplitude of the acoustic oscillations increases the acoustically induced thermal conduction[6,8,9,31,32] and the acoustically induced mass flow.[15,16,19] Both these effects (proportional to $|\tilde{p}|^2$ to a first approximation) cause the nonlinear increase in the heat transport between the limits $x=0$ and $x=H_S$ of the thermoacoustic stack. (3) In the case when the external thermal action on the system is fixed (if, for example, the total heat flux supplied to the hot end of the stack is fixed), the increased acoustically induced heat transport through the gas induces a smoothing of the temperature gradient across the stack and a reduction of the sound amplification. (4) The growth of the amplitude of the pressure oscillations and the smoothing of the temperature distribution compensate each other when the total amplification of the acoustic waves in the prime-mover becomes equal to zero. At this point, the steady-state equation (15) is valid.

Consequently, in the case of the thermoacoustic prime-mover, the temperature distribution in Eq. (15) itself depends, in fact, on the acoustic field. For example, in order to find the amplitude of the stationary wave in the system, the equation for the temperature distribution (which includes acoustically induced heat transport) should be added to Eq. (15). In order to study the stability of the stationary oscillations, these coupled equations should be modified to include the transient processes (for example, the process of the acoustic streaming development[10,20,33]). However, both these physical problems are far beyond the scope of the present paper (though they provide clear a perspective for future extensions of the theory).

For the analysis presented below, it is important that Eq. (15) can be used to derive some general conclusions concerning the acoustic streaming even without finding a specific solution. In fact, the solutions for v_x, v_y, T, and ρ in the lowest-order acoustic mode of the resonator (i.e., in the mode which becomes a homogeneous plane wave if the viscosity and the thermal conductivity are negligible) are the known functions of p distribution[24] (see Appendix A). So, we are able to present the density of sources $s(x)$ as a function of $\tilde{p}$, $d\tilde{p}/dx$, and $d^2\tilde{p}/dx^2$, and to estimate the sources of the mass flow in the different parts of the annular device. It is possible to estimate the mass flow and to compare the enthalpy fluxes carried by the directional mass flow and by the acoustically induced thermal conduction.

Using the results (A1)–(A4) presented in Appendix A and the relation $\langle gh\rangle=(1/2)\mathrm{Re}(\tilde{g}\tilde{h}^*)$ (where * denotes the complex conjugate, and g, h are the arbitrary functions), we are able to find all time-averaged nonlinear terms contributing to Eq. (13). For the sources of the mass flow we derive the following general formula:

$$
\begin{aligned}
s=&-\frac{1}{2\omega^2\rho_m}\mathrm{Re}\left\{\frac{1}{T_m}\frac{dT_m}{dx}\left|\frac{d\tilde{p}}{dx}\right|^2\overline{\int_{-1}^{\eta}d\eta'\left\{\left(\frac{\beta+1}{2}\right)\left[\eta' F_\nu-\left(1+\left(\frac{D}{(1+i)\delta_\nu}\right)^2 f_\nu\right)\Phi_\nu-\eta' F_\nu F_\nu^*+\left(1+\left(\frac{D}{(1+i)\delta_\nu}\right)^2 f_\nu\right)\Phi_\nu F_\nu^*\right]\right.}\right.\\
&\overline{+\left[\eta'-\frac{\sigma}{\sigma-1}\Phi_\nu+\frac{1}{\sigma-1}\Phi_k-\eta' F_\nu^*+\frac{\sigma}{\sigma-1}\Phi_\nu F_\nu^*-\frac{1}{\sigma-1}\Phi_k F_\nu^*\right]-\beta\left[\Phi_\nu^*-\frac{\sigma}{\sigma-1}F_\nu\Phi_\nu^*+\frac{1}{\sigma-1}F_k\Phi_\nu^*\right]}\\
&+\frac{d^2\tilde{p}}{dx^2}\frac{d\tilde{p}^*}{dx}\overline{\int_{-1}^{\eta}d\eta'(\eta'-\Phi_\nu-\eta' F_\nu^*+\Phi_\nu F_\nu^*)}\\
&\left.+\left(\frac{\omega}{a}\right)^2\tilde{p}\frac{d\tilde{p}^*}{dx}\overline{\int_{-1}^{\eta}d\eta'[\eta'-\eta' F_\nu^*+(\gamma-1)\Phi_k-(\gamma-1)\Phi_k F_\nu^*+\beta(\gamma-1)\Phi_\nu^*-\beta(\gamma-1)F_k\Phi_\nu^*]}\right\}\\
&+\frac{1}{2\omega^2\rho_m}\mathrm{Re}\left\{\frac{1}{T_m}\frac{dT_m}{dx}\left|\frac{d\tilde{p}}{dx}\right|^2\frac{1}{(\sigma-1)}\left[\frac{(1+i)\delta_\nu}{D}\right]^2\overline{(F_k+F_\nu^*-F_k F_\nu^*)}\right.\\
&\left.+\left(\frac{\omega}{a}\right)^2\tilde{p}\frac{d\tilde{p}^*}{dx}\left[\frac{(1+i)\delta_\nu}{D}\right]^2\overline{[1-F_\nu^*+(\gamma-1)F_\nu-(\gamma-1)F_k F_\nu^*]}\right\}. \qquad (17)
\end{aligned}
$$

In Eq. (17), the last term in curly brackets (which does not contain any other integration except the averaging over the cross section of the channel) describes the contribution $(4\nu_m/D^2)\langle\overline{\rho v_x}\rangle$ to the sources s. We have not done any significant regrouping of the terms in Eq. (17) with the intention of facilitating for the reader a verification of the solution if desired. The only rearrangement during the derivation of Eq. (17) is the separation of the terms proportional to $(1/T_m)(dT_m/dx)|d\tilde{p}/dx|^2$, to $(d^2\tilde{p}/dx^2)(d\tilde{p}^*/dx)$, and to $(\omega/a)^2\tilde{p}(d\tilde{p}^*/dx)$. The reason for this procedure will become clear later.

All the spatial integrals necessary for the evaluation of Eq. (17) are presented in Appendix B. Consequently, Eqs. (12) and (17), together with the results presented in the Ap-

pendix B, provide the description of the circulating mass flow in annular thermoacoustic prime-movers in terms of the temperature distribution and the acoustic pressure distribution inside the resonator.

For example, with the help of Eqs. (B7) and (B8) the solution for the component $\langle \overline{\rho v_x} \rangle$ of the total mass flow can be presented in a compact form

$$\langle \overline{\rho v_x} \rangle = -\frac{1}{2\omega^3 \rho_m} \left\{ \frac{1}{(\sigma^2-1)} \frac{1}{T_m} \frac{dT_m}{dx} \left| \frac{d\tilde{p}}{dx} \right|^2 \mathrm{Im}(f_k + \sigma f_\nu^*) + \left(\frac{\omega}{a} \right)^2 \mathrm{Im}\left[\left(1 - f_\nu^* + \frac{\gamma-1}{\sigma+1}(f_k - f_\nu^*) \right) \tilde{p} \frac{d\tilde{p}^*}{dx} \right] \right\}. \tag{18}$$

As we have described earlier, the distribution of the mean temperature in the system and the acoustic field themselves depend on the mass flow $\overline{M}$ because $\overline{M}$ contributes to the enthalpy transport in the system. So, the derived description obtained in Eqs. (12) and (17) is not, in fact, a solution for the mass flow but just one of the equations in the set describing the thermoacoustic processes in the system. In general, these equations should be solved simultaneously with the equation for the heat transport in the thermoacoustic prime-mover. However, it appears that using Eqs. (12) and (17) we can reach conclusions concerning the direction of the mass flow and the acoustic streaming velocity in the system without needing to solve the general problem.

For this purpose we analyze separately the limiting cases of the quasiadiabatic and of the quasi-isothermal interaction of the acoustic and thermal waves inside the stack ($0 \leqslant x \leqslant H_S$). Note that in both cases under consideration the thermoacoustic interactions in the rest of the resonator are assumed to be quasiadiabatic (in accordance with all known experimental configurations[9,19]).

III. QUASIADIABATIC REGIME

In the quasiadiabatic (QA) regime of the interaction between the acoustic waves and the solid surfaces (which is characterized by the fact that the widths of the viscous and thermal boundary layers are significantly narrower than the distance D between the channel walls, i.e., $\delta_{\nu,k} \ll D_{S,W}$), there is a small parameter in the theory

$$|f_\nu| \approx \sqrt{\sigma}|f_k| \approx \left| \frac{(1+i)\delta_\nu(x)}{D(x)} \right| \equiv \frac{\sqrt{2}\delta_\nu}{D} \propto \frac{\delta_\nu}{D} \ll 1. \tag{19}$$

Retaining in Eq. (15) only those terms not smaller than the first order in the parameter $|f_\nu| \ll 1$, we get

$$\frac{d^2\tilde{p}}{dx^2} + \left[1 - \frac{(\beta+1)}{2}(2c+d) f_\nu \right] \frac{1}{T_N} \frac{dT_N}{dx} \frac{d\tilde{p}}{dx} + k_c^2 (1 + 2c f_\nu) \frac{1}{T_N} \tilde{p} = 0. \tag{20}$$

Here, $c \equiv (1/2)[1 + (\gamma-1)/\sqrt{\sigma}]$ is Kirchhoff's constant and $d \equiv (1/\sqrt{\sigma})[2(1+\beta)^{-1}(1+\sqrt{\sigma})^{-1} - (\gamma-1)]$ is Kramer's constant.[25,26]

In the cold part of the resonator, i.e., where $H_S + H_W \leqslant x \leqslant L$, $dT_N/dx \equiv 0$ ($T_N \equiv 1$) and hence Eq. (20) is additionally simplified

$$\frac{d^2\tilde{p}}{dx^2} + k_c^2 (1 + 2c f_\nu) \tilde{p} = 0. \tag{21}$$

In this cold part of the resonator the most important (leading) contributions both to $\langle \overline{\rho v_x} \rangle$ and to $\overline{M}$ can be found by retaining in Eqs. (18) and (17) only those terms of zero order in the small parameter $|f_\nu| \ll 1$.

$$\langle \overline{\rho v_x} \rangle_c^{\mathrm{QA}} \cong -\frac{1}{2\omega \rho_c a_c^2} \mathrm{Im}\left(\tilde{p} \frac{d\tilde{p}^*}{dx} \right) \cong \frac{1}{2\rho_c a_c^3} |\tilde{p}|^2, \tag{22}$$

where the relation between $d\tilde{p}/dx$ and $\tilde{p}$ is approximated by the form it has in a wave traveling in the positive x direction

$$d\tilde{p}/dx \approx i k_c \tilde{p}. \tag{23}$$

Note that we are using the suffix c (''cold'') for functions and parameters in the homogeneous region of the device ($H_S + H_W \leqslant x \leqslant L$). The description of the density of the mass flow sources following from Eq. (17) for the cold region is

$$s_c^{\mathrm{QA}} \cong \frac{1}{6\omega^2 \rho_c} \mathrm{Re}\left[\frac{d^2\tilde{p}}{dx^2} \frac{d\tilde{p}^*}{dx} + \left(\frac{\omega}{a_c} \right)^2 \tilde{p} \frac{d\tilde{p}^*}{dx} \right]. \tag{24}$$

With the help of Eq. (21) and then Eq. (23) [taking into account the inequality (19)], we obtain the density of the sources of the mass flow in the cold part of the device Eq. (24) in the final form

$$s_c^{\mathrm{QA}} \cong -\frac{c\omega}{3\rho_c a_c^3} \left(\frac{\delta_{\nu c}}{D_W} \right) |\tilde{p}|^2, \tag{25}$$

where $\delta_{\nu c}$ is the thickness of the viscous boundary layer in the cold part of the resonator. Note that the effective sources in the cold part of the resonator are directed in a direction opposite to that of the traveling wave.

For the estimates in the heated part ($0 \leqslant x \leqslant H_S + H_W$) of the thermoacoustic device, we adopt a strategy based on the following assumptions. First, we assume that the heated region is acoustically thin

$$(H_S + H_W)/\lambda \ll 1, \tag{26}$$

where λ is the acoustic wavelength inside the stack. Second, because the heated region should provide the amplification of the acoustic wave, we assume for Eq. (20) that the term related to the temperature gradient (the second term) dominates over the last term. Thus, we approximate Eq. (20) in the inhomogeneously heated region by

$$\frac{d^2\tilde{p}}{dx^2} + \left[1 - \frac{(\beta+1)}{2}(2c+d) f_\nu \right] \frac{1}{T_N} \frac{dT_N}{dx} \frac{d\tilde{p}}{dx} \cong 0. \tag{27}$$

To get a rough condition for the validity of the transition from Eq. (20) to Eq. (27) [i.e., for neglecting terms proportional to $(\omega/a)^2 \propto 1/\lambda^2$], we can approximate $d\tilde{p}/dx$ in the acoustically thin heated region by its value at the stack edge (at $x=0$, for example). The latter can be found from the condition of continuity of the average velocity $\overline{\tilde{v}_x}$ at $x=0$

$[\overline{\tilde{v}_x}(x=0-0)=\overline{\tilde{v}_x}(x=0+0)]$. This condition can be presented as

$$\frac{1}{i\omega\rho_c}(1-f_\nu^W)\frac{d\tilde{p}}{dx}(0-0)=\frac{1}{i\omega\rho_c}(1-f_\nu^S)\frac{d\tilde{p}}{dx}(0+0). \quad (28)$$

We see then that Eqs. (28) and (23) do indeed show that the relation $d\tilde{p}/dx\approx ik_c\tilde{p}$ [Eq. (23)] holds in the first approximation in a QA stack and hence the ratio of the second and the third terms in Eq. (20) can be estimated as $\propto(dT_N/dx)/k_c$. Substituting the crude estimate $(dT_N/dx)\propto[(T_H-T_c)/T_c]/H\equiv(\Delta T/T_c)/H(H\equiv H_{S,W})$, we come to the conclusion that the approximate equation (27) is valid if

$$(\Delta T/T_c)\geqslant k_cH\equiv 2\pi(H/\lambda_c). \quad (29)$$

This inequality holds in thermoacoustic prime-movers because in the experiments $\Delta T\geqslant T_c$ above the threshold[18] and the inhomogeneously heated region is acoustically thin [see Eqs. (3), (6), and (26)].

When condition (29) is satisfied we can also neglect the terms $\propto(\omega/a)^2\propto 1/\lambda^2$ in Eqs. (17) and (18) when evaluating $\langle\overline{\rho v_x}\rangle_h$ and s_h^{QA}. Here, we introduced the suffix h (''heated'') for functions in the heated part of the thermoacoustic device. From Eq. (18)

$$\langle\overline{\rho v_x}\rangle_h^{\mathrm{QA}}\cong\frac{1}{2\omega^3\rho_m}\frac{\sigma+\sqrt{\sigma}+1}{\sqrt{\sigma}(\sqrt{\sigma}+1)(\sigma+1)}\left(\frac{\delta_\nu}{D}\right)\frac{1}{T_N}\left(\frac{dT_N}{dx}\right)\left|\frac{d\tilde{p}}{dx}\right|^2$$
$$\cong\frac{1}{2\omega\rho_ca_c^2}\frac{\sigma+\sqrt{\sigma}+1}{\sqrt{\sigma}(\sqrt{\sigma}+1)(\sigma+1)}\left(\frac{\delta_\nu}{D}\right)\left(\frac{dT_N}{dx}\right)|\tilde{p}|^2$$
$$\propto\mathrm{sign}\left(\frac{dT_N}{dx}\right)\frac{1}{4\pi\rho_ca_c^3}\left(\frac{\delta_{\nu c}}{D}\right)\left(\frac{\lambda_c}{H}\right)|\tilde{p}|^2. \quad (30)$$

The final order-of-magnitude estimate in Eq. (30), obtained making the approximations $\sigma\propto T_N\propto(\Delta T/T_c)\propto 1$, $dT_N/dx\propto(\Delta T/T_c)/H\approx$const, is the most rough.

To find the density of the sources of the mass flow, we derive from Eq. (17)

$$s_h^{\mathrm{QA}}\cong\frac{1}{6\omega^2\rho_m}\mathrm{Re}\left(\frac{1}{T_m}\frac{dT_m}{dx}\left|\frac{d\tilde{p}}{dx}\right|^2+\frac{d^2\tilde{p}}{dx^2}\frac{d\tilde{p}^*}{dx}\right). \quad (31)$$

Note that it was sufficient to retain in Eq. (31) only those terms of zero order in the small parameter $|f_\nu|\ll 1$. First with the aid of Eq. (27), and then using Eq. (23), we transform Eq. (31) into

$$s_h^{\mathrm{QA}}\cong\frac{(c+d/2)}{3\omega^2\rho_m}\left(\frac{\beta+1}{2}\right)\left(\frac{\delta_\nu}{D}\right)\frac{1}{T_m}\frac{dT_m}{dx}\left|\frac{d\tilde{p}}{dx}\right|^2$$
$$\cong\frac{(c+d/2)}{3\rho_ca_c^2}\left(\frac{\beta+1}{2}\right)\left(\frac{\delta_\nu}{D}\right)\frac{dT_N}{dx}|\tilde{p}|^2$$
$$\propto\mathrm{sign}\left(\frac{dT_N}{dx}\right)\frac{1}{3\rho_ca_c^2}\left(\frac{\delta_{\nu c}}{D}\right)\frac{1}{H}|\tilde{p}|^2. \quad (32)$$

In the final order-of-magnitude estimate, we have assumed that $c\propto\beta\propto T_N\propto(\Delta T/T_c)\propto 1$ and $|d/2c|\ll 1$.

The derived expressions (25) and (32) allow us to find the source of the mass flow in the nominator of Eq. (12) in the case of the quasiadiabatic stack (note that the waveguide is always quasiadiabatic)

$$S^{\mathrm{QA}}\equiv\oint s^{\mathrm{QA}}dx$$
$$=\int_0^{H_S}s_h^{\mathrm{QA}}(D=D_S)dx+\int_{H_S}^{H_S+H_W}s_h^{\mathrm{QA}}(D=D_W)dx$$
$$+\int_{H_S+H_W}^{L}s_c^{\mathrm{QA}}(D=D_W)dx$$
$$\cong\frac{|\tilde{p}|^2}{3\rho_ca_c^2}\left\{\left(c+\frac{d}{2}\right)\left(\frac{\beta+1}{\beta-1}\right)\left[\frac{\delta_{\nu c}}{D_S}-\frac{\delta_{\nu c}}{D_W}\right]\left[\left(\frac{T_H}{T_c}\right)^{(\beta-1)/2}-1\right]\right.$$
$$\left.-c\,\frac{2\pi(L-H_S-H_W)}{\lambda_c}\left(\frac{\delta_{\nu c}}{D_W}\right)\right\}. \quad (33)$$

From Eq. (33) it follows that, because of condition (2), the effective Reynolds stresses across the region $dT_m/dx\leqslant 0$ are negligible in comparison with the stresses across the stack. Thus, Eq. (33) can be rewritten in the form

$$S^{\mathrm{QA}}=\oint s^{\mathrm{QA}}dx$$
$$\cong\frac{2\pi c|\tilde{p}|^2}{3\rho_ca_c^2}\left(\frac{\delta_{\nu c}}{D_W}\right)\left\{\frac{1}{2\pi}\left(1+\frac{d}{2c}\right)\left(\frac{\beta+1}{\beta-1}\right)\left(\frac{D_W}{D_S}\right)\right.$$
$$\left.\times\left[\left(\frac{T_H}{T_c}\right)^{(\beta-1)/2}-1\right]-1\right\},$$

where the approximation of Eq. (26) and also that $L\approx\lambda_c$ have been applied. Because typically[25,26] $|d/2c|\ll 1$ and $|(\beta-1)/2|\ll 1$ ($\propto 0.1$ in air), we can additionally simplify the expression as follows:

$$S^{\mathrm{QA}}\equiv\oint s^{\mathrm{QA}}dx$$
$$\cong\frac{2\pi c|\tilde{p}|^2}{3\rho_ca_c^2}\left(\frac{\delta_{\nu c}}{D_W}\right)\left\{\frac{1}{2\pi}\left(\frac{\beta+1}{2}\right)\left(\frac{D_W}{D_S}\right)\ln\left(\frac{T_H}{T_c}\right)-1\right\}. \quad (34)$$

In accordance with Eq. (34) the direction of the net stress difference (inducing the mass flow) depends, in general, on the stack heating. For example, it follows formally from Eq. (34) that the direction of the net stress difference in the case $T_H\to T_c$ is opposite to the direction of the amplified wave (and, consequently, $\bar{M}$ will be directed anticlockwise in Fig. 1). However, in typical thermoacoustic prime-movers, above the threshold the inequality $T_H\geqslant 2T_c$ holds. Then, because of condition (2) the net source of the mass flow is positive. It is controlled by the stack region, and can be approximated by

$$S^{\mathrm{QA}} \cong \frac{(c+d/2)}{3\rho_c a_c^2}\left(\frac{\beta+1}{\beta-1}\right)\left(\frac{\delta_{\nu c}}{D_S}\right)^{\mathrm{QA}} \times\left[\left(\frac{T_H}{T_c}\right)^{(\beta-1)/2}-1\right]|\tilde{p}|^2 \approx \frac{c(\beta+1)}{6\rho_c a_c^2}\left(\frac{\delta_{\nu c}}{D_S}\right)^{\mathrm{QA}} \ln\left(\frac{T_H}{T_c}\right)|\tilde{p}|^2. \tag{35}$$

The superscript QA is used in Eq. (35) to denote that corresponding functions are evaluated in the quasiadiabatic regime. Note the weak (logarithmic) explicit dependence of the streaming source on the maximum normalized heating (T_H/T_c). However, we should not forget that there is also in Eq. (35) an implicit dependence on temperature, because the amplitude of the pressure oscillations depends on the external heating.

The denominator in Eq. (12), which plays a role of the total hydrodynamic resistance for the induced mass flow, can be rewritten in the form

$$\oint\left(\frac{2}{D}\right)^2 \nu_m\, dx = \left(\frac{2}{D_S}\right)^2 \nu_c \int_0^{H_S} T_N^{\beta+1}(x)dx \times\left\{1+\left(\frac{D_S}{D_W}\right)^2 \frac{\int_{H_S}^{H_S+H_W} T_N^{\beta+1}(x)dx}{\int_0^{H_S} T_N^{\beta+1}(x)dx} +\left(\frac{D_S}{D_W}\right)^2 \frac{L}{\int_0^{H_S} T_N^{\beta+1}(x)dx}\right\}. \tag{36}$$

Using a linear model for the temperature distribution in the regions $0\leqslant x\leqslant H_S$ and $H_S\leqslant x\leqslant H_S+H_W$, the second term in the curled brackets in Eq. (36) can be approximated as $\approx(D_S/D_W)^2(H_W/H_S)\ll(D_S/D_W)^2(L/H_S)$. This latter strong inequality follows from condition (6). Using the inequality $T_N\geqslant 1$ the third term in the curled brackets in Eq. (36) is estimated similarly as $\leqslant(D_S/D_W)^2(L/H_S)$. Consequently, because of the assumption (4) both the second and the third terms in Eq. (36) are negligible and hence we see that the dominant contribution to the total hydrodynamic resistance is provided by the stack since

$$\oint\left(\frac{2}{D}\right)^2 \nu_m\, dx = \left(\frac{2}{D_S}\right)^2 \nu_c \int_0^{H_S} T_N^{\beta+1}(x)dx. \tag{37}$$

Combining results (37) and (35), we derive the following solution for the average directional mass flow in the case of the quasiadiabatic stack:

$$\overline{M}^{\mathrm{QA}} \cong \frac{(c+d/2)}{12\pi\rho_c a_c^3}\frac{\left(\frac{\beta+1}{\beta-1}\right)\left[\left(\frac{T_H}{T_c}\right)^{(\beta-1)/2}-1\right]}{\frac{1}{H}\int_0^{H_S} T_N^{\beta+1}(x)dx}\left(\frac{\lambda_c}{H_S}\right) \times\left(\frac{D_S}{\delta_{\nu c}}\right)|\tilde{p}|^2 \propto \frac{1}{12\pi\rho_c a_c^3}\left(\frac{\lambda_c}{H_S}\right)\left(\frac{D_S}{\delta_{\nu c}}\right)^{\mathrm{QA}}|\tilde{p}|^2. \tag{38}$$

The latter rough, order-of-magnitude estimate is obtained under the assumptions $c\propto\beta\propto T_N\propto(\Delta T/T_c)\propto 1$. Equation (38), together with the solutions (22) and (30), provides in accordance with Eq. (14) a description of the streaming velocity. However, a comparison of $\overline{M}^{\mathrm{QA}}$ [Eq. (38)] with $\langle\overline{\rho v_x}\rangle^{\mathrm{QA}}$ both in the cold [Eq. (22)] and in the heated [Eq. (30)] parts of the annular prime-mover demonstrates that in the QA regime $\overline{M}^{\mathrm{QA}}$ significantly exceeds $\langle\overline{\rho v_x}\rangle^{\mathrm{QA}}$ everywhere in the device $[\overline{M}^{\mathrm{QA}}/\langle\overline{\rho v_x}\rangle_c^{\mathrm{QA}}\propto(1/6\pi)(L/H_S)(D_S/\delta_{\nu c})\gg 1$, $\overline{M}^{\mathrm{QA}}/\langle\overline{\rho v_x}\rangle_h^{\mathrm{QA}}\geqslant(D_S/\delta_{\nu c})^2\gg 1$, because of assumptions (3) and (19)]. In these latter estimates we have omitted for compactness the factors of the order of 1 contributed by the combinations of the dimensionless parameters c, β, γ, and σ (all of which are of the order of 1). Consequently, in the QA regime of the interaction between the acoustic and the thermal waves inside the stack, the velocity of the acoustic streaming can be approximated by

$$v_{xm}^{\mathrm{QA}} \cong \overline{M}^{\mathrm{QA}}/\rho_m. \tag{39}$$

The very rough order-of-magnitude estimate $\overline{v}_{xm}^{\mathrm{QA}}\propto(1/12\pi)\times(\lambda_c/H_S)(D_S/\delta_{\nu c})^{\mathrm{QA}}(|\tilde{v}_x|^2/a_c)$ which can be used for the streaming velocity follows from Eq. (39), Eq. (38), and the relation $|\tilde{p}|\approx a_c\rho_c|\tilde{v}_x|$ (approximately valid in the QA regime). This estimate mainly demonstrates the dependence of the streaming velocity on the acoustic thickness of the stack $[(\lambda_c/H_S)\gg 1$ because of the condition (26)] and on the parameter of the adiabaticity $[(D_S/\delta_{\nu c})\gg 1$ because of the assumption (19)] when the amplitude of the acoustic oscillations is fixed. As an example, we have estimated that, for the reported magnitude of the oscillating component of particle velocity $|\tilde{v}_x|\approx 3$ m/s in the annular prime-mover described in Ref. 18, this theoretical result predicts unidirectional streaming with a characteristic velocity $\overline{v}_{xm}^{\mathrm{QA}}\approx 0.1$ m/s.

IV. QUASI-ISOTHERMAL REGIME

In the quasi-isothermal (QI) regime of the interaction between the acoustic waves and the surfaces of the stack (which is characterized by the fact that the width of the viscous and thermal boundary layers is significantly larger than the distance between the channel walls $\delta_{\nu,k}\gg D_S$), there is a further small parameter in the theory

$$(D_S/\delta_\nu)^2\ll 1. \tag{40}$$

Under condition (40) the functions $f_{\nu,k}$ [Eq. (16)] controlling the thermoacoustic interaction inside the stack can be approximated by

$$f_\nu \cong 1+i\frac{1}{6}\left(\frac{D_S}{\delta_\nu}\right)^2-\frac{1}{30}\left(\frac{D_S}{\delta_\nu}\right)^4-i\frac{17}{9\cdot 8\cdot 7\cdot 5}\left(\frac{D_S}{\delta_\nu}\right)^6+\cdots, \qquad f_k \cong 1+i\frac{\sigma}{6}\left(\frac{D_S}{\delta_\nu}\right)^2-\frac{\sigma^2}{30}\left(\frac{D_S}{\delta_\nu}\right)^4-i\frac{17\sigma^3}{9\cdot 8\cdot 7\cdot 5}\left(\frac{D_S}{\delta_\nu}\right)^6+\cdots. \tag{41}$$

Retaining in Eq. (15) only the predominant terms important for the subsequent analysis, we may rewrite it in the form

$$\frac{d^2\tilde{p}}{dx^2}-(\beta+1)\frac{1}{T_N}\frac{dT_N}{dx}\frac{d\tilde{p}}{dx}+ik_c^2(6\gamma)\left(\frac{\delta_\nu}{D}\right)^2\frac{1}{T_N}\tilde{p}=0. \tag{42}$$

For the net amplification, the effects inside the stack caused by the presence of the temperature gradients should dominate over the effects of thermoviscous attenuation described by the last term in Eq. (42). The estimate of the pressure gradient inside the acoustically thin QI stack [which follows from the boundary condition (28) and the relation (23) in the cold part of the resonator] is

$$d\tilde{p}/dx \propto -6k_c(\delta_{\nu c}/D_S)^2\tilde{p}. \tag{43}$$

With the aid of Eq. (43), we can compare the second and the third term in Eq. (42) and we arrive at the conclusion that under condition (29) (where $H=H_S$) the wave equation inside the QI stack can be approximated by

$$\frac{d^2\tilde{p}}{dx^2}-(\beta+1)\frac{1}{T_N}\frac{dT_N}{dx}\frac{d\tilde{p}}{dx}\cong 0. \tag{44}$$

Under the assumption (29) of the dominance of the dT_N/dx-dependent effects, we neglect the terms $\propto(\omega/a)^2 \propto 1/\lambda^2$ in Eqs. (17) and (18) when evaluating $\langle\overline{\rho v_x}\rangle_h^{\mathrm{QI}}$ and s_h^{QI}. From Eq. (18) [using the expansions (41) to evaluate the functions presented in Appendix B], we obtain

$$\begin{aligned}\langle\overline{\rho v_x}\rangle_h^{\mathrm{QI}} &\cong \frac{17\sigma}{10\cdot 9\cdot 8\cdot 7}\frac{1}{\omega^3\rho_m}\left(\frac{D_S}{\delta_\nu}\right)^6\frac{1}{T_N}\frac{dT_N}{dx}\left|\frac{d\tilde{p}}{dx}\right|^2\\ &\cong \frac{17\sigma}{280}\frac{1}{\omega\rho_c a_c^2}\left(\frac{D_S}{\delta_\nu}\right)^2\frac{dT_N}{dx}|\tilde{p}|^2\\ &\propto \frac{17}{140}\frac{1}{\pi\rho_c a_c^2}\left(\frac{D_S}{\delta_{\nu c}}\right)^2\frac{\lambda_c}{H_S}|\tilde{p}|^2.\end{aligned} \tag{45}$$

The second form in Eq. (45) is obtained with the aid of Eq. (43). The third form in Eq. (45) is an order-of-magnitude estimate under the crude assumptions $\sigma\propto T_N\propto(\Delta T/T_c)\propto 1$. The solution for the density of the mass flow sources in the QI stack, obtained from an evaluation of Eq. (17), is

$$\begin{aligned}s_h^{\mathrm{QI}} \cong -\frac{1}{\omega^2\rho_m}\frac{1}{9\cdot 8\cdot 7\cdot 5}\left(\frac{D_S}{\delta_\nu}\right)^4 \mathrm{Re}\Bigg(\frac{1}{T_N}\frac{dT_N}{dx}\left|\frac{d\tilde{p}}{dx}\right|^2\\ \times[9(\beta+1)-8\beta\sigma-17\sigma]-\frac{d^2\tilde{p}}{dx^2}\frac{d\tilde{p}^*}{dx}\Bigg).\end{aligned}$$

Using Eqs. (44) and (43), we can transform this latter solution into the final form

$$\begin{aligned}s_h^{\mathrm{QI}} &\cong \frac{(8\beta+17)\sigma}{9\cdot 8\cdot 7\cdot 5}\frac{1}{\omega^2\rho_m}\left(\frac{D_S}{\delta_\nu}\right)^4\frac{1}{T_N}\frac{dT_N}{dx}\left|\frac{d\tilde{p}}{dx}\right|^2\\ &\cong \frac{(8\beta+17)\sigma}{70}\frac{1}{\rho_c a_c^2}\frac{dT_N}{dx}|\tilde{p}|^2.\end{aligned}$$

Consequently, the difference in the Reynolds stresses across the QI stack contributing to the total source of the mass flow in Eq. (12) can be found

$$S_h^{\mathrm{QI}} \equiv \int_0^{H_S} s_h^{\mathrm{QI}}dx \cong \frac{(8\beta+17)\sigma}{70}\frac{1}{\rho_c a_c^2}\frac{\Delta T}{T_c}|\tilde{p}|^2. \tag{46}$$

Using Eq. (33), the source of the mass flow in the inhomogeneously heated quasiadiabatic region $H_S\leqslant x\leqslant H_S+H_W$ is estimated to be roughly a factor of $(\delta_{\nu c}/D_W)^{\mathrm{QA}}$ smaller than the source described by Eq. (46) and, consequently, since this factor is $\ll 1$, this source of mass flow can be neglected. Then, the net Reynolds stresses inducing the mass flow can be presented in the form

$$\begin{aligned}S^{\mathrm{QI}} &\equiv \int_0^{H_S} s_h^{\mathrm{QI}}dx+\int_{H_S+H_W}^{L} s_c^{\mathrm{QA}}\,dx\\ &\cong \frac{2\pi c}{3\rho_c a_c^2}\left(\frac{\delta_{\nu c}}{D_W}\right)^{\mathrm{QA}}\Bigg\{\frac{3(8\beta+17)\sigma}{140\pi c}\left(\frac{D_W}{\delta_{\nu c}}\right)^{\mathrm{QA}}\\ &\quad\times\left[\frac{T_H}{T_c}-1\right]-1\Bigg\}|\tilde{p}|^2.\end{aligned} \tag{47}$$

It follows from Eq. (47) that the direction of the net source of the mass flow depends, in general, on the stack heating. However, in typical thermoacoustic prime-movers it is true that $(\Delta T/T_c)\geqslant 1$ above the threshold and, as a result [because of the inequality (19) $(\delta_{\nu c}/D_W)^{\mathrm{QA}}\ll 1$] the net source of the mass flow is approximately equal to the difference in the Reynolds stresses across the stack [Eq. (46)]. Consequently, the flow is directed clockwise. Substituting Eqs. (46) and (37) into Eq. (12), we derive the solution for the mass flow

$$\overline{M}^{\mathrm{QI}} \cong \frac{(8\beta+17)\sigma}{280\pi\rho_c a_c^3}\left(\frac{\lambda_c}{H_S}\right)\left[\left(\frac{D_S}{\delta_{\nu c}}\right)^2\right]^{\mathrm{QI}}\frac{\left(\frac{T_H}{T_c}-1\right)}{\frac{1}{H_S}\int_0^{H_S}T_N^{\beta+1}(x)dx}|\tilde{p}|^2. \tag{48}$$

A comparison of the solution (48) with solutions (22), (30), and (45) for $\langle\overline{\rho v_x}\rangle$ in various parts of the device demonstrates that the direction of the acoustic streaming velocity $\overline{v}_{xm}$ [Eq. (14)] can be different in the different parts of the resonator. It is always clockwise in the heated quasiadiabatic region $H_S\leqslant x\leqslant H_S+H_W$, because here, in accordance with Eq. (30), $\langle\overline{\rho v_x}\rangle$ is negative (while $\overline{M}^{\mathrm{QI}}$ is always positive). In the cold part of the resonator the direction of $\overline{v}_{xm}$ depends on the magnitude of the parameter $(\lambda_c/H_S)(D_S/\delta_{\nu c})^2$, which in turn is controlled by the inequalities (26) and (40). By substituting Eqs. (48) and (45) into Eq. (14), we find that inside the QI stack

$$\begin{aligned}\overline{v}_{xm}^{\mathrm{QI}} &\cong \frac{\sigma}{280\pi\rho_m\rho_c a_c^3}\left(\frac{\lambda_c}{H_S}\right)\left[\left(\frac{D_S}{\delta_{\nu c}}\right)^2\right]^{\mathrm{QI}}\\ &\quad\times\frac{\left\{8\beta+17\left[1-\frac{1}{H_S}\int_0^{H_S}T_N^{\beta+1}(x)dx\right]\right\}}{\frac{1}{H_S}\int_0^{H_S}T_N^{\beta+1}(x)dx}\left(\frac{T_H}{T_c}-1\right)|\tilde{p}|^2.\end{aligned} \tag{49}$$

Note that the term in the square bracket in Eq. (49) is negative for $\beta>-1$, because the inequality $T_N\geqslant 1$ holds inside the stack. Consequently, both the magnitude and the direction of the velocity in Eq. (49) depend significantly on the temperature distribution inside the stack.

V. COMPARISON OF THE ACOUSTICALLY INDUCED AND STREAMING-INDUCED ENTHALPY FLOW

The solutions for the mass flow $\overline{M}$ [the general solution (12) and (17), and the solutions in particular limiting cases (38) and (48)] provide an opportunity to describe the associated enthalpy flow $\overline{M}c_pT_m$ in the resonator. Though, in principle, this enthalpy flow is induced originally by the acoustic waves we call it here the *streaming-induced* enthalpy flow in order to distinguish it from the enthalpy flow $\rho_m c_p\langle\overline{v_xT}\rangle$, for which we retain the term *acoustically induced* enthalpy flow.

Using the descriptions of the velocity and temperature oscillations presented in Appendix A [Eqs. (A1) and (A3)], the acoustically induced enthalpy flow can be presented in the form

$$\rho_m c_p\langle\overline{v_xT}\rangle=\frac{c_pT_m}{2\omega^3\rho_m}\left\{\frac{1}{T_m}\frac{dT_m}{dx}\left|\frac{d\tilde{p}}{dx}\right|^2\times\mathrm{Im}\left(\overline{\frac{1}{\sigma-1}F_k-\frac{\sigma}{\sigma-1}F_\nu-F_\nu^*-\frac{1}{\sigma-1}F_kF_\nu^*}\right)-(\gamma-1)\left(\frac{\omega}{a}\right)^2\times\mathrm{Im}\left[\overline{(1-F_k-F_\nu^*+F_kF_\nu^*)}\tilde{p}\frac{d\tilde{p}^*}{dx}\right]\right\}. \tag{50}$$

With the aid of Eqs. (B7) and (B8) from Appendix B, we can present Eq. (50) in the form

$$\rho_m c_p\langle\overline{v_xT}\rangle=\frac{c_pT_m}{2\omega^3\rho_m}\left\{\frac{1}{\sigma^2-1}\frac{1}{T_m}\frac{dT_m}{dx}\left|\frac{d\tilde{p}}{dx}\right|^2\mathrm{Im}(f_k+\sigma f_\nu^*)-(\gamma-1)\left(\frac{\omega}{a}\right)^2\mathrm{Im}\left[\left(1-\frac{1}{\sigma+1}f_k-\frac{\sigma}{\sigma+1}f_\nu^*\right)\tilde{p}\frac{d\tilde{p}^*}{dx}\right]\right\}. \tag{51}$$

This solution coincides with one presented in Ref. 1 [Eq. (A30)] (for the case of an ideal gas and negligible temperature oscillations of the walls and stack).

In general, the expression for the acoustically induced enthalpy flow (51) should be compared with the solution for the streaming-induced enthalpy flow which follows from Eqs. (12) and (17). Here, to make the comparison of $\rho_m c_p\langle\overline{v_xT}\rangle$ and $\overline{M}c_pT_m$ for inside the thermoacoustic stack ($0\leqslant x\leqslant H_S$), we use the approach proposed in previous sections (Secs. III and IV). Applying condition (29), we retain in expression (51) only the terms induced by the inhomogeneous heating [i.e., we neglect the terms $\propto(\omega/a)^2\propto 1/\lambda^2$]. Then, a comparison of this approximate solution with expression (18) for $\langle\overline{\rho v_x}\rangle$ (similarly simplified) provides

$$\rho_m c_p\langle\overline{v_xT}\rangle\cong -c_pT_m\langle\overline{\rho v_x}\rangle. \tag{52}$$

Consequently, a comparison of the streaming-induced and of the acoustically induced enthalpy flows $\overline{M}c_pT_m$ and $\rho_m c_p\langle\overline{v_xT}\rangle$ can be obtained comparing the total mass flow $\overline{M}$ with its component $\langle\overline{\rho v_x}\rangle$. This has been done earlier in Secs. III and IV. Using the results in these sections and Eq. (52), we conclude that in the quasiadiabatic regime of the thermoacoustic interaction inside the stack

$$\left|\frac{\overline{M}c_pT_m}{\rho_m c_p\langle\overline{v_xT}\rangle}\right|^{\mathrm{QA}}\cong\left|\frac{\overline{M}^{\mathrm{QA}}}{\langle\overline{\rho v_x}\rangle_h^{\mathrm{QA}}}\right|\propto\left[\left(\frac{D_S}{\delta_{\nu c}}\right)^2\right]^{\mathrm{QA}}\gg 1. \tag{53}$$

Thus, in the QA regime the streaming-induced enthalpy flow significantly exceeds the acoustically induced enthalpy flow.

In the quasi-isothermal regime of the thermoacoustic interaction inside the stack

$$\left|\frac{\overline{M}c_pT_m}{\rho_m c_p\langle\overline{v_xT}\rangle}\right|^{\mathrm{QI}}\cong\left|\frac{\overline{M}^{\mathrm{QI}}}{\langle\overline{\rho v_x}\rangle_h^{\mathrm{QI}}}\right|\propto 1+\frac{8\beta}{17}\propto 1.$$

Consequently, in this QI regime the streaming-induced and the acoustically induced enthalpy flows are comparable in magnitude. More precise formulas comparing $\overline{M}c_pT_m$ and $\rho_m c_p\langle\overline{v_xT}\rangle$ can be obtained straightforwardly from the solutions derived in the previous sections, if required.

It is important to note that under the assumption of condition (29) the net enthalpy flow inside the thermoacoustic stack (i.e., the sum of the streaming-induced and the acoustically induced enthalpy flows) is controlled only by the hydrodynamic velocity $\overline{v}_{xm}$ of the acoustic streaming. In fact, with the aid of Eqs. (14) and (52) we obtain

$$\overline{M}c_pT_m+\rho_m c_p\langle\overline{v_xT}\rangle\cong c_pT_m(\overline{M}-\langle\overline{\rho v_x}\rangle)=c_pT_m\rho_m\overline{v}_{xm}=c_pT_c\rho_c\overline{v}_{xm}. \tag{54}$$

This result underlines the importance of the evaluation of the streaming velocity $\overline{v}_{xm}$ carried out in Secs. II, III, IV. Of particular significance, Eq. (49) demonstrates the dependence of the direction of the net enthalpy flow inside the QI stack on the heating of the stack.

VI. CONCLUSIONS

We have investigated the possibility of the generation of acoustic streaming in an annular thermoacoustic prime-mover. The theory developed predicts that an annular thermoacoustic prime-mover above the threshold of traveling wave excitation provides circulation of fluid (without using any moving parts and without an externally induced pressure gradient). The solutions show that the direction of the total cross-sectional average mass flow $\overline{M}$ and the direction of the cross-sectional average streaming velocity $\overline{v}_{xm}$ coincide with the direction of propagation of the amplified acoustic wave (clockwise in Fig. 1) in the case of the quasiadiabatic thermoacoustic interaction inside the stack. In the quasi-isothermal stack, the direction of $\overline{v}_{xm}$ depends on the level of stack heating and the parameter β [see Eq. (49)].

It is demonstrated that inside the stack during operation the net cross-sectional average flow of the enthalpy [i.e., the sum of the streaming-induced and acoustically induced enthalpy flows, Eq. (54)] is controlled by the streaming velocity $\overline{v}_{xm}$. In particular, the results of Secs. III and V [see Eqs. (39) and (53)] show that the enthalpy flow carried by the streaming significantly exceeds the enthalpy flux associated

with the acoustically induced thermal conductivity of the gas. Thus, the acoustic streaming is predicted to play the dominant role in the enthalpy transport.

It should be noted that the analytical predictions obtained for the quasiadiabatic regime are currently the most interesting for applications, because they can also be used for order-of-magnitude estimates for the intermediate regime $D_S \geqslant \delta_{\nu,k}$ (which has been the most practical for thermoacoustic devices so far[1]). In fact, the lowest threshold for development of thermoacoustic instability in an annular thermoacoustic prime-mover was reported[18] to be for $D_S \approx 3\delta_k$. In Ref. 1, from both analytical and numerical analysis the conclusion is drawn that ''a practical engine can be expected... to have all available cross-sectional area filled with plates spaced from $2\delta_k$ to $4\delta_k$.''

Finally, it can be expected that the theory developed here will provide the necessary background for a theoretical modeling of a traveling wave thermoacoustic device having a quasi-isothermal stack as realized experimentally in Ref. 19. However, clearly, for this application the theory will require modification to take into account the additional acoustic elements in the setup of Ref. 19 that are not in the basic scheme (Fig. 1) we have considered.

ACKNOWLEDGMENTS

This work was supported by DGA (Contract No. 99.34.072). We acknowledge the constructive criticism of the preliminary versions of the developed theory by the referees.

APPENDIX A: THE RELATIONS OF THE OSCILLATIONS OF THE PARTICLE VELOCITY, OF THE TEMPERATURE AND OF THE DENSITY IN THE ACOUSTIC FIELD WITH THE PRESSURE OSCILLATIONS

The components $\tilde{v}_x, \tilde{v}_y$ of the particle velocity, the complex amplitudes $\tilde{T}, \tilde{\rho}$ of the temperature, and the density variations can be presented as function of the pressure oscillations $\tilde{p}$ and its axial gradient $d\tilde{p}/dx$ as follows:[1,24]

$$\overline{v_x} = \frac{1}{i\omega\rho_m}[1-F_\nu]\frac{d\tilde{p}}{dx}, \tag{A1}$$

$$\tilde{v}_y = -\frac{1}{i\omega\rho_m}\left(\frac{D}{2}\right)\left\{\frac{d}{dx}\left[(\eta-\Phi_\nu)\frac{d\tilde{p}}{dx}\right] + \left(\eta - \frac{\sigma}{\sigma-1}\Phi_\nu + \frac{1}{\sigma-1}\Phi_k\right)\frac{1}{T_m}\frac{dT_m}{dx}\frac{d\tilde{p}}{dx} + \frac{\omega^2}{a^2}[\eta+(\gamma-1)\Phi_k]\tilde{p}\right\}, \tag{A2}$$

$$\tilde{T} = -\frac{1}{\omega^2\rho_m}\left[1-\frac{\sigma}{\sigma-1}F_\nu + \frac{1}{\sigma-1}F_k\right]\frac{dT_m}{dx}\frac{d\tilde{p}}{dx} + \frac{1}{\rho_m c_p}[1-F_k]\tilde{p}, \tag{A3}$$

$$\tilde{\rho} = -\frac{\rho_m}{T_m}\tilde{T} + \frac{\gamma}{a^2}\tilde{p} = \frac{1}{\omega^2}\left[1-\frac{\sigma}{\sigma-1}F_\nu + \frac{1}{\sigma-1}F_k\right] \times \frac{1}{T_m}\frac{dT_m}{dx}\frac{d\tilde{p}}{dx} + \frac{1}{a^2}[1+(\gamma-1)F_k]\tilde{p}. \tag{A4}$$

Here, $\rho \equiv \nu_m/k_m$ is the Prandtl number, a is the speed of sound (adiabatic), γ is the ratio of isobaric to isochoric specific heats. The functions $F_{\nu,k}$ and $\Phi_{\nu,k}$ which describe the transverse distribution of the acoustic field in the channels are defined by

$$F_{\nu,k} \equiv \frac{\cosh\left[\frac{D}{(1+i)\delta_{\nu,k}}\eta\right]}{\cosh\left[\frac{D}{(1+i)\delta_{\nu,k}}\right]}, \tag{A5}$$

$$\Phi_{\nu,k} \equiv \left[\frac{(1+i)\delta_{\nu,k}}{D}\right]\frac{\sinh\left[\frac{D}{(1+i)\delta_{\nu,k}}\eta\right]}{\cosh\left[\frac{D}{(1+i)\delta_{\nu,k}}\right]}, \tag{A6}$$

and, consequently,

$$F_{\nu,k} = \partial\Phi_{\nu,k}/\partial\eta. \tag{A7}$$

Note that the functions $f_{\nu,k}$ defined in Eq. (16) are particular values of the functions $\Phi_{\nu,k}$, i.e., $f_{\nu,k} = \Phi_{\nu,k}\ (\eta = 1)$. We remind here that $\eta \equiv 2y/D$.

APPENDIX B: SOME SPATIAL INTEGRALS OF THE FUNCTIONS CHARACTERIZING THE THERMOACOUSTIC INTERACTION

The following integrals contribute to solution (17) for the sources of the directional mass flow:

$$\overline{\int_{-1}^{\eta} d\eta'\,\Phi_{\nu,k}} = -\left[\frac{(1+i)\delta_{\nu,k}}{D}\right]^2(1-f_{\nu,k}), \tag{B1}$$

$$\overline{\int_{-1}^{\eta} d\eta'\,\eta' F_\nu} = 2\left[\frac{(1+i)\delta_\nu}{D}\right]^2(1-f_\nu) - f_\nu, \tag{B2}$$

$$\overline{\int_{-1}^{\eta} d\eta'\,\eta' F_\nu F_\nu^*} = f_\nu f_\nu^* - \frac{1}{2}(f_\nu + f_\nu^*) + \frac{1}{2}\left[\frac{(1+i)\delta_\nu}{D}\right]^2(f_\nu - f_\nu^*), \tag{B3}$$

$$\overline{\int_{-1}^{\eta} d\eta'\,\Phi_\nu F_\nu^*} = -\frac{1}{2}f_\nu f_\nu^* - \frac{1}{2}\left[\frac{(1+i)\delta_\nu}{D}\right]^2(1-f_\nu^*), \tag{B4}$$

$$\overline{\int_{-1}^{\eta} d\eta'\,\Phi_k F_\nu^*} = -\frac{1}{(\sigma+1)}f_k f_\nu^* - \frac{1}{(\sigma+1)}\left[\frac{(1+i)\delta_\nu}{D}\right]^2 \times\left(1-\frac{\sigma-1}{\sigma+1}f_k - \frac{2}{\sigma+1}f_\nu^*\right). \tag{B5}$$

Note that it is sufficient to assume in Eq. (B5) that $\sigma=1$ in order to get Eq. (B4). In order to get from Eq. (B5) the formula for

$$\overline{\int_{-1}^{\eta} d\eta'\ \Phi_\nu^* F_k},$$

it is sufficient to interchange $\nu \leftrightarrow k$ (and, consequently, σ should be replaced by $1/\sigma$) and take the complex conjugate

$$\overline{\int_{-1}^{\eta} d\eta' \Phi_\nu^* F_k} = -\frac{\sigma}{(\sigma+1)} f_k f_\nu^* + \frac{1}{(\sigma+1)} \times \left[\frac{(1+i)\,\delta_\nu}{D}\right]^2 \left(1+\frac{\sigma-1}{\sigma+1} f_\nu^* - \frac{2\sigma}{\sigma+1} f_k\right). \tag{B6}$$

The simplest integral contributing to Eq. (17) is

$$\overline{\int_{-1}^{\eta} \eta'\, d\eta'} = -\frac{1}{3}.$$

For an evaluation of the contribution of the term $(4\nu_m/D^2)\langle\overline{\rho v_x}\rangle$ to the mass flow source, we also use

$$\overline{F_k F_\nu^*} = \frac{\sigma}{\sigma+1} f_k + \frac{1}{\sigma+1} f_\nu^*, \tag{B7}$$

$$\overline{F_{\nu,k}} = f_{\nu,k}. \tag{B8}$$

[1] G. W. Swift, ''Thermoacoustic engines,'' J. Acoust. Soc. Am. **84**, 1145–1180 (1988).
[2] J. Wheatley, T. Hofler, G. W. Swift, and A. Migliori, ''Understanding of simple phenomena in thermoacoustics with applications to acoustical heat engines,'' Am. J. Phys. **53**, 147–162 (1985).
[3] A. A. Atchley, H. E. Bass, and T. J. Hofler, ''Development of nonlinear waves in a thermoacoustic prime mover,'' in *Frontiers in Nonlinear Acoustics: Proceedings of the 12th ISNA*, edited by M. F. Hamilton and D. T. Blackstock (Elsevier Science, London, 1990), pp. 603–608.
[4] G. Müller and W. Lauterborn, ''Experiments with thermoacoustic oscillator physical and musical,'' in *Proceedings of the International Symposium on Musical Acoustics* (Le Normont, Dourdan, France, 1995), pp. 178–183.
[5] V. A. Müller and E. Lang, ''Experimente mit thermisch getriebenen gasflüssigheits-schwingungen,'' J. Appl. Math. Phys. (ZAMP) **36**, 358–366 (1985).
[6] V. E. Gusev, H. Bailliet, P. Lotton, and M. Bruneau, ''Asymptotic theory of nonlinear acoustic waves in a thermoacoustic prime-mover,'' Acust. Acta Acust. **86**, 25–38 (2000).
[7] L. D. Landau, A. I. Akhiezer, and E. M. Lifshitz, *General Physics (Mechanics and Molecular Physics)* (Pergamon, Oxford, 1967).
[8] D. D. Hebert and A. A. Atchley, ''Measurements of the evolution of the temperature profile in a parallel plate stack,'' J. Acoust. Soc. Am. **100**, 2846 (1986).
[9] T. Yazaki, A. Tominauga, and Y. Narahara, ''Large heat transport due to spontaneous oscillation induced in a tube with steep temperature gradients,'' J. Heat Transfer **105**, 889–894 (1983).
[10] O. V. Rudenko and S. I. Soluyan, *Theoretical Foundations of Nonlinear Acoustics* (Consultants Bureau, New York, 1977).
[11] N. Rott, ''The influence of heat conduction on acoustic streaming,'' J. Appl. Math. Phys. (ZAMP) **25**, 417–421 (1974).
[12] D. F. Gaitan, A. Gopinath, and A. A. Atchley, ''Experimental study of acoustic turbulence and streaming in a thermoacoustic stack,'' J. Acoust. Soc. Am. **96**, 3220(A) (1994).
[13] T. Yazaki and A. Tominaga, ''Measurement of sound generation in thermoacoustic oscillations,'' Proc. R. Soc. London, Ser. A **454**, 2113–2122 (1998).
[14] J. R. Olson and G. W. Swift, ''Acoustic streaming in pulse tube refrigerators: Tapered pulse tubes,'' Cryogenics **37**, 769–776 (1997).
[15] D. Gedeon, ''DC gas flows in Stirling and pulse-tube Cryocoolers,'' in *Cryocoolers*, edited by R. G. Ross (Plenum, New York, 1997), Vol. 9, pp. 385–392.
[16] G. W. Swift, D. L. Gardner, and S. Backhaus, ''Acoustic recovery of lost power in pulse tube refrigerators,'' J. Acoust. Soc. Am. **105**, 711–724 (1999).
[17] P. H. Ceperley, ''Gain and efficiency of a short traveling wave heat engine,'' J. Acoust. Soc. Am. **77**, 1239–1244 (1985), and references therein.
[18] T. Yazaki, A. Iwata, T. Maekawa, and A. Tominaga, ''Traveling wave thermoacoustic engine in a looped tube,'' Phys. Rev. Lett. **81**, 3128–3131 (1998).
[19] S. Backhaus and G. W. Swift, ''A thermoacoustic Stirling heat engine,'' Nature (London) **399**, 335–338 (1999).
[20] O. V. Rudenko and A. A. Sukhorukov, ''Nonstationary Eckart streaming and pumping of liquid in an ultrasonic field,'' Akust. Zh. **44**, 653–658 (1998) [Sov. Phys. Acoust. **44**, 565–570 (1998)].
[21] J. E. Piercy and J. Lamb, ''Acoustic streaming in liquids,'' Proc. R. Soc. London, Ser. A **226**, 43–50 (1954).
[22] L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Media 1: Plane and Quasiplane Waves* (Springer, Berlin, 1991).
[23] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Pergamon, Oxford, 1982).
[24] N. Rott, ''Thermoacoustics,'' Adv. Appl. Mech. **20**, 135–175 (1980).
[25] N. Rott, ''Damped and thermally driven acoustic oscillations in wide and narrow tubes,'' Z. Angew. Math. Phys. **20**, 230–243 (1969).
[26] H. A. Kramers, ''Vibrations of a gas column,'' Physica (Amsterdam) **15**, 971–978 (1949).
[27] A. Larraza and A. A. Atchley, ''Analysis of a thermoacoustic prime mover above onset of self-oscillation,'' J. Acoust. Soc. Am. **94**, 1773 (1993).
[28] H. L. Kurtz, D. T. Blackstock, ''Attenuation of intense sinusoidal waves in air-saturated, bulk porous materials,'' J. Acoust. Soc. Am. **81**, 1723–1731 (1987).
[29] R. F. Lambert and J. D. Mcintosh, ''Nonlinear wave propagation through rigid porous materials. II. Approximate analytical solutions,'' J. Acoust. Soc. Am. **88**, 1950–1959 (1990).
[30] J. R. Olson and G. W. Swift, ''Energy dissipation in oscillating flow through straight and coiled pipes,'' J. Acoust. Soc. Am. **100**, 2123–2131 (1996).
[31] G. Mozurkewich, ''Time average temperature distribution in a thermoacoustic stack,'' J. Acoust. Soc. Am. **103**, 380–388 (1998).
[32] G. Mozurkewich, ''A model for transverse heat transfer in thermoacoustics,'' J. Acoust. Soc. Am. **103**, 3318–3326 (1998).
[33] V. E. Gusev and O. V. Rudenko, ''Nonsteady quasi-one-dimensional acoustic streaming in unbounded volumes with hydrodynamic nonlinearity,'' Akust. Zh. **25**, 875–881 (1979) [Sov. Phys. Acoust. **25**, 493–497 (1979)].

Measurement of a third-order elastic constant in silicon by a comparison method

J.-Y. Duquesne[a)]
L.M.D.H., Université Pierre et Marie Curie/CNRS (UMR 7603), Boîte 86, 4 place Jussieu, 75252 Paris Cedex 05, France

(Received 15 December 1999; accepted for publication 19 May 2000)

The third-order elastic constant C_{111} of silicon is measured with a comparison method, based on the interaction between longitudinal acoustic waves. The value is in good agreement with published values and demonstrates the validity of the method. © *2000 Acoustical Society of America.* [S0001-4966(00)01209-1]

PACS numbers: 43.25.Zx [MAB]

INTRODUCTION

A new method for the measurement of the nonlinear elastic properties of liquids or solids was recently proposed.[1] This method studies the interaction of longitudinal acoustic waves propagating successively in a liquid and in a solid. The nonlinear elastic properties of both media are then compared. This allows either the measurement of the nonlinear elastic constant of the liquid (if the nonlinear elastic constant of the solid is known) or, conversely, the measurement of a nonlinear elastic constant of the solid, in the relevant direction (if the nonlinear elastic constant of the liquid is known). This method is sensitive to both sign and magnitude of the nonlinear elastic constants and requires no measurement of the absolute amplitude of the fundamental and harmonic waves. We present here a refinement of this method, allowing a better sensitivity, and we apply it to a silicon/H_2O system.

I. THEORY

Figure 1 is a sketch of the setup. The notations are defined in Ref. 1. At the transducer T_2, the magnitude of the harmonic wave (with frequency $F_3=F_1+F_2$ or F_1-F_2) produced by the interaction of two fundamental longitudinal waves (with frequencies F_1 and F_2) propagating successively in the liquid and the solid sample, along the $I_1I_2T_2$ path, is

$$p_3'^{[I]}(z)=C\left(f(z)+K\frac{\beta_S}{\beta_L}g(z)\right), \tag{1}$$

where β_S and β_L depend on the linear and nonlinear properties of the liquid and the solid sample, respectively. We get

$$\frac{\beta_S}{\beta_L}=-\frac{V_L}{V_S}\frac{A^2}{\tilde{C}_{11}^2}\frac{3\tilde{C}_{11}+\tilde{C}_{111}}{2A+B}. \tag{2}$$

A and B are the linear and nonlinear elastic constants of the liquid. $\tilde{C}_{11}$ and $\tilde{C}_{111}$ are the second-order and the Brugger's third-order elastic constants of the solid, relevant to the propagation direction. C is proportional to β_L and K depends on T (the transmission coefficient for pressure waves from the liquid to the solid sample) and on the acoustic attenuation in the liquid and the solid sample,

$$K=T\frac{\alpha_{L_3}-\alpha_{L_2}-\alpha_{L_1}}{\alpha_{S_3}-\alpha_{S_2}-\alpha_{S_1}}(1-\exp[(\alpha_{S_3}-\alpha_{S_2}-\alpha_{S_1})d]). \tag{3}$$

$f(z)$ and $g(z)$ only depend on the acoustic attenuation in the liquid,

$$f(z)=\exp[-(\alpha_{L_2}+\alpha_{L_3})z]-\exp[-(\alpha_{L_1}+2\alpha_{L_2})z], \tag{4}$$

$$g(z)=\exp[-(\alpha_{L_1}+2\alpha_{L_2})z]. \tag{5}$$

The attenuation in the liquid versus frequency F can be expressed as

$$\alpha_L=\gamma F^2. \tag{6}$$

In principle, the value of $K\beta_S/\beta_L$ can be found by fitting the experimental curves $p_3'^{[I]}(z)$, with the theoretical law (1). This has been performed in fused silica.[1] However, in case of low frequencies (and then low attenuation in the solid sample) or short solid samples, we get

$$K\simeq\pm 2Td\gamma F_1F_2 \tag{7}$$

(+ or − for difference and sum frequencies harmonic waves, respectively). At low fundamental frequencies or for short solid samples K goes to zero. In those cases, the $g(z)$ term in (1) turns out to be negligible with respect to the $f(z)$ term. Physically, the interaction along the $I_1I_2T_2$ path is more effective in the liquid than in the solid. The shape of the $p_3'^{[I]}(z)$ curve is then mainly determined by the attenuation in the liquid. Then, at low fundamental frequencies or for short solid samples, fitting $p_3'^{[I]}(z)$ with Eq. (1) leads to a poor accuracy on $K\beta_S/\beta_L$.

To overcome this drawback, we suggest in the present paper to consider, in addition, the interaction of the fundamental longitudinal waves along the I_2T_2 path. A straightforward calculation shows that the magnitude of the resulting harmonic wave at the receiving transducer T_2 is

$$p_3'^{[II]}(z)=-CK\frac{\beta_S}{\beta_L}\frac{1}{t_2}\sqrt{\frac{1-t_2}{1-t_1}}\exp(-\alpha_{L_1}z), \tag{8}$$

where t_1 and t_2 are the power transmission factor at the liquid/buffer and liquid/sample interfaces, respectively.

a)Electronic mail: duquesne@ccr.jussieu.fr

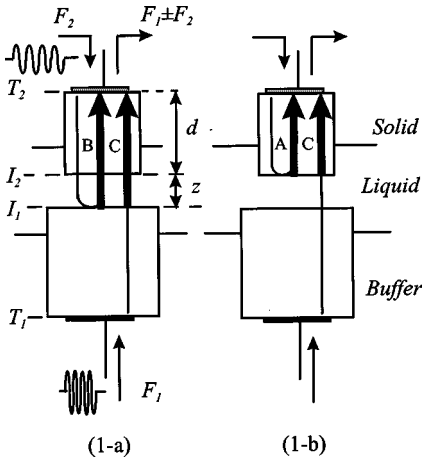


FIG. 1. Sketch of the setup. The time delay between the F_1 and F_2 pulses is adjusted to promote either the interaction of the acoustic pulses B and C [interaction along the $I_1I_2T_2$ path / harmonic wave $p_3'^{[I]}$ / Fig. 1(a)] or the interaction of the acoustic pulses A and C [interaction along the I_2T_2 path / harmonic wave $p_3'^{[II]}$ / Fig. 1(b)].

Along the I_2T_2 path, the fundamental waves only interact in the solid and $p_3'^{[II]}(z)$ does not depend on β_L [Eq. (8) is in accordance with this point since C is proportional to β_L].

We assume that the attenuation in the liquid and the transmission coefficients t_1 and t_2 are measured independently. Then, fitting both curves $p_3'^{[I]}(z)$ and $p_3'^{[II]}(z)$ with the same set of parameters C and $K\beta_S/\beta_L$ provides a good accuracy on $K\beta_S/\beta_L$. Roughly speaking, if $K\beta_S/\beta_L$ is small, the fit of $p_3'^{[I]}(z)$ sets C and the fit of $p_3'^{[II]}(z)$ sets $CK\beta_S/\beta_L$, from which $K\beta_S/\beta_L$ can be accurately determined.

II. EXPERIMENT

To illustrate the previous calculation, we present here an experiment performed in a H_2O/Si system. The Si sample is a single crystal, with a low impurity content (dopant: B, resistivity $>100\ \Omega$ cm). Two (100) faces are polished, flat, and parallel. The acoustic waves propagates parallel to the [100] axis, then: $\tilde{C}_{11}=C_{11}$ and $\tilde{C}_{111}=C_{111}$. The length of the sample is $d=9.941$ mm. The experiments are performed at 21.5 °C. The mass densities of H_2O and Si were found in literature: 997 and 2329 kg m^{-3}, respectively. The velocities in H_2O and Si[100] were measured: 1488 and 8431 m s^{-1}, respectively. The power transmission factor at the liquid/buffer and liquid/sample interfaces was measured: $t_1=0.378$, $t_2=0.263$. The pressure transmission factor at the liquid/sample interface is deduced from the acoustic impedances Z_L and Z_2 of the liquid and of the sample: $T=2Z_2/(Z_L+Z_2)=1.86$. The frequencies of the fundamental waves are $F_1=102$ and $F_2=45$ MHz and we investigate the $F_3=F_1+F_2=147$ MHz harmonic wave. The diameter of the transducers is 3 mm.

Figure 2 shows the magnitude of the harmonic waves $p_3'^{[I]}$ and $p_3'^{[II]}$ versus the liquid thickness z. The two experimental curves are fitted with the same set of parameters: C, $K\beta_S/\beta_L$, and γ. The agreement between the theoretical and experimental curves is very good. We find: $K\beta_S/\beta_L=-3.040\,10^{-3}$ and $\gamma=2.365\,10^{-5}$ mm^{-1} MHz^{-2}. The

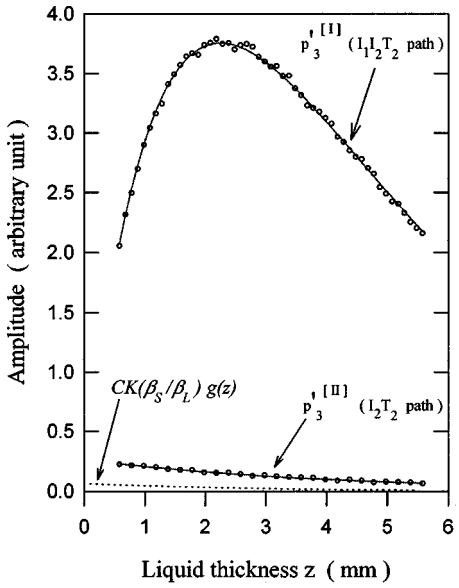


FIG. 2. Magnitude of the harmonic waves $p_3'^{[I]}$ and $p_3'^{[II]}$ in the H_2O/Si[100] system. Fundamental waves: $F_1=102$, $F_2=45$ MHz. Harmonic wave: $F_3=F_1+F_2=147$ MHz. Circles: experimental points. Full lines: fit of the experimental data, using Eqs. (1), (8), and $K\beta_S/\beta_L=-3.040\,10^{-3}$, $\gamma=2.365\,10^{-5}$ mm^{-1} MHz^{-2}. Dashed line: $g(z)$ contribution to $p_3'^{[I]}$ [Eq. (1)].

value of C is irrelevant. Figure 2 also displays the $(CK\beta_S/\beta_L)g(z)$ term in Eq. (1) computed from the preceding values. It can be seen that this term is small and cannot be extracted with good accuracy from the fit of $p_3'^{[I]}$ alone.

To extract the value of β_S/β_L, the value of K must be known. Then, the attenuation in the solid must be measured. The uncertainty on the attenuation in the solid sample is the main source of uncertainty on C_{111}. However, the acoustic attenuation in Si in this frequency range is very low so that the value of K is not very sensitive to the attenuation. From our measurements, we estimate $0<(\alpha_{S_3}-\alpha_{S_2}-\alpha_{S_1})<1.15\ 10^{-2}$ mm^{-1} (1 dB cm^{-1}) in the whole frequency range. Then we derive $K=-4.12\pm0.12$ and $\beta_S/\beta_L=(7.37\pm0.21)\cdot10^{-4}$. Using[2] $B/A=5.03$, we deduce the value of C_{111} in silicon: $C_{111}=-860\pm10$ GPa.

III. CONCLUSION

The value of C_{111} deduced from our experiment is in good agreement with published values: -825, -795, -834 GPa.[3–5] This demonstrates the validity of the comparison method. As compared to the currently used method based on the harmonic generation,[6] the present method requires no measurement of the absolute amplitude of the fundamental and harmonic waves and is sensitive to the sign of the third-order elastic constants.[1] Additional experiments along the symmetry axis [111] and [110] could give two other combinations of third-order elastic constants:[6] $(C_{111}+6C_{112}+2C_{123}+12C_{144}+24C_{155}+16C_{456})$ and $(C_{111}+3C_{112}+12C_{155})$. The complete set of third-order elastic constants in Si could then be derived from those results and from published measurements on the variation of ultrasonic velocities with hydrostatic pressure.[3]

[1] J.-Y. Duquesne and D. A. Parshin, J. Appl. Phys. **82**, 3275–3280 (1997).
[2] T. R. Beyer, in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, New York, 1998), p. 34.
[3] H. J. McSkimin and P. Andreatch, J. Appl. Phys. **35**, 3312–3319 (1964).
[4] J. J. Hall, Phys. Rev. **161**, 756–761 (1967).
[5] L. Philip and M. A. Breazeale, J. Appl. Phys. **52**, 3383–3387 (1981).
[6] M. A. Breazeale and J. Philip, in *Physical Acoustics*, edited by W. P. Mason and R. N. Thurston (Academic, New York, 1984), Vol. XVII, Chap. 1.

Effective impedance of rough porous ground surfaces

Keith Attenborough[a)]
School of Engineering, The University of Hull, Hull HU6 7RX, United Kingdom

Tim Waters-Fuller
School of Built Environment, Napier University, 219 Colinton Road, Edinburgh, Scotland, United Kingdom

(Received 20 May 1999; revised 6 June 2000; accepted 19 June 2000)

Laboratory measurements at grazing angles of the impedance of sand surfaces containing artificial two-dimensional (2D) roughness are compared with predictions of an extended boss theory, which allows for incoherent scatter, randomness, and finite impedance effects. The data and predictions are found to be in reasonable agreement. The rough surface impedance model may be used, as an alternative to layering, to explain measurements where the reactance exceeds the resistance at low frequencies. The model also explains instances where the measured ground impedance tends to zero at higher frequencies. In addition, the model is found to predict features of measurements made with an elevated source and vertically separated receivers at a range of 30 m over cultivated ground surfaces. © *2000 Acoustical Society of America.* [S0001-4966(00)04609-9]

PACS numbers: 43.28.En, 43.28.Fp [LCS]

I. INTRODUCTION

Many natural outdoor ground surfaces give appreciable attenuation of audio-frequency sound from compact sources in excess of that due to wave front spreading and atmospheric absorption. Often, such surfaces are rough as well as porous. In this paper, an extended boss model that allows for incoherent scatter is used to investigate the influence of roughness scales small compared with the smallest wavelength of interest, on the effective admittance of porous surfaces. The boss model due to Biot/Tolstoy allows for the prediction of propagation from a point source over rough finite impedance surfaces through an effective admittance.[1] However, it does not include incoherent scatter effects. Lucas and Twersky's boss model[2] includes incoherent scatter but was developed initially for acoustically hard surfaces and restricted to scatterers with semicircular or semielliptical cross section. A heuristic extension of this model to arbitrary scatterer shapes has been found to enable reasonable agreement with data over various rough hard surfaces, as long as the mean roughness height and spacing are small compared to a wavelength.[3] For a mean scatterer spacing comparable to the incident wavelengths, a further heuristic extension to include diffraction grating effects has been found useful.[3] As a result of studying the attenuation and properties of the boundary wave over randomly rough *hard* surfaces near grazing incidence, Medwin *et al.*[4] have deduced empirical adjustments to the Biot/Tolstoy boss model involving average height and root-mean-square (rms) slope for steep-sloped roughness elements. In the boss models considered, the plane of the effective admittance is placed at the base of the scatterers, i.e., at the imbedding plane. Studies of ultrasonic surface wave propagation over periodically corrugated *hard* surfaces, have shown that the Biot/Tolstoy boss theory gives good agreement with data only if the assumed depth of the corrugation is less than the actual depth.[5] In the other studies of grazing incidence propagation over *soft* rough surfaces,[6] using an approximate three-dimensional (3D) version of the Biot/Tolstoy boss theory,[1] it has been found necessary to use an impedance plane at the top of the scatterers to achieve best agreement with data. However, these aspects are not pursued here.

In Sec. II, we review previous formulations of the boss model for 2D roughness and present a new heuristic form that enables predictions of the effective admittance of rough finite impedance surfaces and includes incoherent scatter. In Sec. III, the resulting predictions are compared with data obtained over rough sand surfaces in the laboratory. In Sec. IV, the predicted influences of roughness are found to be consistent with data obtained over several surfaces outdoors.

II. EFFECTIVE ADMITTANCE FORMULATIONS USING BOSS MODELS

According to the Biot/Tolstoy boss model,[1] the effective normalized admittance β^* of an acoustically soft surface containing regularly spaced 2D scatterers with the same density and sound speed as the surrounding medium, is given by the sum of the smooth surface impedance and a correction for roughness. Hence

$$\beta^* = -ik_0 \cos^2(\varphi)\varepsilon_{01}^* + \beta, \tag{1}$$

where

$$\begin{aligned}
&\varepsilon_{01}^* = a_{01}^* - b_{01}^*, \\
&a_{01}^* = 2V\left[\frac{\rho_1-\rho_0}{\rho_1+\rho_0}\right]\frac{s_{2D}}{v_{2D}}, \quad b_{01}^* = V(1-m_1/m_0), \\
&s_{2D} = \frac{\rho_0+\rho_1}{\rho_1+K\rho_0}s_2, \quad v_{2D} = 1+\frac{2\pi V}{3Nb^2}\left[\frac{\rho_1-\rho_0}{\rho_1+\rho_0}\right]s_{2D}, \\
&s_2 = \frac{1}{2}(1+K), \quad m_i = \frac{1}{\rho_i c_i^2}, \quad i = 0,1.
\end{aligned} \tag{2}$$

a)Author to whom correspondence should be addressed.

Here, β represents the normalized admittance, $=\rho_0 c_0/\rho_1 c_1$ at the smooth interface between a lower fluid half-space with density and sound speed ρ_1 and c_1, and an upper fluid half-space characterized by ρ_0 and c_0. N is the number of cylindrical scatterers per unit length. The cylindrical scatterers have height h (h represents the rms height if the scatterers are not identical), center-to-center spacing b, and occupy a total cross-sectional area V per unit length above the plane. The incident wavelength is assumed to be greater than h or b. Specifically $k_0 h \ll 1$ and $k_0 b \leqslant 1$, where $k_0 = \omega/c_0$, the wave number in the upper fluid half-space. K is a hydrodynamic drag factor for steady flow past a scatterer (equal to 0.5 for a hemisphere, 1 for a semicylinder and the eccentricity for an elliptic semicylinder), φ is the angle between the source-receiver axis and the normal to the scatterer axis.

If the embedding material and scatterers are rigid and porous, then they can be treated, acoustically, as if they are fluids but with complex densities and sound speeds. In addition, if $|\rho_2| \gg \rho_1$, as is the case, for example, in air-saturated sand, and shape and close interaction effects are ignored, Eqs. (1) and (2) may be approximated by[1]

$$\beta^* = -ik_0 \cos^2(\varphi) V + \beta(1 - ik_1 V), \tag{3}$$

where k_1 is the complex wave number within the imbedding material. For an air-filled high-flow resistivity porous medium and at low frequencies,[7,8] $\beta k_1 \approx \gamma\Omega$, where γ is the ratio of specific heats in air and Ω is the porosity. Hence Eq. (3) implies that the presence of small-scale roughness only changes the imaginary part of the surface admittance.

However, these results ignore incoherent scatter. To include incoherent scatter, we combine results deduced from Tolstoy for the rough interface between two fluids[1] and results deduced from Lucas and Twersky[2] for a hard rough surface.[3] These results allow for random spacing also. Hence the effective normalized surface admittance of a porous surface containing 2D roughness becomes

$$\beta^* = \eta - ik_0 \cos^2(\varphi)\varepsilon_{01}^* + \beta, \tag{4}$$

where η is given by

$$\eta(\alpha,\varphi) \approx \frac{k_0^3 b V^2}{2}(2 - W^2)\left\{(1 - \sin^2\alpha \sin^2\varphi) \times \left[1 + \left(\frac{\delta^2}{2}\cos^2\varphi - \sin^2\varphi\right)\sin^2\alpha\right]\right\} + O(k_0^5), \tag{5}$$

$(1-W)^2$ is a packing factor introduced for random distributions, $W = b^*/b$, b^* is the minimum (center-to-center) separation between two scatterers, and b is the mean spacing. $\delta = 2s_{2D}/v_{2D}$, s_{2D} and v_{2D} being calculated from (2) or the corresponding approximations for $|\rho_2| \gg \rho_1$, k_0 is the wave number in air and α is the angle of incidence from the normal to the plane. Note that the predicted effective admittance is a function of the observer geometry. For $\varphi = 0$, $\alpha \neq \pi/2$ (normal to scatterer axes, but nongrazing incidence),

$$\beta^* = \frac{k_0^3 b V^2}{2}(1 - W^2)\left[1 + \left(\frac{\delta^2}{2}\right)\sin^2\alpha\right] - ikV[\delta \sin^2\alpha - 1 + \gamma\Omega] + \beta. \tag{6}$$

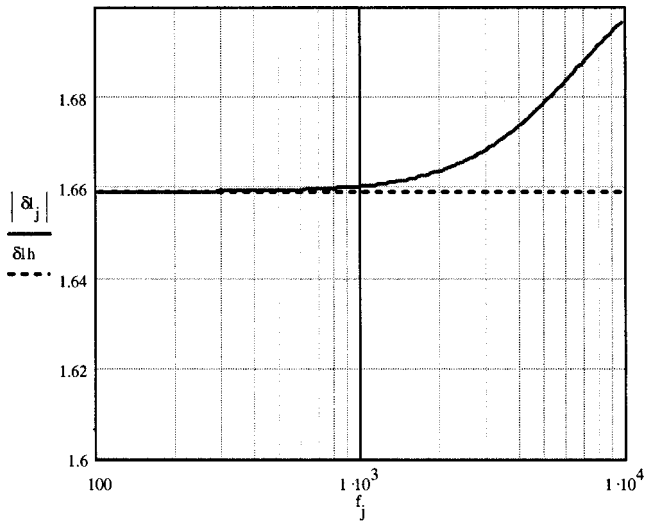


FIG. 1. A comparison of $|\delta| = |2s_{2D}/v_{2D}|$ (continuous line) and $\delta = 2s_2/v_2$ (broken line) for parameter values: flow resistivity 100 kPa s m^{-2}, porosity 0.4, tortuosity 2.5, and semicylindrical roughness radius 0.0075 m, mean spacing 0.03 m.

For $\alpha = \pi/2$, $\varphi \neq 0$ (grazing incidence but not normal to scatterer axes),

$$\beta^* = \frac{k_0^3 b V^2}{2}(1 - W^2) - ikV[2(\delta - 1)\cos^2\varphi - 1 + \gamma\Omega] + \beta. \tag{7}$$

If the roughness is periodic, then $W=1$ and only the imaginary part of admittance is changed. If the packing is random then $W=0$ and the real part of the admittance changes by a factor that increases with f^3. For nongrazing incidence, Eq. (6), with $W=0$, predicts the maximum roughness effect for a given value of Ω, V, and β. Roughness is predicted to be particularly significant for high-flow resistivity or hard ground surfaces near grazing incidence. For ''hard'' rough surfaces, i.e., where $\beta = 0$,

$$s_{2D} = s_2, \quad v_{2D} = v_2 = 1 + \frac{2\pi V}{3Nb^2}s_2 \quad \text{and} \quad \delta = 2s_2/v_2. \tag{8}$$

The model for the effective admittance of the rough interface between two fluids was developed for semi-infinite media.[1] However, for a high-flow resistivity medium $|\rho_2| \gg \rho_1$, and the difference between the values of δ calculated from $2s_{2D}/v_{2D}$ and $2s_2/v_2$ is small. This is illustrated in Fig. 1, in which calculated values of $|\delta| = 2|s_{2D}/v_{2D}|$ and $\delta = 2s_2/v_2$ are compared for a flow resistivity of 1000 kPa s m^{-2}. Consequently, (4), (6), and (7) may be evaluated by adding the normalized admittance β, calculated for any of the existing models for surface impedance (including near-surface layering, for example see Refs. 7–9, to the effective admittance predicted for acoustically hard rough ground.

Figure 2 shows predictions of the impedance of various rough finite impedance surfaces for a grazing angle of 5°. At all frequencies, and for the listed parameters, it is predicted that there is a reduction in the real part of impedance compared with the imaginary part. Such behavior at low frequencies is predicted also in the impedance of smooth ground surfaces if there is layering.[6] For larger roughness and lay-

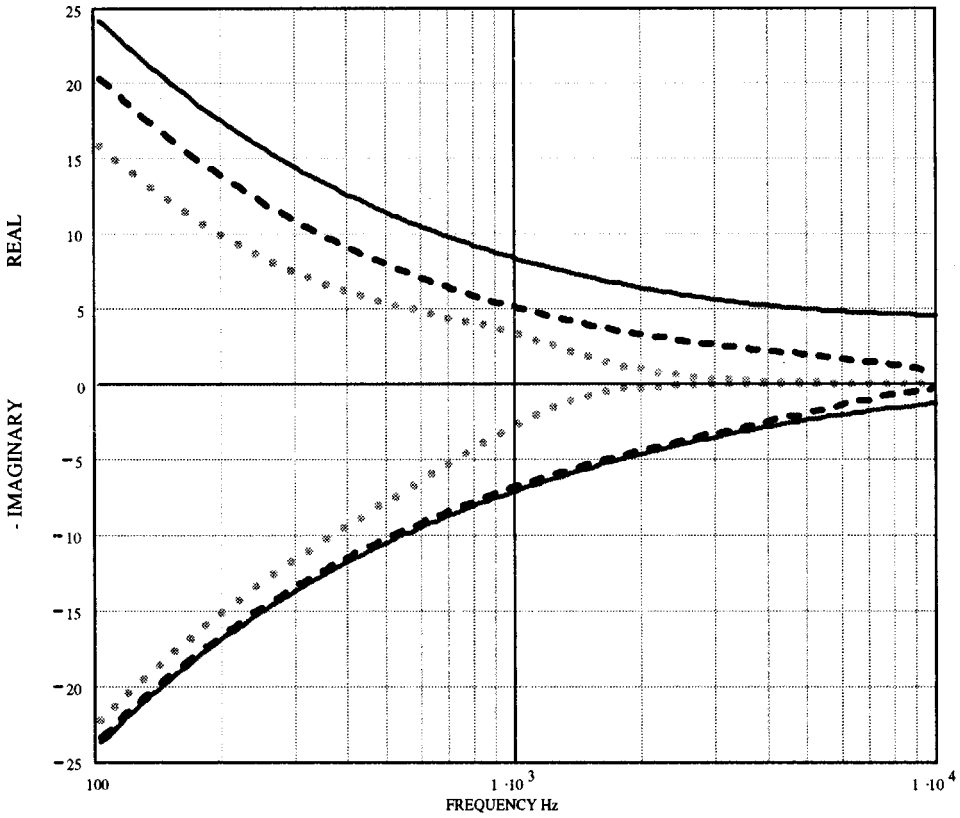


FIG. 2. Predicted importance of smooth and rough finite impedance surfaces. The porous medium is assumed semi-infinite with uniform triangular pores, flow resistivity 500 kPa s m^{-2}, porosity 0.4, and tortuosity =2.5. The solid lines represent the predicted relative impedance of the smooth porous surface. The broken lines correspond to random semicylindrical roughness 0.0075-m high with mean spacing 0.03 m. The dotted lines correspond to random semielliptical roughness, height 0.025 m, eccentricity 1.5, and mean spacing 0.06 m. The rough surface impedance is predicted for a grazing angle of 5°.

ered ground, more complicated relationships between real and imaginary parts of impedance may result (see Fig. 9 for example). Nevertheless, a common feature of the predictions is that, at high frequencies, the real part of impedance of rough finite impedance surfaces is less than would be expected for a smooth porous surface with the same pore structure. The high-frequency limit of the surface impedance of a smooth porous medium is determined by the ratio of tortuosity to porosity and, therefore, is always greater than 1. The incoherent component of scatter [represented by the first term in Eq. (6) or (7)] ensures that the effective impedance of a randomly rough finite impedance surface tends to zero at high frequencies. It should be mentioned also that roughness is predicted to cause the imaginary part of the impedance to be greater than the real part at low frequencies and this is a condition for the creation of surface waves above a finite impedance boundary.[12]

Note that the model assumptions, $kh<kb\leqslant 1$, are violated above 1000 Hz for the largest roughness case considered in Fig. 2.

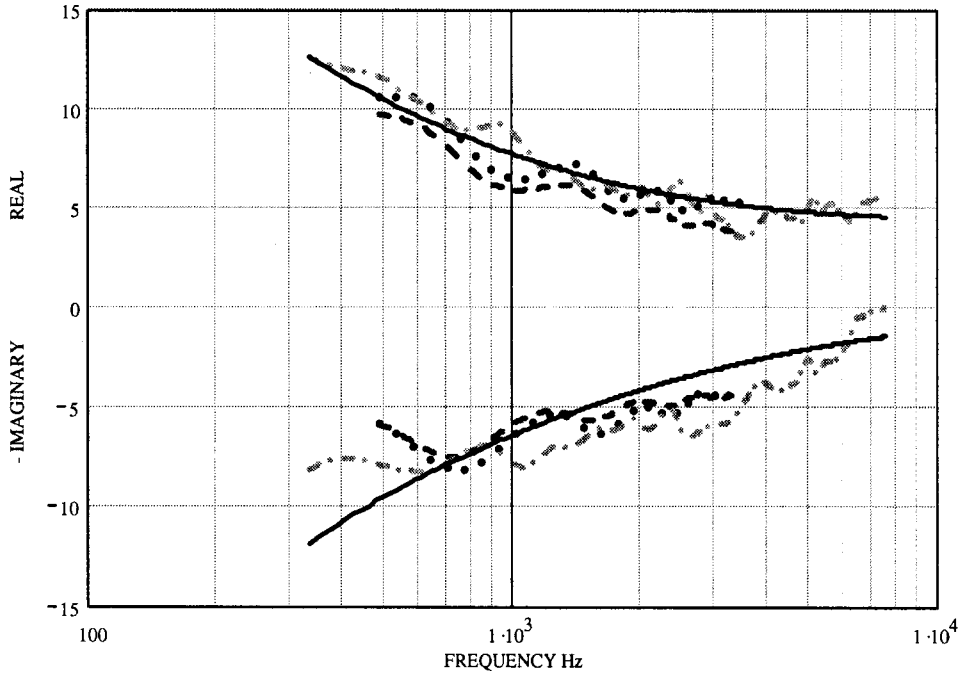


FIG. 3. Impedance deduced from three complex excess attenuation measurements with a point source over smoothed sand [source height =receiver height=0.1 m (dotted and dot-dash lines), source height =receiver height=0.05 m (dashed lines), range=1 m]. The predictions (solid lines) use a semi-infinite slit-pore model (Ref. 7), measured flow resistivity 420 kPa s m^{-2}, assumed porosity=0.4, and tortuosity=1/porosity.

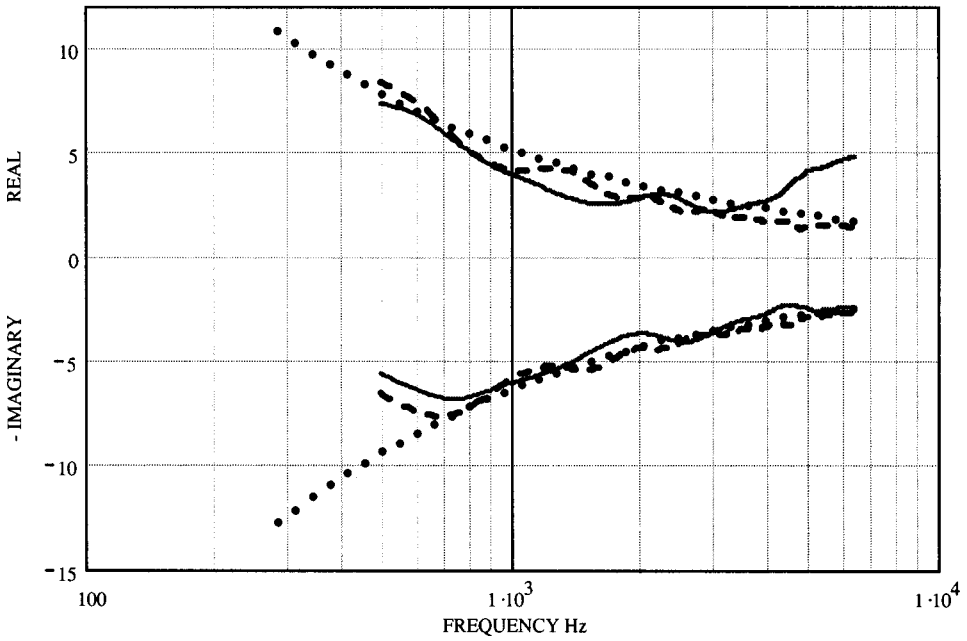


FIG. 4. Comparison of measured (solid lines) and predicted (dotted and broken lines) impedance of a sand surface containing (approximately) semicylindrical roughness, height 0.008 m, and center-to-center spacing 0.04 m. Predictions assume semicylindrical roughness with given dimension and spacing and either predicted (dotted line) or measured (broken line) smooth surface impedance (see Fig. 3).

III. COMPARISONS WITH LABORATORY DATA

A series of measurements has been made over a 0.3-m-deep tray of sand in an anechoic chamber. A point source consisting of a Tannoy driver fitted with a 1-m-long tube of 3-cm internal diameter has been used in the experiments. It was supported at the chosen height above the surface by a stand. The receiver was a Bruel & Kjaer type 4311 1.3-cm (0.5-in.) diameter condenser microphone fitted with a preamplifier. Usually, the microphone was supported by a retort stand at the same height as the source. Signal processing and signal generating were carried out using MLSSA, a maximum length sequence system analyzer. MLSSA enables background noise effects to be reduced by analyzing the microphone signal with respect to the known output sequence. A reference measurement was made with source and receiver at the maximum heights allowed by the supports and separated by 1 m. The source and receiver were then lowered to the heights of interest (for example 0.1 m) and the resulting received (complex) signal divided by the reference to obtain the complex excess attenuation spectrum. It should be noted that the point source used in the measurements reported here gave relatively little output below 400 Hz and that the impedance values deduced for frequencies lower than 400 Hz are not reliable. Initial measurements of complex excess attenuation were made over the sand surface which had been smoothed and slightly wetted. Subsequently, a comblike shaper was used to create parallel semicylinders in the surface, and further measurements of complex excess attenuation were made with the same source-receiver geometry. The sides of the sand semicylinders tended to crumble so that the resulting roughness shapes were only approximately semicylindrical.

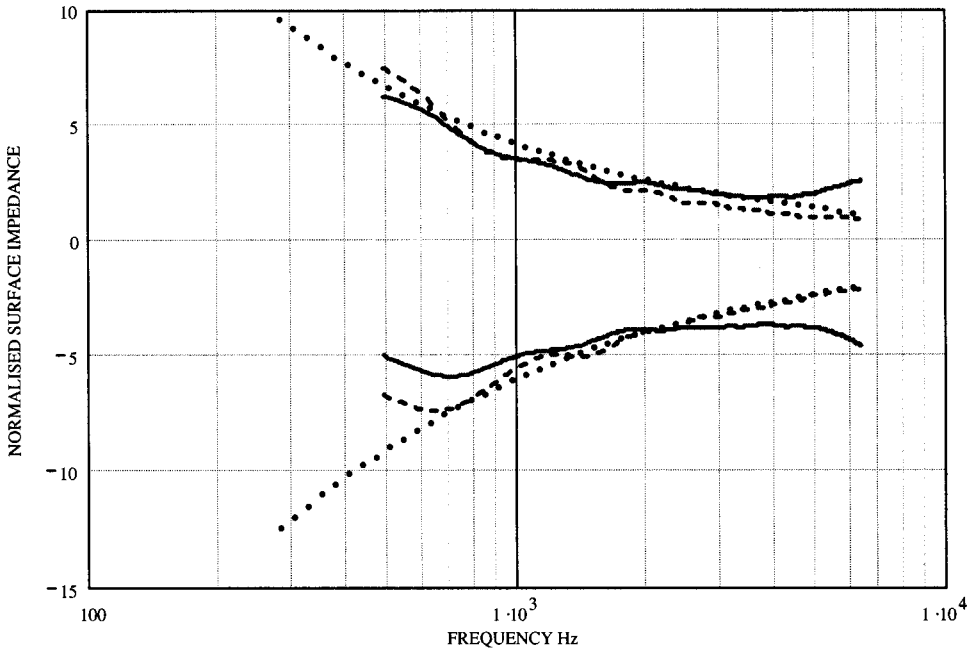


FIG. 5. Comparison of measured (solid lines) and predicted (dotted and broken lines) impedance of a sand surface containing (approximately) semicylindrical roughness, height 0.004 m and center-to-center spacing 0.02 m. Predictions assume semielliptical roughness (eccentricity 2) with given dimensions and either predicted or measured smooth surface impedance (see Fig. 2).

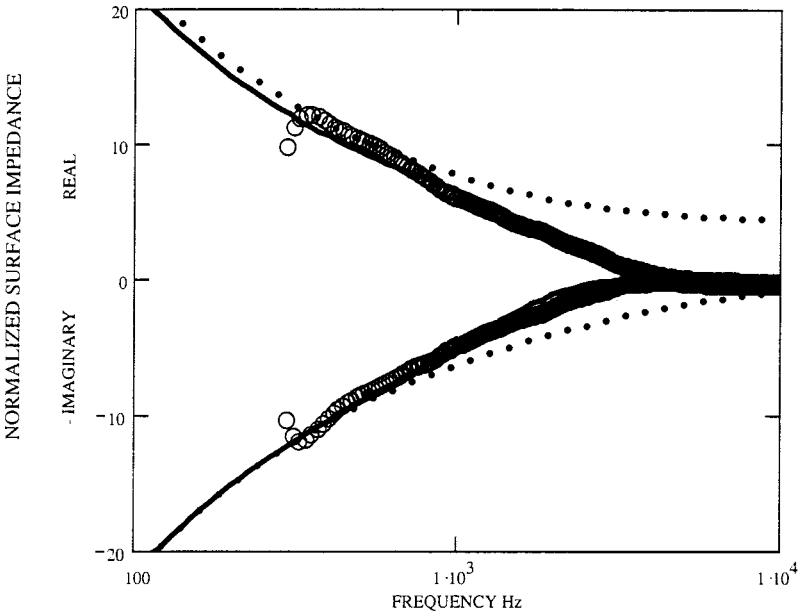


FIG. 6. Normalized impedance data (open circles) obtained from complex excess attenuation measurements over established grassland at Silsoe, Beds, UK. The theoretical predictions (dotted lines) are for the impedance of a semi-infinite slit-pore medium, with flow resistivity 400 kPa, s m^{-2}, porosity 0.4, tortuosity=1/porosity. Predictions including a partially random (randomness parameter $W=0.5$) close-packed semicylindrical roughness with 0.02-m radius are represented by the solid lines.

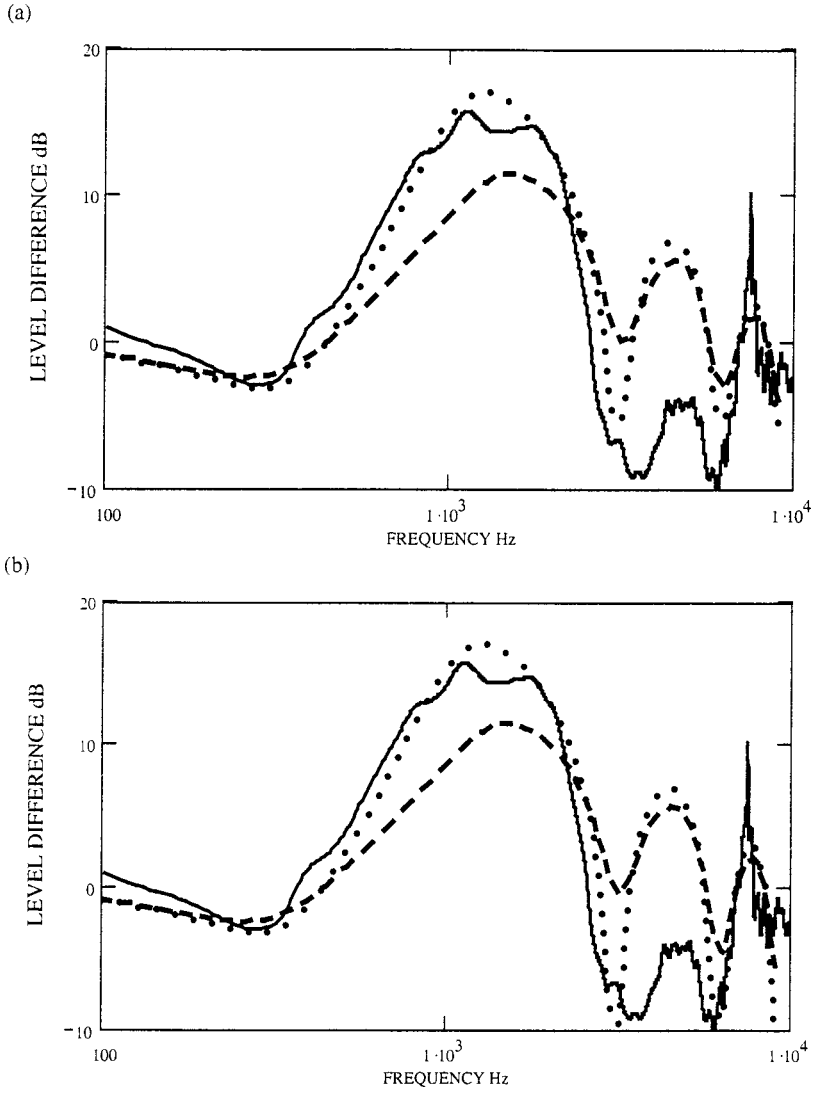


FIG. 7. (a) Average level difference spectrum obtained with a loudspeaker source at 1.65 m height and vertically separated microphones at 1- and 0.1-m height at a range of 30 m over sub-soiled ground (solid line). Predictions assume either the smooth surface of a hard-backed slit pore medium with flow resistivity 80 kPa s m^{-2}, porosity 0.4, tortuosity 2.5, and thickness 0.08 m (broken line) or a surface with these parameters plus 0.05-m-high close-packed but partially random (randomness factor 0.75) semicylindrical roughness (dotted line). Turbulence is characterized by $\langle \mu^2 \rangle = 10^{-5.75}$, $L_0 = 1$ m (see Ref. 12 for example). (b) Predictions excluding turbulence.

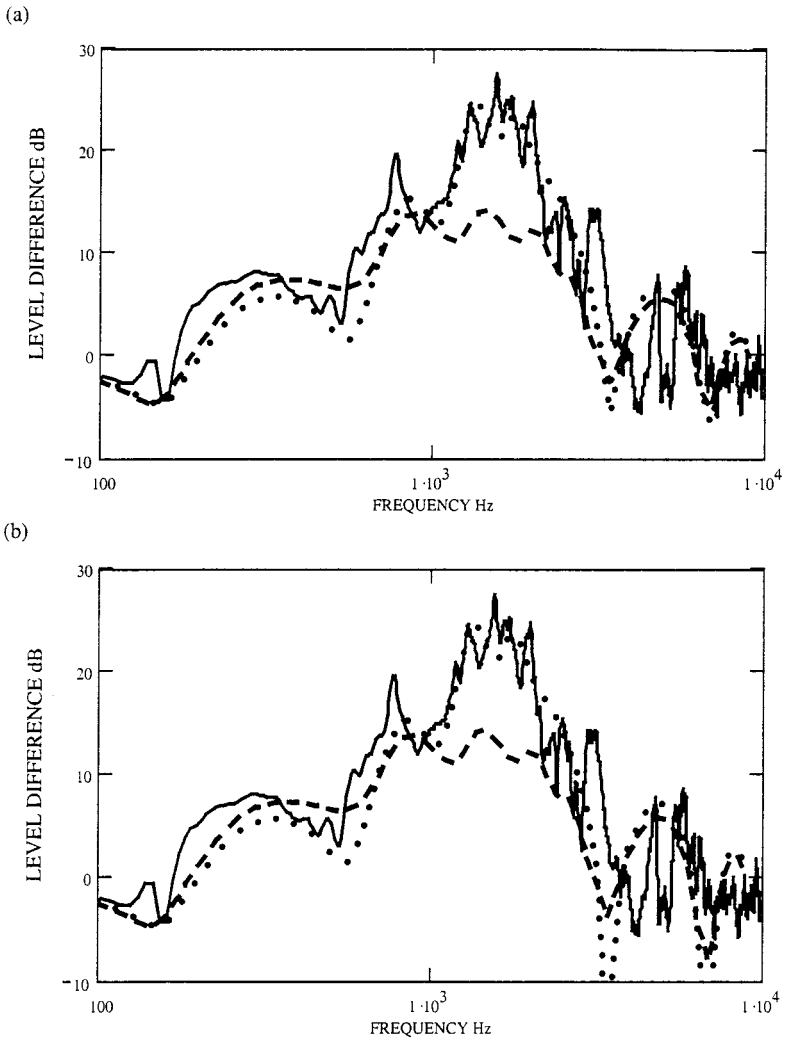


FIG. 8. (a) Average level difference spectrum obtained at a range of 30 m over ploughed ground (solid line). Predictions assume either the smooth surface of a hard-backed slit pore medium with flow resistivity 10 kPa s m^{-2}, porosity 0.4, tortuosity 2.5, and thickness 0.175 m (broken line) or a rough surface with these parameters plus 0.08-m-high, close-packed, semielliptical cylinder (eccentricity 1.6) roughness, (dotted line). The turbulence parameters assumed are the same as for Fig. 7. (b) Predictions excluding turbulence.

A method of deducing surface impedance, from measurements of complex excess attenuation,[10] has been used to analyze the resulting data. Figure 3 shows results of measurements over the smooth sand surface. Also shown in this figure are predictions based on the measured flow resistivity of the sand (420 kPa s m^2) assumed porosity of 0.4, assumed tortuosity of 1/porosity, and a semi-infinite slit-pore impedance model. The relatively high flow resistivity of the sand layer enables the finite depth of the sand (0.3 m) to be ignored.

Figures 4 and 5 show the impedance deduced from measurements of complex excess attenuation over two different rough sand surfaces. These figures also show predictions based either on Eq. (5) and the predictions for the smooth surface impedance shown in Fig. 3 (dotted lines) or on the measured smooth surface impedance plus a ''hard'' rough surface correction [obtained from Eqs. (4) and (5)—broken lines]. It should be noted that $kh>0.5$ at 3.4 kHz for the larger roughness (0.008-m high) in the experiments and that the rough surface model assumption, that $kh<kb\leqslant 1$, is violated at the higher frequencies of the measurements. The smaller (0.004-m high) semicylinders were assumed to be semielliptical with eccentricity (K) set equal to 2 to improve agreement with data.

IV. THE EFFECTIVE IMPEDANCE OF SOME OUTDOOR SURFACES

Figure 6 shows an example of a measured grassland impedance which tends to zero above 3 kHz. The data were obtained over an established area of grass at Silsoe, UK. The boss model, including randomness and incoherent scatter, may be used to give a close fit with these particular measurements of grassland impedance. The assumed flow resistivity and porosity are within the expected range. However, small-scale roughness must be included to provide a reasonable fit to these data (solid lines). The effect of not including roughness is shown by the dotted lines.

Potentially, cultivation practices have important influences on ground effect since they change both the roughness and surface properties (including the air permeability) of the ground. Aylor[11] noted a significant change in the excess attenuation at a range of 50 m over a soil after disking without any noticeable change in the meteorological conditions. Another cultivation practice is sub-soiling. It is intended to break up soil compaction 300 mm or more beneath the ground surface, caused, for example, by repeated passage of heavy vehicles. It is achieved by creating cracks in the compacted layer by means of a single- or double-bladed tine with sharpened leading edges. Sub-soiling has only a small effect

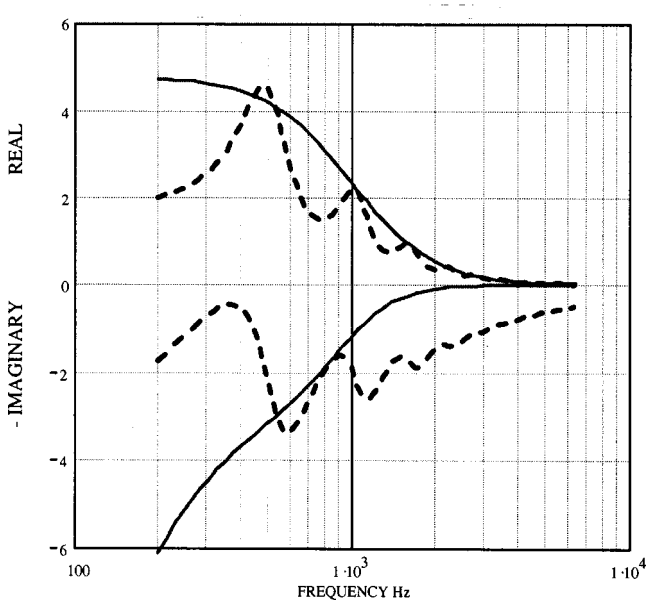


FIG. 9. Surface impedance of sub-soiled (continuous lines) and ploughed clay soil (broken lines) predicted using the parameters obtained by fitting the data shown in Figs. 7 and 8.

on the surface profile of the ground. This is in contrast to ploughing, which turns the soil surface over to a depth of about 0.15 m. Figures 7 and 8 show spectra of the difference in sound pressure levels received by vertically separated microphones over two conditions of heavy (boulder) clay. The data in Fig. 7 were obtained after sub-soiling and those in Fig. 8 were obtained after ploughing.

The microphones were nominally at 1 and 0.1-m heights and located 30 m from the Electro-Voice loudspeaker source (center at 1.65-m height) on a sunny day with light wind. The nominal values were difficult to maintain over the ploughes ground since the tripods' bearing source and microphones tended to sink. Vertical level differences are surrogates for excess attenuation since they are independent of the source spectrum. There are clear differences in the ground effects. Figures 7 and 8 also show theoretical predictions for the level difference spectra at 30 m. Since both scattering due to roughness and turbulence may influence the total field by reducing the coherence between direct and ground-reflected contributions, an attempt has been made to include turbulence effects. The method of allowing for turbulence introduces two parameters, the mean-square refractive index caused by local fluctuations in sound speed and a length representing the outer scale of turbulence, and assumes a Gaussian spectrum.[13] Predictions are shown both for rough and smooth surfaces both with and without turbulence. At the short range of measurements, the influence of turbulence is confined, essentially, to the higher-order dips associated with destructive interference between direct and ground-reflected contributions to the total field and dictated mainly by the source-receiver geometry. The fitted turbulence values are consistent with values expected for low turbulence conditions. Some of the changes in ground effect between the two soil states are predicted to result from differences in flow resistivity and surface layer depth. The main influence of these parameters is on the first level difference dip. However, improved agreement with data, particularly between 500 Hz and 2 kHz, is obtained in both cases by including roughness effects and by attributing different surface roughness to each ground condition. The roughness size assumed to fit the data over ploughed ground is greater than that assumed for the sub-soiled ground. This is consistent with the fact that sub-soiling produces less surface roughness than ploughing. the fitted thickness (0.175 m) for the surface layer of the ploughed ground is consistent with the expected depth of the plough pan and the fitted spacing (0.16 m) of the roughness is consistent with the size of furrows expected from ploughing. The fitted parameters for the ploughed ground include regularly spaced roughness and slightly lower source and upper receiver heights than the nominal values. Figure 9 shows the impedance predictions, corresponding to the parameters used to fit the level difference data in Figs. 7 and 8. The impedance predicted for the ploughed field is affected by layer resonance, as a consequence of the apparent low flow resistivity. That predicted for the sub-soiled ground has the ''rough finite impedance'' form, in which the impedance tends to zero and the imaginary part exceeds the real part at low frequencies (see also Fig. 6).

V. CONCLUDING REMARKS

At audio frequencies, small-scale roughness on laboratory and outdoor surfaces is predicted and measured to result in a change in the effective impedance. Tolerable predictions of the change may be obtained from heuristic extensions of the Biot/Tolstoy and Lucas and Twersky boss models. Small-scale roughness may influence the impedance of grassland surfaces. The modified boss model also predicts some of the features of measurements up to 30 m over sub-soiled and ploughed soil surfaces.

ACKNOWLEDGMENT

The work was supported in part by BBSRC (UK) Grant No. CTE02469.

[1] K. Attenborough and S. Taherzadeh, ''Propagation from a point source over a rough finite impedance boundary,'' J. Acoust. Soc. Am. **98**, 1717–1722 (1995).
[2] R. J. Lucas and V. F. W. Twersky, ''Coherent response to a point source irradiating a rough plane,'' J. Acoust. Soc. Am. **76**, 1847–1863 (1984).
[3] P. Boulanger, K. Attenborough, S. Taherzadeh, T. F. Waters-Fuller, and K. M. Li, ''Ground effect over hard rough surfaces,'' J. Acoust. Soc. Am. **104**, 1474–1482 (1998).
[4] H. Medwin and G. L. D'Spain, ''Near-grazing, low-frequency propagation over randomly rough, rigid surfaces,'' J. Acoust. Soc. Am. **79**, 657–665 (1986).
[5] J. F. Allard, L. Kelders, and W. Lauriks, ''Ultrasonic surface waves above a doubly-periodic grating,'' J. Acoust. Soc. Am. **105** (4), 2528–2531 (1999).
[6] J. P. Chambers, J. M. Sabatier, and R. Raspet, ''Grazing incidence propagation over a soft rough surface,'' J. Acoust. Soc. Am. **102** (1), 55–59 (1997).
[7] K. Attenborough, ''Acoustical impedance models for outdoor ground surfaces,'' J. Sound Vib. **99**, 521–544 (1985).
[8] K. Attenborough, ''Acoustical characteristics of fibrous and granular materials,'' J. Acoust. Soc. Am. **73**, 785–799 (1983).
[9] K. Attenborough, ''Models for the acoustical properties of air-saturated granular materials,'' Acta Acust. **1**, 213–226 (1993).

[10] S. Taherzadeh and K. Attenborough, ''Deduction of ground impedance from measurements of excess attenuation spectra,'' J. Acoust. Soc. Am. **105** (3), 2039–2042 (1999).
[11] D. Aylor, ''Noise reduction by vegetation and ground,'' J. Acoust. Soc. Am. **51** (1), 197–205 (1972).
[12] K. Attenborough, S. I. Hayek, and J. M. Lawther, ''Propagation from a point source over a porous half-space,'' J. Acoust. Soc. Am. **68**, 1493–1501 (1980).
[13] K. Attenborough, ''A note on short-range ground characterization,'' J. Acoust. Soc. Am. **95**, 3103–3108 (1994).

Acoustic field variability induced by time evolving internal wave fields

Steven Finette, Marshall H. Orr, and Altan Turgut
Acoustics Division, Naval Research Laboratory, Washington, DC 20375-5320

John R. Apel
Global Ocean Associates, Silver Spring, Maryland 20908

Mohsen Badiey
Graduate College of Marine Studies, University of Delaware, Newark, Delaware 19716

Ching-sang Chiu
Naval Post Graduate School, Monterey, California 93943

Robert H. Headrick, John N. Kemp, James F. Lynch, Arthur E. Newhall, and Keith von der Heydt
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Bruce Pasewark and Stephen N. Wolf
Acoustics Division, Naval Research Laboratory, Washington, DC 20375-5320

Dirk Tielbuerger
Forschnungsanstalt der Bundeswehr fur Wasserschall und Geophysik, Klausdorfer Weg 2-24, 24148 Kiel, Germany

(Received 19 March 1999; revised 14 April 2000; accepted 7 June 2000)

A space- and time-dependent internal wave model was developed for a shallow water area on the New Jersey continental shelf and combined with a propagation algorithm to perform numerical simulations of acoustic field variability. This data-constrained environmental model links the oceanographic field, dominated by internal waves, to the random sound speed distribution that drives acoustic field fluctuations in this region. Working with a suite of environmental measurements along a 42-km track, a parameter set was developed that characterized the influence of the internal wave field on sound speed perturbations in the water column. The acoustic propagation environment was reconstructed from this set in conjunction with bottom parameters extracted by use of acoustic inversion techniques. The resulting space- and time-varying sound speed field was synthesized from an internal wave field composed of both a spatially diffuse (linear) contribution and a spatially localized (nonlinear) component, the latter consisting of solitary waves propagating with the internal tide. Acoustic simulation results at 224 and 400 Hz were obtained from a solution to an elastic parabolic equation and are presented as examples of propagation through this evolving environment. Modal decomposition of the acoustic field received at a vertical line array was used to clarify the effects of both internal wave contributions to the complex structure of the received signals. [S0001-4966(00)03109-X]

PACS numbers: 43.30.Bp, 43.30.Pc, 43.30.Re [SAC-B]

I. INTRODUCTION

Oceanographic processes such as fronts, eddies, and internal waves, can mediate random variations in temperature and salinity fields. These variations induce space–time changes in the water column sound speed distribution that can be treated as a random field. The variability may span several space–time scales[1] and is a major factor in determining the statistical properties of an acoustic field propagating through an oceanic waveguide.

This article constructs a model for a randomly varying ocean environment and is derived from some basic oceanographic measurements in shallow water. Numerical simulations of acoustic field propagation through this shallow water waveguide are then presented to illustrate the influence of the oceanographic features causing variability in the acoustic field. While the relative variations in sound speed, δc, attributed to fluctuations in temperature and salinity are typically small perturbations of the mean sound speed $\overline{c}$ (for the environment discussed here, $[\delta c/\overline{c}]_{\max} \approx 10^{-2}$), these range- and time-dependent fluctuations significantly affect acoustic field properties. For example, acoustic mode coupling due to range-dependent spatial gradients in sound speed will redistribute modal energy within the water column and, in conjunction with mode stripping, alter the range dependence of the transmission loss, field coherence, and the normalized intensity fluctuations(scintillation index) for acoustic waves propagating in a random waveguide.[2,3]

The environmental variability discussed here is caused by internal wave fields and in shallow water they are distinguished from their deep water counterpart in several ways. Internal waves in shallow water are often nonstationary in

time and azimuthally anisotropic due to the presence of a contribution from spatially coherent, nonlinear wave packets. These packets propagate predominantly in the first internal wave mode.[4,5] For deep ocean environments, internal waves are reasonably well described statistically by the Garrett–Munk spectrum; this representation is stationary and azimuthally isotropic, reasonable assumptions when measurements are made in areas exterior to source-generating regions. The depth dependence for deep water internal waves is not limited to the first mode. While the above comments highlight differences between shallow and deep water internal wave fluctuations, they share the important property that the horizontal spatial scale of ocean fluctuations is much larger than the vertical scale; this implies a strong anisotropy between vertical and horizontal motions.[1]

Our approach includes several important features of shallow water internal wave fields: statistical variability, nonstationary field structure, nonlinear wave packets, and mode 1 dominance. The nonlinear component is treated deterministically and is combined with a random term representing a shallow water analog of the (linear) deep water internal wave field. These two contributions are mapped into the water column sound speed distribution. Anisotropic internal wave structure due to azimuthal variation of the wave packets is neglected; an acoustic propagation study including this structure requires a full three-dimensional acoustic treatment and is currently under investigation. The complete environmental model then consists of the space–time distribution of sound speed in the water column and the spatial distribution of sediment/bottom sound speed, attenuation, and density.

All parameters are constrained by data acquired from the SWARM (Shallow Water Acoustics in Random Media) experiment, performed during July–August 1995 on the continental shelf off the coast of New Jersey. This experiment was designed to understand some of the relationships between oceanographic and acoustic field variation by performing simultaneous oceanographic and acoustic measurements in a region of ubiquitous internal wave activity.[4] We have used a subset of those oceanographic measurements to construct the environmental model based on data collected along the SWARM acoustic propagation track. This model is then combined with an acoustic propagation code in order to perform numerical simulations of the space–time evolution of acoustic fields arriving at a vertical array. The usefulness of the model is illustrated by presenting some numerical simulations of acoustic field variability arising from random sound speed variations induced by the internal waves. A single realization of the field is considered here, which does not allow for direct comparison with the experimental acoustic data; computational efficiency limitations precluded the ensemble averaging of single, time-evolving realizations needed to arrive at appropriate statistics for data comparison. At this stage of development the model is, however, useful for developing some insight on acoustic variability caused by time-evolving internal waves in shallow water under complex environmental conditions.

Theoretical background for identification of relevant parameters of the environmental model is discussed in Sec. II. Next, we give a brief description of the experimental assets and methods used to estimate these oceanographic parameters, followed by some results of this procedure in Sec. III. Acoustic simulations incorporating the environmental model are presented in Sec. IV for frequencies of 224 and 400 Hz. Full field and modal intensity fluctuations are considered. The final section gives a summary and conclusions.

II. THEORETICAL BACKGROUND

Two sets of parameters need to be determined for a full characterization of the acoustic environment, one set for the water column sound speed distribution and another for the sediment/bottom properties. The computations of internal wave field parameters responsible for the water column sound speed perturbations are based on an extension of the theoretical approach described in Ref. 3. A brief summary is given in this section. Both the linear isopycnal displacements that form a spatially diffuse component and the nonlinear wave packet contribution associated with a spatially localized component differ somewhat from previous work and additional discussion of these terms is warranted. Parameters specifying the sediment/bottom characteristics are briefly mentioned at the end of the section; the estimation of these latter quantities are described in more detail in Ref. 4.

Let $r=\sqrt{x^2+y^2}$ denote the range measured along a line connecting the source point $r=0$ and the receiving array $r=r_{\text{array}}$. It is assumed that internal wave dynamics at the SWARM site are weakly nonlinear and that the total isopycnal displacement $\eta(r,z,t)$ can be decomposed into two independent contributions, one from a homogeneous, linear, and spatially diffuse internal wave field $\eta_d(r,z,t)$ and a second contribution $\eta_p(r,z,t)$ associated with nonlinear, spatially localized wave packets: $\eta=\eta_d+\eta_p$. The presence of nonlinear wave packets on the New Jersey shelf is well known.[4,6] A note on terminology is appropriate here. While the term ''nonlinear wave packet'' describes a more general wave field than the term ''soliton,'' the latter is commonly used in the oceanographic literature to loosely describe spatially localized nonlinear internal waves, even though these packets often do not strictly conform to the rigorous definition of a soliton. We will accept this point of view here, and use the phrases ''solitary wave packet'' or ''nonlinear wave packet'' interchangeably, bearing in mind that in a spatially inhomogeneous medium these waves can, for example, undergo shape changes as they propagate. They are distinguished from the linear internal wave field by traveling with a well-defined group velocity and retention of spatial coherence over a time scale of many hours. The sound speed distribution in the water column, $c(r,z,t)$, is assumed to be separable into a deterministic, time-independent ''background'' field $\overline{c}(r,z)$ and a random contribution $\delta c(r,z,t)$, which is small relative to the background field ($\delta c/\overline{c} \leqslant 10^{-2}$).

The relationship between the sound speed fluctuations and the isopycnal displacements due to internal waves is given by

$$\delta c(r,z,t)=\int_{z}^{(z+\eta(r,z,t))} \overline{c}(r,z')G(r,z')N^2(r,z')\,dz', \qquad (1)$$

where $G(r,z)$ is a function of the ratio of the depth gradients of salinity to temperature, and $N(r,z)=\sqrt{-((g/\rho)\partial\rho/\partial z)}$ is the buoyancy frequency.[3] In the latter expression ρ and g are the density of seawater and acceleration due to gravity, respectively. The description of δc given by Eq. (1) is more general than the standard expression[1] relating the sound speed perturbation to a small isopycnal displacement; that expression, in which $\delta c \propto \eta$, is not valid for large displacements such as those associated with the nonlinear packets.

A. Spatially diffuse, linear internal waves

For linear internal waves in a source-free region, the displacement of the isopycnal, η_d, by this diffuse contribution can be written explicitly as a sum over plane waves.[1] Because temporal evolution of internal wave fields is simulated in this article and used in the computation of 2-D acoustic fields in an (r,z) plane, it is necessary to consider a generalization of η_d that accounts for internal waves propagating from arbitrary directions in the (x,y) plane. One then chooses a representative 2-D range-depth slice of the time-evolving isopycnal displacement for use in acoustic propagation studies. This approach properly accounts for the superposition of amplitude and phase of the plane wave components of η_d. The distribution of energy among the plane wave amplitudes is expressed in wave number space (k_x,k_y) by a power spectral density $\langle|F(k_x,k_y,j)|^2\rangle$, where the index j corresponds to the jth mode. The depth dependence is described by the internal wave eigenmodes $W(k_h,j,z)$ with horizontal wave numbers $k_h=\sqrt{k_x^2+k_y^2}$. Therefore, we can write

$$\eta_d(x,y,z,t)=\sum_j \int_{k_x}\int_{k_y} F(k_x,k_y,j)W(k_h,j,z) \times e^{i[k_x x+k_y y-\omega(k_h,j)t]}\,dk_x\,dk_y, \tag{2}$$

where the weighting factors F are assumed to be zero-mean Gaussian random variables whose variance is the power spectral density associated with the diffuse contribution to the isopycnal displacement, $\omega=\omega(k_h,j)$ is a dispersion relation relating temporal frequency and wave number. The space and time evolution of the diffuse internal wave field are described by Eq. (2) provided (a) we ignore nonlinear interactions associated with the convective derivative in Euler's equation (i.e., we consider small amplitude displacements) and (b) assume invicid, incompressible flow within the Boussinesq approximation. If shear currents are neglected and the buoyancy frequency is range independent, the eigenmodes and dispersion relation are computed by solving the following Sturm–Liouville equation,

$$\frac{d^2}{dz^2}W(k_h,j,z)+k_h^2\left[\frac{N^2(z)-\omega^2}{\omega^2-\omega_c^2}\right]W(k_h,j,z)=0, \tag{3}$$

subject to boundary conditions $W(z=0)=W(z=H)=0$. In this eigenvalue problem, H is the water depth and ω_c represents the coriolis (inertial) frequency. When the power spectral density is given in the temporal frequency domain, i.e., obtained from time series measured at a single spatial location, the dispersion relation computed from a solution to Eq. (3) can be used to map a frequency spectrum into the spatial wave number domain as required in Eq. (2). The mapping is specified by requiring that $|F(\omega,j)|^2\partial\omega/\partial k_h|_{\omega=\omega(k)}=|F(k_h,j)|^2$ and assumes that ω is a single-valued function of k_h (see Ref. 7 for more complex scenarios). In the absence of information concerning the diffuse component field directionality, we assume it is isotropic. With this assumption, the spatial wave number spectrum can be written as $|F(k_x,k_y,j)|=|F(k_h,j)|/\sqrt{2\pi k_h}$. This completes the description of the linear component.

B. Spatially localized, nonlinear internal waves

It is convenient to orient the acoustic propagation path in the $y=0$ plane so that 2-D acoustic propagation is described by coordinates $r=\{x,z\}$. The solitary wave packet contribution to the sound speed variability, discussed below, will be described as a plane wave whose wavefront is perpendicular to the x axis with a wave vector parallel to that axis. The wave packet displacement of the isopycnal, η_p, is generated by nonlinear interaction of the internal tide with bathymetric features such as the continental shelf edge. The space–time evolution of the wave packet envelope is often described by a member of a family of nonlinear partial differential equations derived by perturbation expansion of Euler's equation. The particular choice depends on a number of factors and constitutes an active area of research in the oceanographic community. For our purposes, we consider a solution based on the generic Korteweg–deVries (KdV) equation discussed later. This solution is valid for mildly nonlinear waves when $\eta_p/h_1 \ll 1$ and $[(h_1+h_2)/\lambda]^2 \ll 1$, where h_1 is the thermocline depth, h_1+h_2 specifies the water column depth, and λ is a representative wavelength.[8,9] A nonlinear wave packet can be written as a sum of products where one term of the product, W, describes the depth dependence via the eigenmode solutions in Eq. (3) and the other term, A, characterizes the evolution of the packet envelop with distance and time and η_{oj} gives the amplitude for the jth mode: $\eta_p(x,z,t)=\sum_{j=1}^{j\,\max}\eta_{oj}W(k_h,j,z)A_j(x,t)$. The first mode usually dominates in shallow water so the sum reduces to only one term; this will be assumed in what follows. Dropping the subscript j for notational simplicity, the evolution function $A(x,t)$ satisfies the KdV equation

$$\frac{\partial A}{\partial t}+c_o\left(\frac{\partial A}{\partial x}+\beta\frac{\partial^3 A}{\partial x^3}+\alpha A\frac{\partial A}{\partial x}\right)=0, \tag{4}$$

where parameters β and α describe the internal wave dispersion and nonlinearity, respectively. The limiting (long wavelength) phase speed, c_o, is obtained from the dispersion relation associated with the solution to Eq. (3). Neglecting shear currents, the environmental parameters are given by the following integrals:

$$\alpha=\frac{3}{2Q}\int_H^0\left(\frac{dW}{dz}\right)^3 dz, \tag{5}$$

$$\beta=\frac{1}{2Q}\int_H^0 W^2\,dz, \tag{6}$$

$$Q=\int_H^0\left(\frac{dW}{dz}\right)^2 dz, \tag{7}$$

where H is the water depth and the eigenmodes are defined by solution of Eq. (3). The KdV equation has the well-known wave solution $A(x,t)=\mathrm{sech}((x-Vt)/\Delta)^2$, and this functional form has been used in several recent numerical studies of acoustic propagation through an internal wave field.[3,10–12] Accumulating experimental evidence at the SWARM site indicates that this function is qualitatively deficient in two respects. It does not allow for the internal tide-related depression and eventual recovery of an isopycnal surface to its unperturbed state nor the addition of new pulses within the packet as it propagates along the leading edge of the internal tide. The resulting depression and recovery formed by the internal tide can exist on the order of half the semidiurnal tidal period and may alter the acoustic propagation path through the waveguide. An attempt is made below to include this behavior through a little-known solution to Eq. (4) termed the ''dnoidal'' solution.[9,13,14] This solution naturally includes packet dispersion, the depression of the isopycnal displacement by the internal tide, and the addition of spatial oscillations to the solitary wave packet as it evolves in time. It is multiplied by an *ad hoc* term that describes the observed isopycnal recovery or ''wake'' to its unperturbed value.

Based on the above discussion, we consider the internal tide as an undulatory bore with the solitary waves at its leading edge (termed a solibore in Ref. 15). The spatial interval between packets is given by the internal tide wavelength on the shelf. A brief description of the wave packet model used in this article is given below; more details can be found in Ref. 9. It should be noted that this function represents an asymptotic solution to the KdV equation and is not valid near or at the source-generating region. The evolution function is given by the expression in brackets:

$$\eta_p(x,z,t)=\eta_o W(k_h,j=1,z)\left\{2dn_s^2\left[\frac{k_o}{2}(x-Vt)\right]-(1-s^2)\right\}I(x,t), \tag{8}$$

where $dn_s(\,)$ denotes a Jacobian elliptic function parametrized by a nonlinear parameter $s=s(x)$ which varies with location within a wave packet ($0\leqslant s\leqslant 1$), and

$$V=c_o\left(1+\left[\frac{1+s^2}{3}\right]\alpha\eta_o\right) \tag{9}$$

is the nonlinear phase speed of the packet. Note that the amplitude η_o represents half the peak-to-trough displacement of the leading pulse in the packet. The short wavelength wave number k_o is related to the KdV parameters by $k_o/2=(\alpha\eta_o/6\beta)^{1/2}$. We use the following *ad hoc* form to describe the recovery of the isopycnal displacements to their equilibrium values: $I(x,t)=\frac{1}{2}\{1+\tanh((x+x_1-c_ot)/L)\}\times e^{-a(x-x_o)}$, where L is the characteristic scale of the recovery phase, x_1, a, x_o are adjustable parameters, and the exponential factor attenuates the waveform at a rate that can be adjusted from observation.

The parameter $s=s(x)$ is obtained by solving a transcendental equation, but can be approximated by a monotonic function varying between zero at the trailing edge of the packet and unity at the leading edge. In the limit $s\to 1$, achieved at the leading edge of the packet where the displacements are large, dn_s approaches the hyperbolic secant solution discussed above, while at the trailing edge, where the displacements are small and $s\to 0$, the solution is given by a sinusoid. The evolution function in Eq. (8) adds depressions and undergoes dispersion as it propagates. For reference, a picture of solitary wave isopycnal displacements is given later (Fig. 6) and discussed there in the context of numerical simulations.

The sum of the displacements in Eqs. (2) and (8) appears in the upper limit of the integral in Eq. (1) which computes the space–time evolution of the sound speed field induced by internal waves. Regardless of the detailed structure of the packet, its isopycnal displacement is added to the ever-present spatially diffuse internal wave contribution η_d and the relative strengths of these two displacements determine the acoustic significance of including such details. The relative importance of the contributions on acoustic field behavior can be evaluated by computer simulation; examples are given in Sec. IV B.

In summary, a number of oceanographic parameters and functions need to be estimated in order to describe the space–time evolution of the sound speed distribution associated with an internal wave field. These include the set $\{W(k_h,j,z),|F(k_h,j)|^2,\omega(k_h,j),N,G,\eta_o,s,\kappa_o,c_o,V,\alpha,\beta\}$. Additionally, bottom properties such as density, compressional and shear speeds, attenuation, and bathymetry estimates along the experimental track need to be included in the sound speed environmental model to fully characterize the waveguide for acoustic propagation studies. Internal wave parameters are estimated in the next section at several spatial locations along the SWARM track. The power spectral density associated with the spatially diffuse internal wave contribution, as well as the bathymetry and bottom parameters for this site, have been estimated elsewhere.[4]

III. ESTIMATION OF MODEL PARAMETERS

This section describes procedures used to compute the model parameters needed to specify the acoustic environmental model. A very brief description of the experimental assets[4] is given first, and then the analysis approach is described. Some representative parameters are discussed in detail, followed by the construction of the environmental model and an example of the perturbed sound speed distribution.

A. Experimental configuration

A subset of the SWARM experimental assets was used in this study and this subset is illustrated in Fig. 1. The R/V Endeavor traversed a 42-km track between two acoustic sources and the vertical receiving line array, performing tow-yo CTD casts over a depth window containing most of the internal wave activity. The acoustic sources emitted at 224 and 400 Hz; they were located at depths of 48 and 29 m, respectively. There were also times when tow-yo casts were not made and during these periods the ship was positioned at stations along the track where it performed stationary CTD casts over the entire water column. A hull-mounted 200-kHz

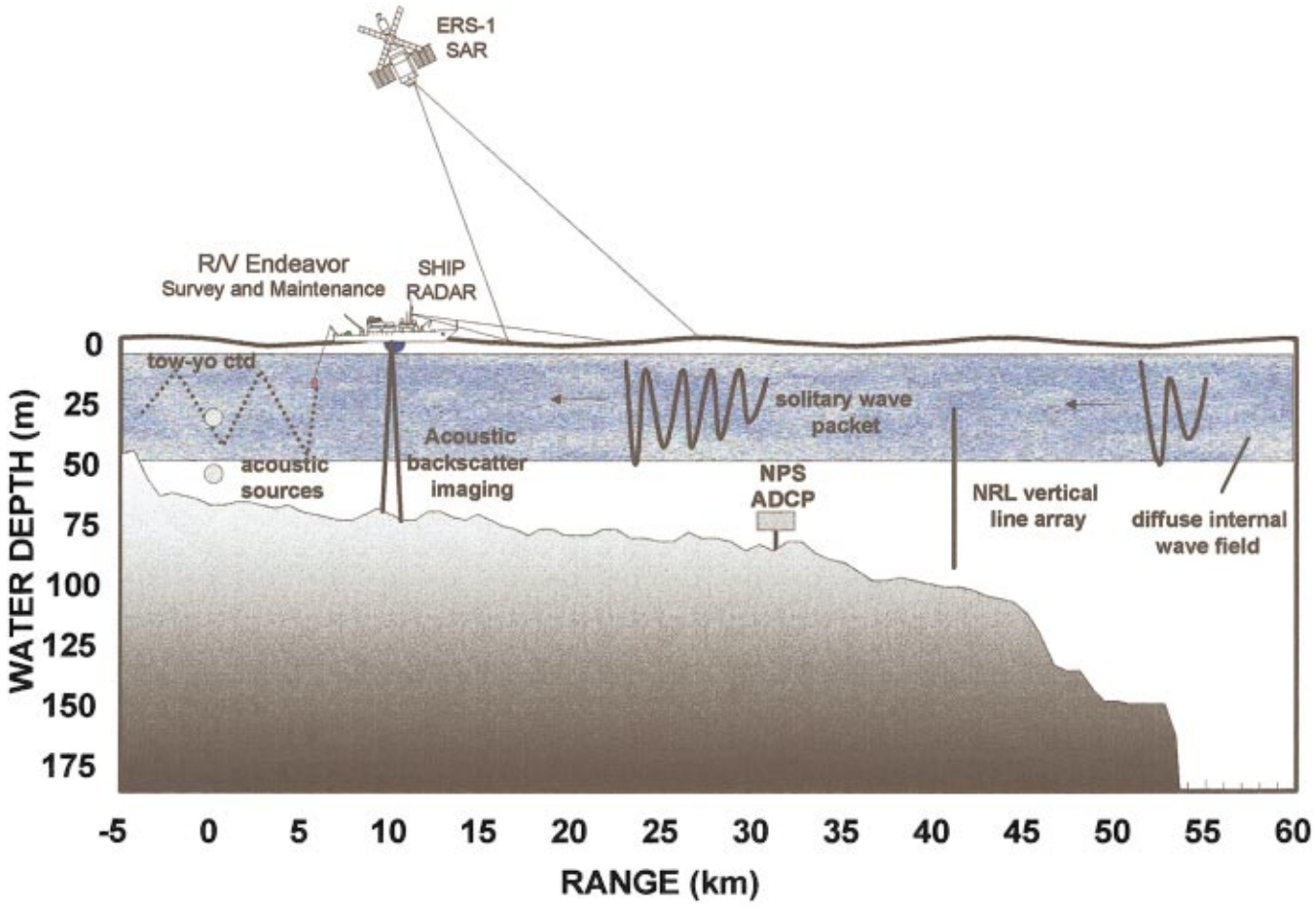


FIG. 1. Experimental assets used in the construction of the environmental model and acoustic propagation studies.

acoustic backscatter sonar was used to image internal wave displacement of scattering layers (correlated with isopycnal displacements) while the ship radar simultaneously acquired images of the surface expression of the solitary wave packets. A bottom-mounted acoustic Doppler current profiler (ADCP) computed water particle displacements which, after filtering out the nonlinear wave packet contribution, were used to estimate the displacement power spectral density for the linear, spatially diffuse internal wave field. A 3.5-kHz center frequency chirp sonar (not illustrated in Fig. 1)

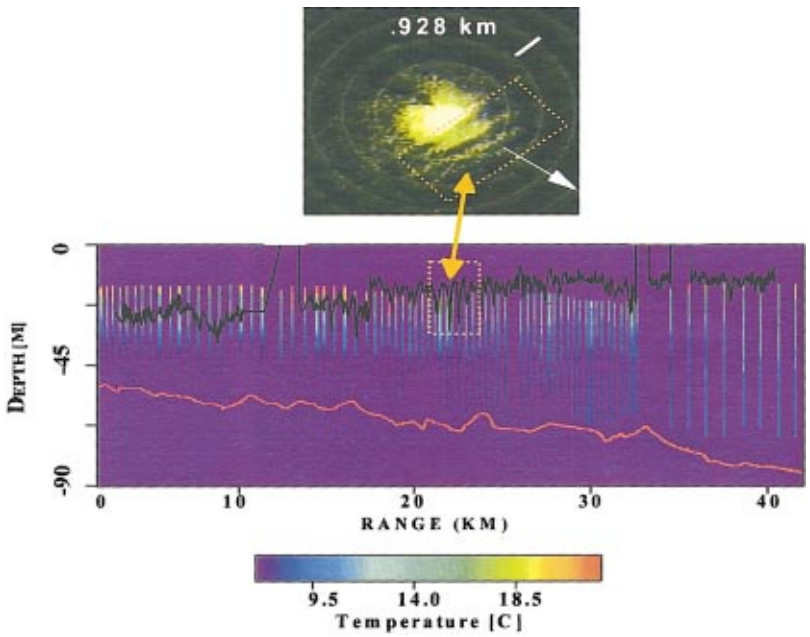


FIG. 2. Overlay of tow-yo temperature data along the SWARM track (color, vertical strips) with the isopycnal displacement (black trace) obtained from high-frequency acoustic backscatter measurements. The radar image shows the surface signature of a solitary wave packet with the simultaneously measured *in situ* image shown in the dotted box. Vertical jumps to zero depth in the backscatter curves bound regions associated with breaks in the data record.

mapped the bathymetry and subbottom structure. The backscattered chirp signal was processed to compute bottom parameters using a poro-elastic Biot model and a simulated annealing inversion technique.[4] Density, shear, and compressional sediment speeds and attenuation coefficients were estimated for unconsolidated sediments with a depth resolution of 1 m and a lateral resolution of about 25 m.

B. Internal wave parameters

The parameters and functions considered in Sec. II are estimated here primarily from a single day of SWARM measurements (3 August, 1827 h–4 August, 0157 h). These quantities are computed along the track at five selected locations, interpolated in range, and mapped into sound speed. We begin with the temperature and salinity measurements since these fields are the basis for the computation of several quantities of interest.

1. Buoyancy frequency

Tow-yo CTD data were mapped onto the SWARM track using global positioning system (GPS)-derived spatial coordinates. For each up and down cast, the temperature and salinity data, $T(x,z), S(x,z)$, were mapped into depth at a range midway between the beginning and the end of the up/down cast, i.e., $x^i = (x_{\text{start}}{}^i + x_{\text{end}}{}^i)/2$ for the ith cast. The data for all the CTD casts acquired during one transit of the 42-km track were then placed onto a common depth grid with $\Delta z = 0.2$ m step size. Data points in the same depth interval for a given range were averaged. For the acoustic propagation frequencies of interest here, 224 and 400 Hz, the effect of this averaging should be small. Variable ship speed and changing water depth caused the spacing of these vertical *T, S* strips to be nonuniform in range and depth coverage of the water column. The tow-yo coverage included depths dominated by internal wave activity, excluding both the upper 10–15 m of the water column and approximately the deepest 10 m relative to the bottom. To allow for full depth coverage of the water column, other information was used. Temperature data points for $z = 5$ m and $z = 1$ m were obtained by thermistors mounted on the ship's hull. Three days before the tow-yo data were acquired, stationary CTD casts were made over the entire water column at several ranges throughout the track when the ship was on station, and these casts were used to compute salinity for the upper 10–15 m and both temperature and salinity for the lower depths not covered by the tow-yo run. An increase in temperature was observed in the lower 10 m of the water column starting about 20 km from the source and extending to the acoustic receiving array location, causing a slight (but acoustically significant) upward refracting contribution to the sound speed profile over that part of the track. This combined *T, S* data set was linearly interpolated from the surface to the bottom for those gridded depths for which data were not available. The hydrostatic water pressure $P(z)$ was computed and used in conjunction with $T(x,z), S(x,z)$ and an equation of state for sea water[16] to compute the density for each vertical strip. Next, the buoyancy frequency $N(z)$ was obtained from the density gradient for each depth strip and smoothed using a running average over a 6-m depth interval. The temperature and salinity gradients were used to compute $G(x,z)$, mentioned earlier in connection with Eq. (1).

2. Internal wave eigenmode, KdV wave packet parameters

Before computing the internal wave eigenmodes using these buoyancy frequencies we note that, while knowledge of the eigenfunctions $W(k_h, j, z)$ is necessary to determine the depth dependence of both η_D and η_S, the modes are computed from Eq. (3), which was derived via linearization of the Euler equation of motion. The buoyancy frequencies along the track are potentially biased by depth variations induced by the nonlinear contribution of the solitary wave packets to the isopycnal displacement. Since only packet-free regions should be used to estimate the modes, the spatial location of nonlinear packets along the track were determined using the tow-yo CTD, surface ship radar, and the acoustic backscatter data.[17] Ship radar images located the surface expression of the coherent nonlinear packets by using electromagnetic backscatter from surface capillary waves perturbed by the internal wave packet currents. Videotape recordings of radar images provided information about the direction of packet propagation, location of the larger amplitude pulses in a packet, and pulse wavelengths within a circular region of radius 5 km centered on the ship position. High-frequency acoustic backscatter images measured scattering layer displacement within the water column and provided a more direct measure of the latter two wave packet properties. The acoustic backscatter data were acquired at a sampling rate of 4 Hz and converted from two-way travel time to depth relative to the surface assuming a constant sound speed of 1500 m/s. For a nominal ship speed of 4–5 knots the smallest horizontal spatial wavelength is depth dependent. At a depth of 20 m the 3-dB points on the beam were about 1 m apart, and this horizontal wavelength was well below that of solitary waves measured at the site. The acoustic backscatter measurements were mapped onto the same grid as the tow-yo data using GPS coordinate information. The combination of these space coincident measurements allowed us to estimate relatively ''quiescent'' periods, corresponding to positions at which the linear internal wave field dominated the sound speed variability. Several favorable locations were found at $x = 2.78$, 4.91, 14.39, 30.1, and 37.55 km. These positions were used to select the CTD data from which the buoyancy frequency and the internal wave modes were computed. An overlay of the isopycnal displacement on the tow-yo temperature strips distributed throughout the track is shown in Fig. 2, along with a radar image that directly corresponds to a segment of the acoustic backscatter image highlighted in the dotted box.

Empirical orthogonal function analysis of the moored acoustic Doppler current profiler (ADCP) data indicated that the diffuse internal wave field was dominated by the first mode,[4] and therefore only the first mode contribution was included here. For the computation of the diffuse internal wave component this simply eliminates the sum over modes in Eq. (2), though the sums over horizontal wave number are required. The η_p contribution also reduces to a single term $\eta_p(x,z,t) = \eta_o W(k_h = k_o, j = 1, z) A(x,t)$, where k_o was esti-

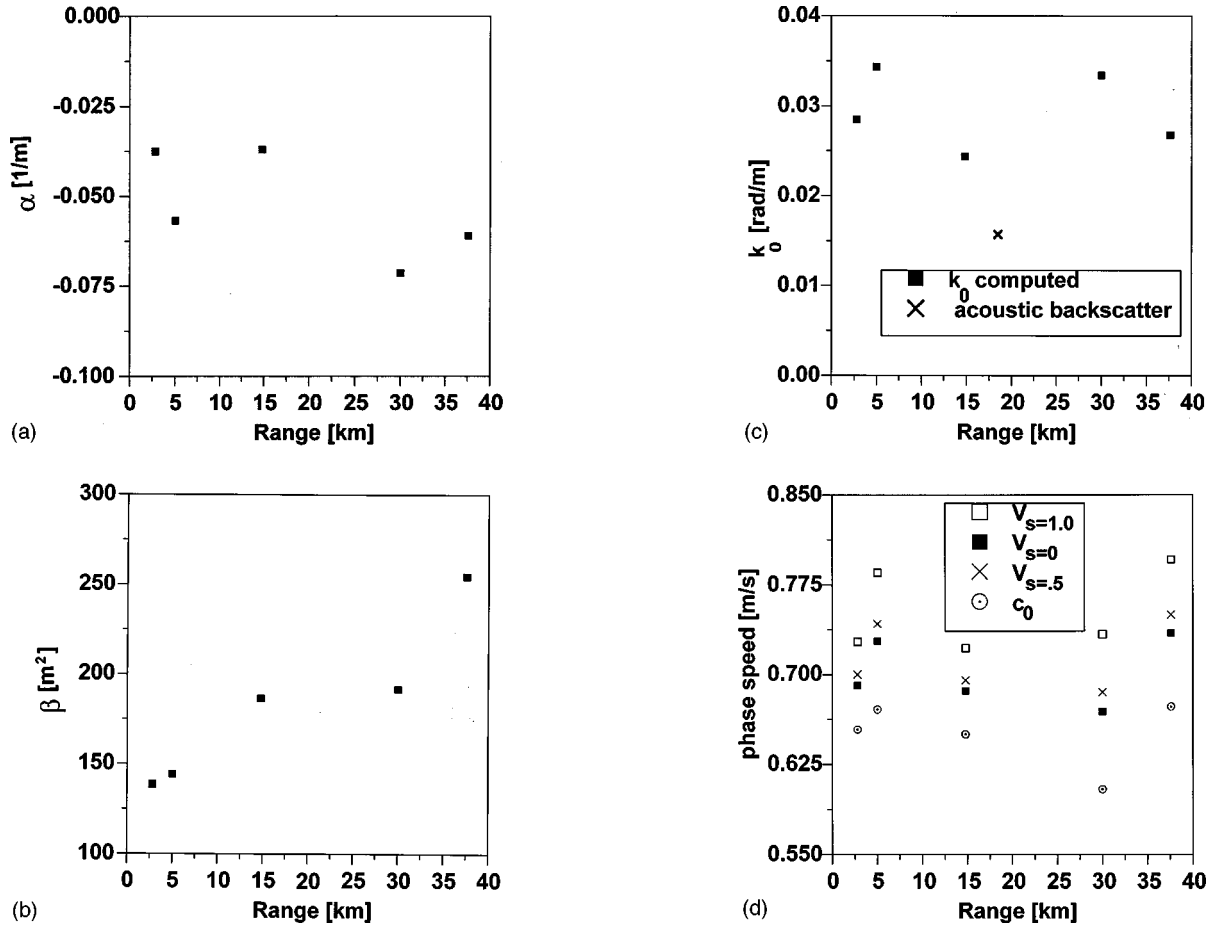


FIG. 3. Environment model parameters as a function of range. (a) Nonlinearity parameter, (b) dispersion parameter, (c) short wavelength wave number, and (d) solitary wave phase speed and diffuse component phase speed.

mated from the limiting distance between successive peaks[9] in the isopycnal displacement of the wave packet viewed in the acoustic backscatter image, in conjunction with use of the radar images as a guide to isolate an appropriate internal wave packet. The resulting internal wave mode determines the depth dependence of the isopycnal displacement associated with a wave packet. The value of k_o estimated in this manner for a packet located at range $x \approx 18.5$ km was approximately $k_o = (2\pi/400)\ \mathrm{m}^{-1} = 0.016\ \mathrm{m}^{-1}$.

Note that this approach yielded a k_o value at a specific point along the track for a given packet. An alternative method of computation for k_o involves the equation linking the short wavelength wave number with η_o, α, γ. The appropriateness of the latter approach may be in question for solitary waves at this site since one of the conditions for the validity of Eq. (4) is $\eta_o/h_1 \ll 1$; for SWARM data, the ratio is about $\frac{1}{2}$. To assess the usefulness of the alternative method, a consistency check was made in the following manner. Using the mode shape obtained for wave number k_o, the nonlinearity and dispersion coefficients were computed by trapezoidal integration of Eqs. (5)–(7). Given these estimates of the KdV parameters, a consistency check was make on k_o by computing this wave number from $\frac{1}{2}k_o = (\alpha\eta_o/6\gamma)^{1/2}$ and using a value for $\eta_o = 4.5$ m obtained from the high-frequency backscatter image. The result is $k_o = 0.02\ \mathrm{m}^{-1}$ at 14.77 km, quite close to the value computed earlier at 18.5 km from the acoustic backscatter data. Given the reasonable comparison between k_o values at similar ranges using the two approaches discussed earlier, we computed $k_o(x)$ using the equation approach at the five ranges selected previously. The nonlinear phase speed of the packet was estimated from Eq. (9) with c_o computed from the dispersion relation at each of the five range points. The resulting parameters $\{\alpha, \beta, k_o, c_o, V\}$ are shown as a function of range in Fig. 3. The trend in Figs. 3(a) and 3(b) is toward decreasing values of the dispersion and nonlinearity parameters with decreasing water depth, consistent with exact two-layer model results.[8] The two-layer model also predicts that the nonlinearity parameter is negative for mode one depressions of the isopycnal. The solitary wave phase speeds computed from Eq. (9) and shown in Fig. 3(d) are consistent with direct measurements of phase speed made at the SWARM site.[4] The highest speed is at the leading edge of the packet ($s = 1$) and the diffuse background phase speeds, c_o, are less than the packet speed, as expected.

3. Spatial frequency spectrum for linear internal waves

A temporal frequency spectrum for the diffuse internal waves was estimated from ADCP data for the first mode.[4] This spectrum was converted to a spatial wave number spectrum by using the group speed $\partial\omega/\partial k_h$ obtained from solving

the eigenvalue equation at each of the five ranges and applying the mapping from frequency to wave number space discussed in Sec. II. The transformation covered the band of frequencies {0.0001, 0.005} Hz measured by the ADCP and this band was mapped into a wave number range {0.0014, 0.122} m^{-1}, corresponding to wavelengths between 52 and 4486 m.

4. Deterministic sound speed field $\bar{c}$

The deterministic contribution to the sound speed distribution in range and depth, $\overline{c}(x,z)$, was obtained by linear interpolation utilizing the five (T,S) profiles found at the ''quiescent'' ranges noted earlier. Assuming quiescence of $\approx$2 km, three adjacent tow-yo depth profiles were averaged at each of the five ranges. A standard phenomenological expression[18] was applied to compute the sound speed profile as a function of T, S, P at those ranges and an estimate the entire sound speed field over the track was obtained by linear interpolation. The solitary wave packet parameters at the five ranges were also linearly interpolated across the track, with the resulting values used to model the evolution of the wave packet(s) in space and time.

5. Examples of synthesized sound speed distributions

Let c_p, c_d represent the sound speed perturbations due to the solitary wave packet(s) and diffuse (linear) component of the internal wave field, respectively. Examples of synthesized sound speed environments at a given instant of time t_o are presented in Fig. 4 which shows the deterministic background profile $\overline{c}(x,z)$ in Fig. 4(a), an environment for $\overline{c}(x,z)+\delta c_p(x,z,t_o)$ in Fig. 4(b), and another for $\overline{c}(x,z)+\delta c_p(x,z,t_o)+\delta c_d(x,z,t_o)$ in Fig. 4(c). The trailing thermocline depression shown in Fig. 4(b) and its relaxation to the unperturbed state at about 25 km are associated with the isopycnal recovery phase. In the next section, some acoustic simulation results are shown using time evolving environments for which Figs. 4(b) and 4(c) are examples.

As a consistency check on the scale and depth dependence of the sound speed fluctuations, the depth-dependent standard deviation of δc_d derived from processing 2 h of moored yo-yo CTD data is shown as a solid curve in Fig. 5. The data were obtained a few kilometers shoreward of the NRL array (see Fig. 1) and include contributions from all the internal wave modes present, though the standard deviation was computed without the solitary wave packet contribution.[19] For comparison, the dotted curve represents the simulated random sound speed perturbations for the linear, mode one contribution of the internal wave field computed by averaging five consecutive 2-h time segments taken at approximately the same range as the fixed yo-yo CTD data. Considering that the experimental data were obtained on a different day from data used to compute the environmental model and that mode one dominates the field in this region, the qualitative structure of the curves are quite similar. Reasonable quantitative agreement occurs though the data indicate somewhat elevated fluctuation regions between 25 and 40 m, with a slight shift in location (in depth) of the peak rms variations in sound speed.

IV. NUMERICAL EXPERIMENTS

A. Acoustic simulation approach

The synthesized sound speed distribution in the water column, along with the bathymetry and bottom parameters were used as input to the wide-angle, elastic parabolic equation code FEPES.[20] Details concerning the implementation follow the discussion in Ref. 3. Comparison of transmission loss results with those using the acoustic parabolic code RAM showed that the effect of shear was not significant for this environment. In this section, the emphasis is on simulations of acoustic propagation through a waveguide in which a single realization of the sound speed distribution evolves in time over an M2 tidal cycle. We interpret the term ''realization'' to imply that the linear component of the internal wave spectrum is associated with a random field (see Sec. II) and sampled using Monte Carlo techniques. The resulting zero mean, Gaussian random variables $F(k_x,k_y,j=1)$ were then substituted into Eq. (2) which describes the evolution of the linear isopycnal component. Other realizations of this component would correspond to different samples of the spectrum; due to computational limitations, no statistical averaging over multiple realizations was performed. In conjunction with the specification of the wave packet parameters, Eqs. (1), (2), and (8) then specify the sound speed perturbation as a function of space and time. This quantity is added to $\overline{c}(x,z)$ to fully characterize the sound speed field in the water column. For each time step $\varepsilon\Delta t(\Delta t=60\,\mathrm{s})$ of the evolving sound speed distribution, a cw acoustic field $P_\varepsilon(x,z,f)$ emitted by a point source at $x=0$ is propagated through the waveguide to 42 km. Calculations were performed at 224 and 400 Hz, with source depths (48 and 29 m, respectively) chosen to correspond to the SWARM experimental arrangement. A static propagation medium is assumed during acoustic transmission. The complex field as a function of depth and time is stored for further processing. Modal quantities obtained from the full acoustic field are estimated by performing a modal decomposition of the time-dependent field at the array; the acoustic eigenmodes ϕ used in the decomposition were computed from the deterministic sound speed profile $\overline{c}(x_{\mathrm{array}},z)$. Since $\delta c(x,z,t)$ is at most 1%–2% of $\overline{c}$, the error incurred is small; a discussion of mode leakage is given in Ref. 3. The resulting mode amplitudes d_ε are given by the weighted inner products

$$d_\varepsilon(j,x_{\mathrm{array}},f)=\int_0^H \frac{\phi(j,z,f)P_\varepsilon(x_{\mathrm{array}},z,f)}{\rho(z)}\,dz, \qquad (10)$$

where the subscript ε denotes an arbitrary time step in the simulation and j is the mode number. Modal intensity is defined by $d_\varepsilon d_\varepsilon^*$, where * is the complex conjugation operator.

B. Simulation results

With the specification of a background sound speed profile, bathymetry, and bottom parameters, numerical experiments were performed for two classes of time-dependent water column environments depending on the type of dynamic isopycnal displacements included in the simulation. One case introduced only the spatially localized solitary wave contri-

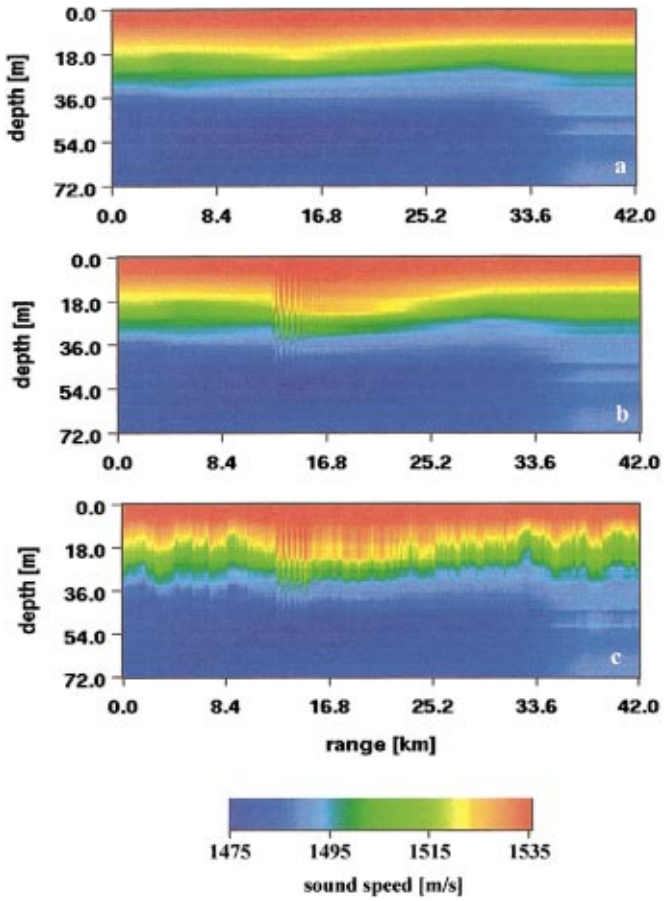


FIG. 4. Range-dependent sound speed distributions reconstructed for fixed time t_o. (a) Deterministic background $\overline{c}(x,z)$, (b) background and solitary wave packet contributions $\overline{c}(x,z)+\delta c_p(x,z,t_o)$, and (c) same as (b) but including the spatially diffuse internal wave component: $\overline{c}(x,z)+\delta c_p(x,z,t_o)+\delta c_d(x,z,t_o)$.

bution to the displacement, while another contained both diffuse and spatially localized contributions. For both cases the internal wave field was updated in 1 min time steps and the acoustic field computed for each of the resulting sound speed environments. As a reference, the approximate positions of the solitary wave packets obtained from Eq. (8) are given throughout the track at several fixed times in Fig. 6. The direction of packet propagation is toward the source located at $r=0$, with the leading edge traveling at speeds of approximately 0.73 to 0.8 m/s as given in Fig. 3(d). Packets are generated every M2 tidal cycle (period 12.42 h) and during the cycle there are intervals when two packets appear simultaneously within the 42-km track. The packets are labeled A and B as an aid for future discussion. For all simulations, the source region for the wave packets is several kilometers seaward of the array. The tail of the wave packets, mapped into sound speed and shown in Fig. 4(b), represents the observed return of the isopycnal depression to the unperturbed state and is approximately 10 km in length for the packets represented in this simulation. The isopycnal depression and subsequent relaxation imply a similar response in the thermocline.

Defining the full field intensity by $I_\varepsilon = P_\varepsilon^* P_\varepsilon$, transmission loss is given by $-10\log I_\varepsilon$ and is shown at 400 Hz in Fig. 7 for loss over the entire track for a particular time

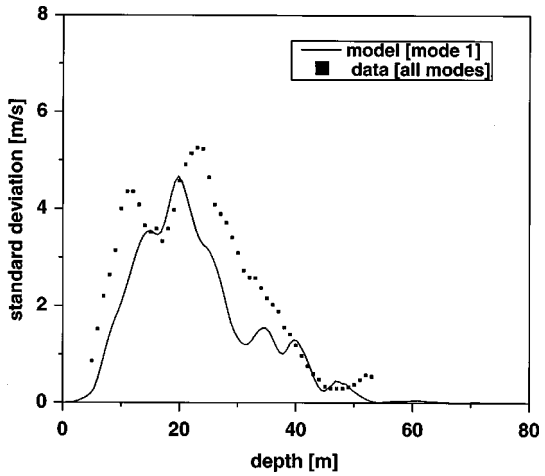


FIG. 5. Measured and modeled depth dependence of the sound speed fluctuations for environment including only the spatially diffuse internal wave field.

instant ($t=600$ min). Note that geometric ($1/R$) spreading loss has been removed to preserve detail in the color scale and this convention is followed throughout this article. A continuous red curve between approximately 55 to 90-m depth represents the measured bathymetric variations along the track.

The transmission loss for the deterministic environment $\overline{c}$ is given in Fig. 7(a) as a reference. Loss for the case which includes both a single solitary wave packet B with its leading edge at range 19 km and the trailing depression of packet A near the source (Fig. 6, $t=600$ min) is shown in Fig 7(b) while Fig. 7(c) contains the environment in Fig. 7(b) as well as the diffuse, linear internal wave field. The presence of multiple acoustic modes causes modal interference patterns in all three cases with the patterns changing from (a) to (c). Based on previous work,[3] the changes indicate a redistribution of energy among the modes due to mode stripping and mode coupling with the latter induced by spatial variations in sound speed along the path. A comparison between Figs. 7(a) and 7(b) shows that an abrupt change in modal structure occurs at about 19 km, the range at which the leading edge of

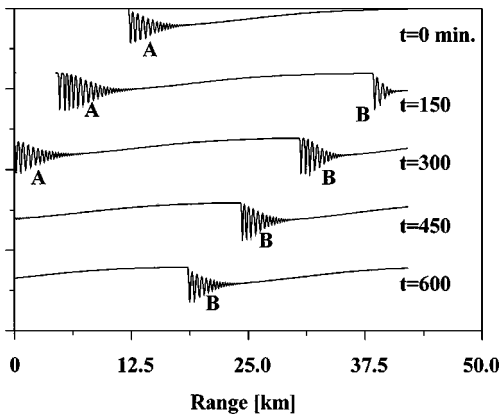


FIG. 6. Illustration of the spatial location of the solitary wave packets along the acoustic propagation path as a function of time. Two packets separated by a diurnal tidal cycle are shown at different times during the simulation. The acoustic sources are located at $r=0$ with the receiving array at $r=42$ km.

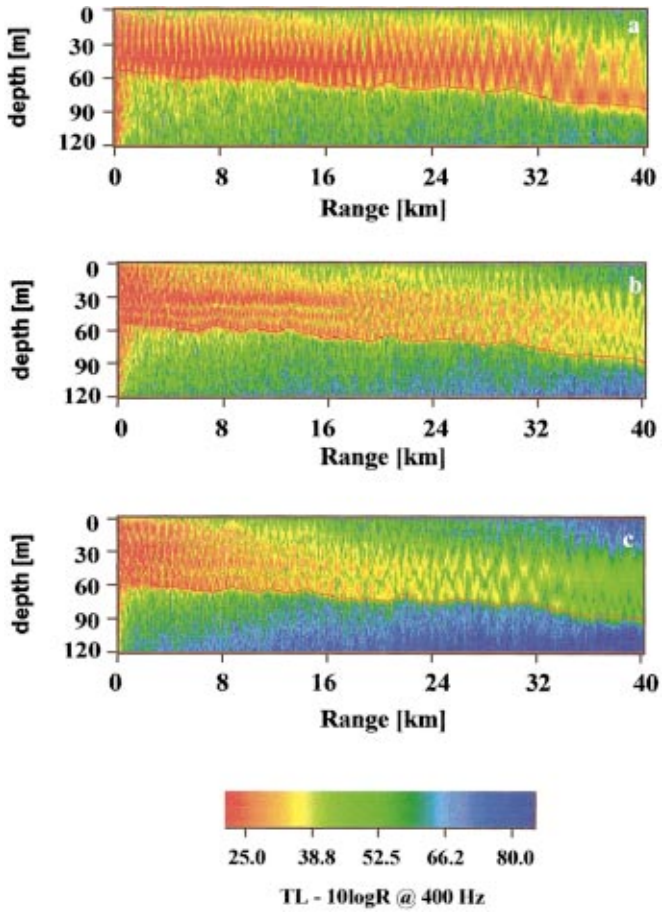


FIG. 7. Range and depth dependence of 400-Hz transmission loss for a fixed time corresponding to $t=600$ min in Fig. 6. (a) Deterministic loss (internal wave field absent), (b) loss in the presence of solitary wave packets and deterministic profile, and (c) loss in the presence of solitary wave packets, diffuse contribution, and deterministic profile.

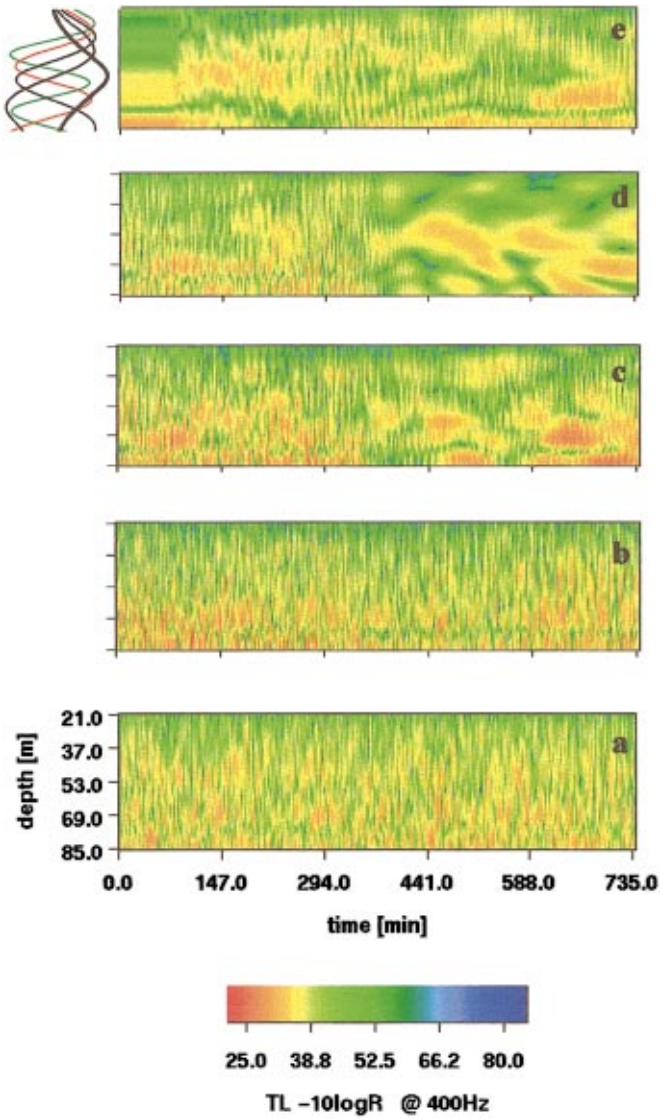


FIG. 8. Depth and time dependence of transmission loss at array location for a 400-Hz signal. All cases include the deterministic sound speed distribution, with various combinations of internal wave components: (a) spatially diffuse, (b) spatially diffuse and solitary wave packets, (c) solitary wave packets A and B, (d) packet A, and (e) packet B.

packet B is located. In addition, the modal structure for the first 19 km is different in Figs. 7(a) and 7(b) due to the influence of the recovery phase of packet A on the modal energy distribution in the vicinity of the source. Theoretical studies[10,11] imply that mode coupling is the cause of these interference changes. One cannot rule out, *a priori*, a component due to adiabatic propagation. The criterion for adiabatic invariance of two modes is dependent on the mode interference length and the overlap between modes at depths where the horizontal gradient of δc is significant.[10] While some adiabatic propagation may occur, simulations discussed later in this section show strong time-dependent anticorrelation between mode pairs, indicating mode coupling is an important characteristic of the propagating field in this environment. Note that only the water column environment is changing from Figs. 7(a)–7(c), though mode coupling also depends on both the bathymetric and sediment/volume variations along the track. Intensity fluctuations as a function of range and depth can be significant here (5–15 dB) and exhibit time dependence as the internal wave field evolves. This is true when either one or both components are present.

Analyses that follow are performed on the field at range $r=42$ km. The time variation of the calculated 400-Hz transmission loss at the location of the NRL array (range 42 km) is given in Fig. 8. Each plot displays depth versus time maps of transmission loss over a 12.4-h period. The loss was computed over 740 separate pressure field calculations taken in 1-min intervals as the internal wave field evolved in space and time. The receiver depth interval (21–85 m) is nominally the interval spanned by the NRL array in the SWARM experiment. In all cases, the deterministic sound speed profile $\overline{c}(x,z)$ is present in combination with different internal wave contributions $\delta c(x,z,t)$. Transmission loss for the case including only the spatially diffuse internal wavefield δc_d is shown in Fig. 8(a), while Fig. 8(b) corresponds to the same realization of the diffuse field but with the inclusion of the nonlinear wave packets A and B represented by δc_p. The spatial location of the packets as a function of time is referenced in Fig. 6. Viewing Fig. 8(a), there is little systematic variation in depth for transmission loss at depths greater than about 35 m, although at depths less than 35 m the transmission loss is, on average, a bit higher over the 12.4-h time

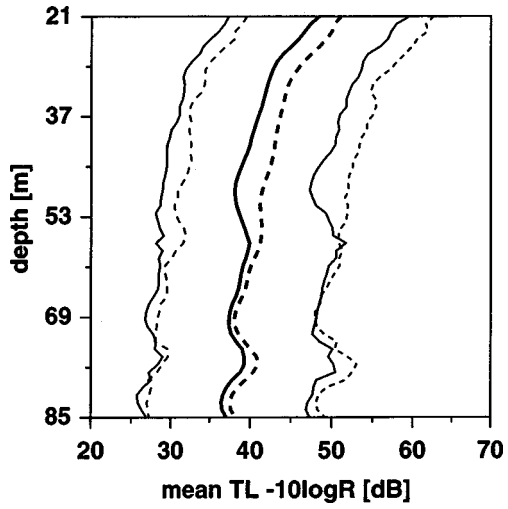


FIG. 9. Range averaged transmission loss bracketed by two standard deviation curves: Solid curves correspond to the case in Fig. 8(a), dotted curves correspond to the case in Fig. 8(b).

window. A comparison of Figs. 8(a) and 8(b) indicates a moderate variation in transmission loss with depth in Fig. 8(b), particularly below about 50 m. In addition, the average loss appears to be slightly greater in the upper 30 m of the array. These subjective observations are illustrated more clearly by Fig. 9, which shows time-averaged transmission loss (center traces) as a function of depth for losses shown in Figs. 8(a) and 8(b) along with two standard deviation (dB average) brackets. In both cases, the rms deviations are about 5 dB. The additional loss in Fig. 8(b), accompanying the introduction of the packets, is caused by energy transfer from mode one to higher order modes. Additionally, the temporal variation in Fig. 8(b) has the appearance of being ''less stationary'' than in Fig. 8(a). These spatial and temporal differences between Figs. 8(a) and 8(b) result from the introduction of the spatially localized packets.

Contrast the result in Fig. 8(a) with that in Fig. 8(c), the latter illustrating the environment for the case where only the solitary wave packets are present in the waveguide. It is helpful to interpret these two figures as loosely representing transmission loss for two extremes of the internal wave ''signal to noise ratio,'' i.e., a measure of the relative ''strength'' or fluctuation energy of the wave packet contribution in the sound speed perturbation field to that of the spatially diffuse contribution. In Fig. 8(a), the ratio is zero (no wave packets present) while in Fig. 8(c) it is infinite (no diffuse component present). Note that as the diffuse contribution to the internal wave field decreases, Fig. 8(b), representing an intermediate signal-to-noise case, transforms into Fig. 8(c) in a complicated manner since the acoustic field intensity is a nonlinear functional of the sound speed distribution. A more pronounced depth and time dependence of the transmission loss is evident in Fig. 8(c), with a qualitative difference in the space–time dependence for periods before and after $t \approx 350$ min. During the first 350 min, transmission loss in Fig. 8(c) is characterized by a relatively rapid temporal modulation throughout depth and a (relatively) weak depth dependence while for time periods greater than 350 min, the modulation period has lessened somewhat and the depth and time dependence have become more pronounced. Referring to Fig. 6, the leading edge of the packets throughout the time window indicates that between $t \approx 250$ and 350 min both the leading edge of packet A and its subsequent solitary wave depressions have passed over the source position. Approximately 350 min into the simulation, the source's local sound speed environment is dominated by the depression of the thermocline associated with the recovery of the internal tide depressed isopycnal surfaces back to their equilibrium positions. During the same time period packet B began to cross the receiving array while moving toward the source position. This situation was often present in the experimental data,[4] with additional packets entering the track from different directions (a spatially anisotropic effect not modeled here). It is difficult to separate the effects of individual packets on transmission loss in the case shown in Fig. 8(c), but it is reasonable to assume that at least some of the qualitative change in the space–time transmission loss illustrated in Fig. 8(c) is related to the change in shape between oscillatory isopycnal depressions caused by the solitary wave packet and the recovery phase of the long period isopycnal depression caused by the internal tide as it moves toward the source. To clarify the effect on transmission loss due to packets at the source and receiver positions, Figs. 8(d) and 8(e) show, respectively, the transmission loss when only the packet approaching and passing over the source is present (packet A) and when only the packet B traversing the receiver position is present. Figure 8(d) clearly indicates some of the differences discussed above for times less than or greater than about 350 min. The high-frequency modulation pattern is absent after that time; the absence is attributed to the smooth nonoscillatory return of the isopycnal surface to equilibrium specified in Eq. (8) by $I(x,t)$, during which time the depression of the thermocline was reduced. For periods greater than 350 min, the high-frequency temporal modulation in Fig. 8(c) is caused by packet B's presence in the transmission path near the array, as shown in Fig. 8(e). Note that for approximately the first 100 min the transmission loss in Fig. 8(e) is depth dependent but time invariant. This result is expected because the packet does not enter the simulation window until that time; for earlier periods, the array senses only the time-invariant field intensity associated with propagation through the deterministic sound speed distribution $\overline{c}(r,z)$. After the packet enters the track and begins to pass by the array, there is a shift in modal structure relative to the time-invariant mode configuration. The insert at the left shows the first five acoustic modes estimated from $\overline{c}$ at the array location and indicates that energy that was initially in mode four is transferred primarily into mode one as the environmental perturbation caused by packet B influences acoustic propagation.

Estimates of the temporal covariance were made from the calculated fields shown in Figs. 8(a) and 8(b) using averages over five consecutive 1-h time series segments. For selected depths, this computation was performed twice for data in each of the two figures, once for the first 300-min time window and once for the last 300-min window. Because of the nonstationary temporal character of the wave packet contribution [Figs. 8(c)–8(e)] it was felt that this separation provided a better averaging sequence than an average over

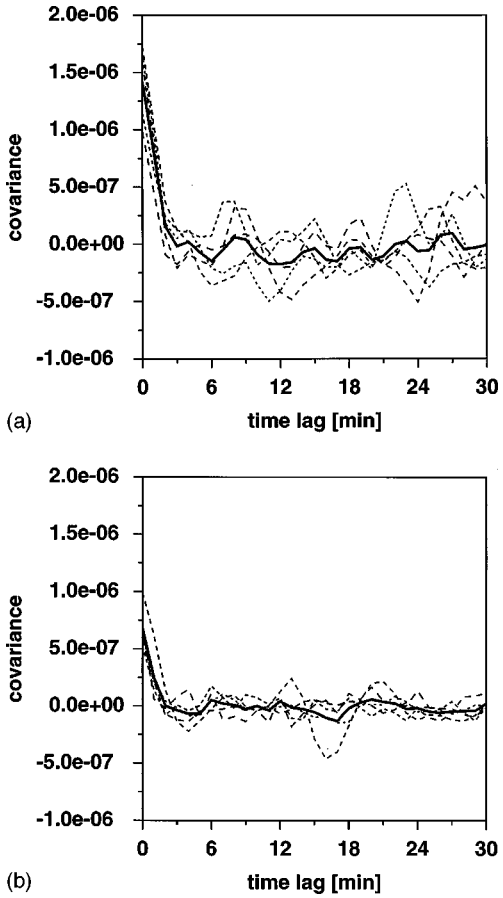


FIG. 10. Averaged temporal covariance (solid lines) computed at $z=53$ m for a 400-Hz signal. (a) Diffuse internal wave field and (b) combination of solitary packets and diffuse component. Dotted curves represent convariance of 1-h segments constituting the averages.

the entire tidal cycle. Typical examples are shown in Figs. 10(a) and 10(b) for a receiver depth of 53 m. The solid lines denote the average covariance, the dotted curves give the individual contributions to that average. Results for a spatially diffuse internal wave contribution are shown in Fig. 10(a) while Fig. 10(b) includes both the solitary wave and diffuse internal wave components. Decorrelation times vary from 2 to 4 min at 400 Hz, consistent with observations from matched filtered SWARM data and other theoretical results.[11] Differences are small between coherence times associated with the two cases discussed in reference to Figs. 8(a) and 8(b). Coherence times associated with Figs. 8(c)–8(e) would be longer over much of the water column for selected periods; those environments were not indicative of the actual oceanographic situation discussed here but could be present in other geographical locations.

Further insight into the intensity variations was obtained by modal decomposition of the acoustic field at the array position. The time-dependent modal intensities were computed from Eq. (10) for several modes at a frequency of 400 Hz and the results are shown in Figs. 11 and 12. The environments in which these plots were obtained are the same as those in Fig. 8 and are shown in the same sequence. The amplitudes of seven modes were computed and it was found that the first five modes had the dominant contributions to the field. We first consider the case where only the spatially diffuse internal wave field is added to the deterministic component. A comparison among modes 1–3 in Fig. 11(a) and modes 4 and 5 in Fig. 12(a) indicates that none of these modes strongly dominates during the entire 12.4 h of the simulation, although on average there is slightly more energy in mode 1 and a relatively weaker contribution from mode 5. This result reflects the general impression given by the full field image shown in Fig. 8(a), where the absence of a strong depth dependence in the transmission loss is indicative of a lack of single mode dominance. A comparison with Figs. 11(b) and 12(b) shows an overall drop in modal intensity for modes 1–3 including an additional drop starting around $t=350$ min, just after the multiple isopycnal depressions in packet A have passed by the source but with the recovery phase of that packet located within the propagation path in the vicinity of the source. This additional drop in modal intensity lasts for approximately 4 h, after which mode one energy is enhanced significantly. Modes 4 and 5 are not strongly affected during this period. The modal energy drop can be viewed as a slight increase in transmission loss in Fig. 8(b) during that time period, relative to transmission loss in adjacent time windows. The modal time variation in the ''high'' signal-to-noise ratio case is given in Figs. 11(c) and 12(c). Several examples of modal energy transfer are illustrated here, though they are somewhat obscured by the presence of both packets. For example, a spike in mode one intensity appearing at about 100 min appears to be correlated with an intensity drop in modes 4 and 5. A clearer instance occurs after 400 min, which shows anticorrelation between mode 2 and mode 3 power levels indicative of energy transfer between these modes. Comments on the temporal modulation pattern of transmission loss discussed earlier are also relevant in the modal intensity plots of Figs. 11(d) and 12(d) for the single wave packet propagating near the source position. The high-frequency modulation is present in the first 6-h segment, and completely absent during the recovery phase of the isopycnal displacement. While the detailed variations of the time series will differ for different parameters of the wave packet, the qualitative behavior should remain intact. During the recovery phase, modes 1 and 4 tend to dominate and exchange energy, as seen by the anticorrelation of their time signatures during this period. Since the modal eigenfunctions are depth dependent, these energy transfers are embedded in the full field transmission loss plots viewed in Figs. 8(c) and 8(e) as time variations in the transmission loss depth dependence. As the leading isopycnal depressions from packet B pass over the array, mode 4 energy [initially the largest contribution—see Fig. 12(e)] drops quickly as it transfers much of its energy into mode 1. Referring to Fig. 8(e), this transfer is seen as a decrease in transmission loss at mid-array depth and is followed by a complex mode coupling arrangement as shown in Figs. 8(e), 11(e), and 12(e).

The situation is somewhat less complex for the case of propagation at 224 Hz, since only three modes contribute significantly to the field at the array location. Time-dependent transmission loss at the array is shown in Fig. 13 with the same format as Fig. 8. Modes 1–3 dominate the arrival structure at range 42 km while the mode insert in the

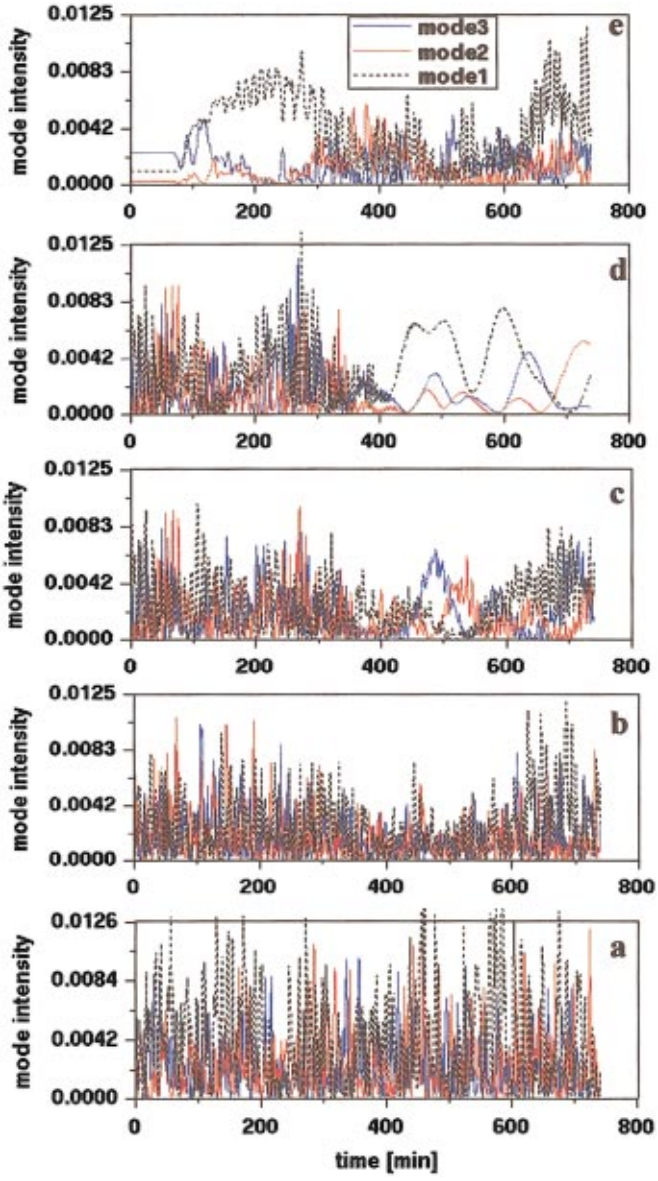


FIG. 11. Time variation of mode intensity at the array for the first three modes at 400 Hz. Internal wave contributions in (a)–(e) in the same format as given in Fig. 8.

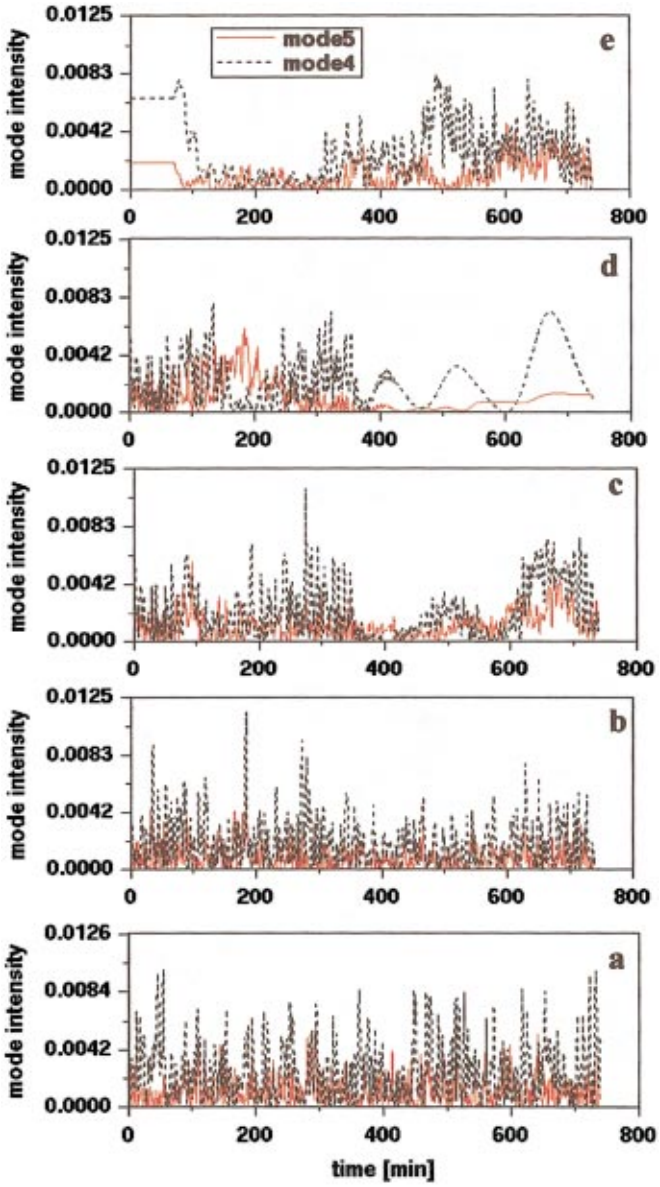


FIG. 12. Time variation of mode intensity at the array for modes 4 and 5 at 400 Hz. Internal wave contributions in (a)–(e) in the same format as given in Fig. 8.

leftmost portion of Fig. 13(e) indicates that, in the deterministic case, mode 2 contained most of the energy at the receiver. The depth dependence of transmission loss is more regular in Figs. 13(a) and 13(b) than in Figs. 8(a) and 8(b) corresponding to the 400-Hz case with a 29-m source depth. A comparison of Figs. 13(a) and 13(b) indicate similar structures where, on average, loss decreases with depth over the array. The time variability of mode intensity at 224 Hz is shown in Fig. 14 for modes 1–3. Anticorrelation between modes 1 and 2 is clearly visible in Figs. 14(c)–14(e) and indicates that these modes are strongly coupled. Temporal decorrelation times, illustrated in Fig. 15 are, on average, somewhat longer than those for the 400-Hz signals discussed above.

A notable feature of the time variation for transmission loss is the quasi-periodic nature of the loss over all depths commonly seen in the full field loss pictures, with the modal variations mirroring this time variation. This is distinct from the bloblike time-dependent changes occurring due to mode coupling discussed above. While a detailed analysis of this effect can be performed for simplified environments[11] using a full coupled mode theory, the complex situation here precludes such an analysis. However, the origin of the time variability can be understood from the following idealized example. Assume a rigid bottom waveguide of constant depth. From $x=0$ to $x=x_o$ the sound speed is a constant c and for $x>x_o$ the sound speed is another constant $c+\delta c$, where δc induces mode coupling at x_o. Consider the nth mode propagating in $\{0,\ x_o\}$ and specified by $A_n e^{ik_n x}$, where A_n contains the depth-dependent mode structure and radial spreading term. Let us further assume that at the point $x=x_o$ this mode transfers all its energy, via mode coupling, into modes l and m so that for $x=R>x_o$, the pressure field p_{ml} at an array located at range R is represented by the sum of these two modes:

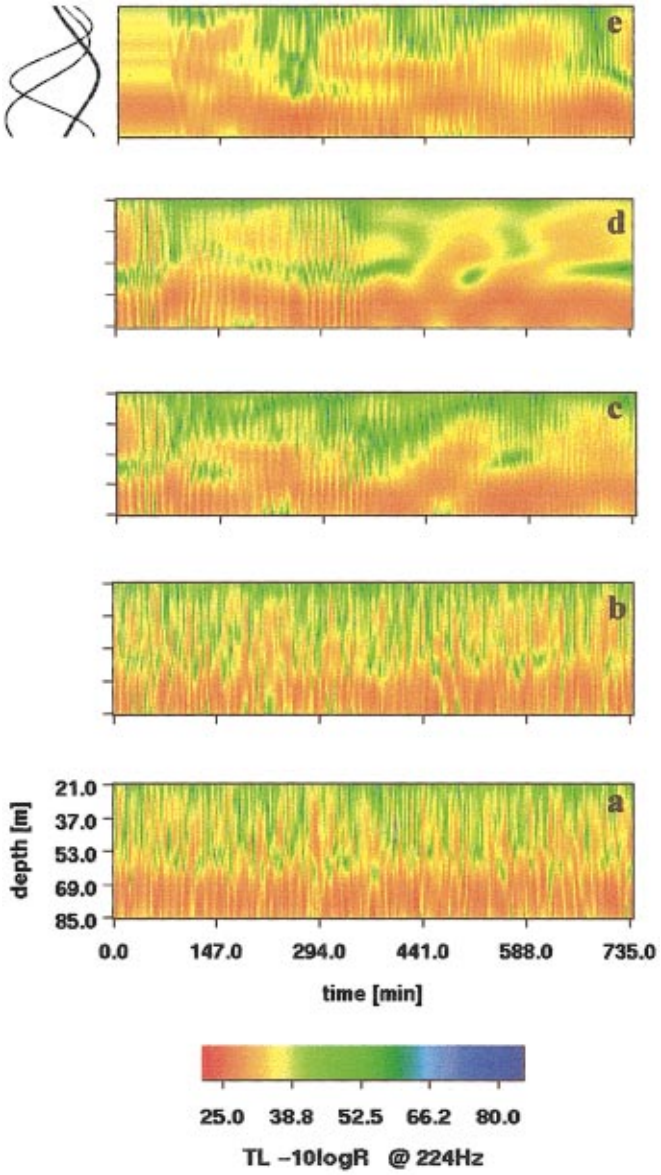


FIG. 13. Depth and time dependence of transmission loss at array location for a 224-Hz signal. Same format as Fig. 8.

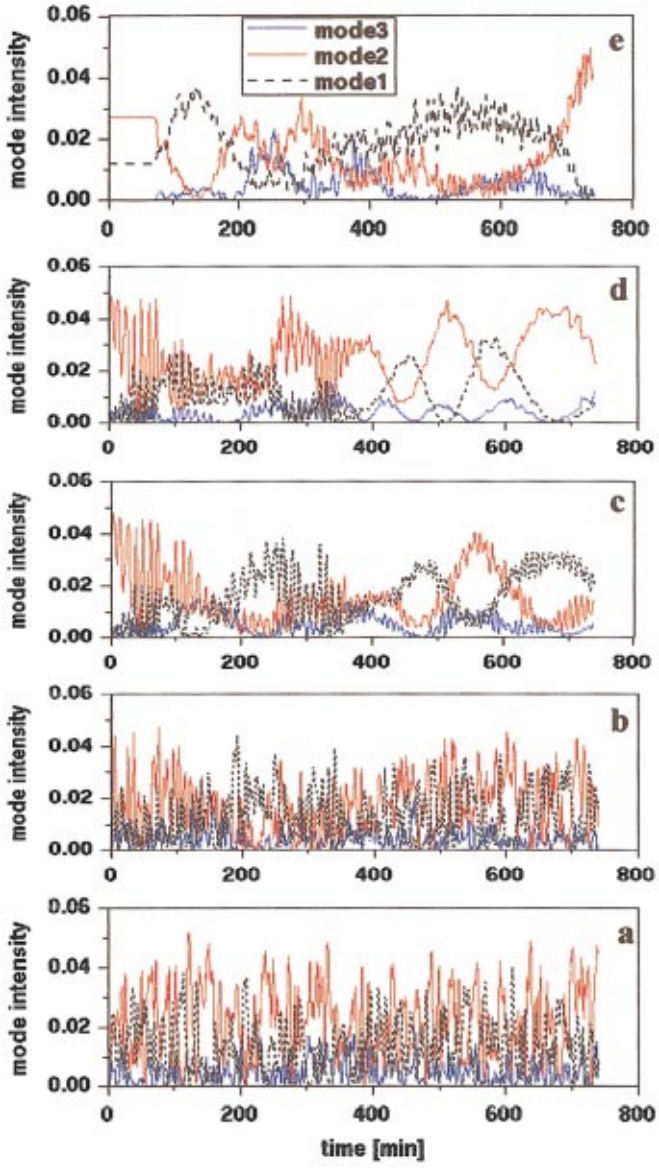


FIG. 14. Time variation of mode intensity at the array for modes 1–3 at 224 Hz. Same format as Fig. 8.

$$p_{ml}(x=R,z)=B_{nl}e^{ik_n x}e^{ik_l(R-x_o)}+C_{nm}e^{ik_n x}e^{ik_m(R-x_o)}. \quad (11)$$

Now let the location of the sound speed jump, originally fixed at location x_o, move at a speed v so that the location where mode coupling occurs changes as $x_o \rightarrow x_o \pm vt$. Under this condition the field intensity at the array, I_{ml}, is given by

$$I_{ml}(x=R,z,t) = |B_{nl}|^2+|C_{nm}|^2+2\,\mathrm{Re}\{B_{nl}C^*_{nm}e^{i[(k_m-k_l)(R-x_o)]} \times e^{\pm i(k_m-k_l)vt}\}. \quad (12)$$

The third term in this expression represents interference between modes *l, m*. A moving boundary describes the time-dependent change in the location of the sound speed gradient where the mode coupling occurs; it introduces a temporal modulation factor whose period depends on the modal interference wavelength and which serves to modulate the depth dependence of the intensity. When $v=0$, Eq. (12) reduces to the range-dependent modal beating pattern seen in multi-mode waveguides. In the waveguide discussed in this article, there are multiple modes present, transferring energy among themselves at many locations between source and receiver due to distributed spatial gradients in sound speed that change in time. This causes additional interference terms to appear with different modulation frequencies and leads to the complex time-dependent pattern seen in the transmission loss plots.

V. SUMMARY AND CONCLUSIONS

This article describes the construction and application of a data-constrained environmental model for a shallow water region dominated by internal wave activity. The resulting water column sound speed distribution included both a linear, random field component that was spatially distributed throughout the waveguide and a spatially localized contribution in the form of nonlinear solitary wave packets. Oceanographic parameters used to construct the internal wave rep-

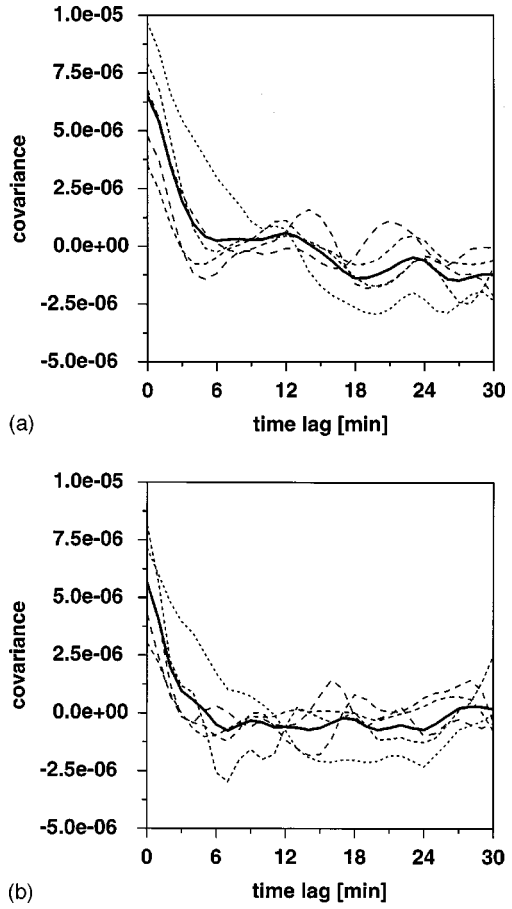


FIG. 15. Averaged temporal covariance (solid lines) computed at $z = 53$ m for a 224-Hz signal. Same format as in Fig. 10.

resentation were obtained from temperature and salinity measurements at selected ''quiescent'' locations along a 42-km track where acoustic backscatter and radar images indicated an absence of strong solitary wave activity. The full environmental model, comprised of the water column sound speed distribution and sediment/bottom parameters, was coupled to an acoustic propagation algorithm and some representative numerical simulations were performed to illustrate the spatial and temporal variability of the acoustic field intensity in a dynamic ocean environment. Consideration of the internal wave components separately and together yielded insight into the relative influence of these factors in characterizing the fluctuations of the acoustic transmission loss. Both components were shown to contribute to the overall structure of the field variability. The nonstationary nature of the field fluctuations was dominant when the diffuse component was absent or, by inference, weak relative to the contribution of the solitary wave packets. Though we have not considered results for the phase structure of the field in this article, our previous work[3] and other simulations based on the environment discussed here indicate that the diffuse internal wave field tends to randomize the phases and degrade the coherence of the acoustic field. Intensity coherence times computed with the diffuse contribution present are typically shorter than those computed when only the solitary wave packet(s) is present. Mode decomposition of the acoustic field indicated that time-varying mode coupling was prevalent and responsible for the temporal variation of the depth dependence of transmission loss. Both the modal and full field intensities were temporally modulated over depth and a simple model was invoked to give a qualitative explanation of this effect in terms of the modal interference wavelength and mode coupling at a moving boundary (sound speed gradient) interface.

A number of parameters had to be specified in order to characterize the sound speed environment. The influence of parameter estimation error on the simulation results was not considered here; a parameter sensitivity study of relationships between parameter variation and field variability was beyond the scope of this work. Due to the complexity of the mapping from sound speed fluctuations to acoustic field statistics, such a study would be important in judging the robustness of any statistical estimates based on this model. Another factor related to an error analysis is the general problem of quantifying the influence of model errors on numerical simulations. This source of error is rarely considered, even though its contribution can exceed that due to errors in parameter estimation.[21] In the approach presented here, the model error would be related to the appropriateness of the mathematical representation of the sound speed via the specific internal wave model discussed in Sec. II. The choice of model is usually dictated by the physics and/or by comparison with experiment, though it is often the case in complex systems that the physics is incompletely understood and that direct comparison with data is difficult. Other choices of internal wave model, e.g., higher order KdV formulations, will influence the space–time nature of the acoustic fluctuations and it would be useful to bound these variations due to modeling error. This form of error is a current subject of research in statistics.[21]

On the other hand, it is reasonable to conclude that for this particular model and a fixed source depth and frequency, the gross features of the acoustic field discussed above should remain intact when wave packet parameters and internal wave mode structure are changed slightly, although the magnitudes and time scales may be altered to some extent. It should also be noted that the acoustic pressure field is nonlinearly related to sound speed through the wave equation and acoustic field intensity does not satisfy the superposition principle, implying that, e.g., the transmission loss figures do not necessarily bear simple relationships to one another. The range-dependent nature of the Green's function for these environments, particularly in the cases where the internal wave packets dominate, precludes detailed analytical modeling of this environment.

Our shallow ocean representation employs a level of detail encompassing a middle ground between highly idealized environments, e.g., a two-layer thermocline model with flat bathymetry and homogeneous bottom parameters on the one hand, and full oceanographic environmental models obtained from direct numerical integration of the equations of fluid mechanics on the other. Oceanographic measurements are often undersampled over the spatial and temporal scales of interest in shallow water acoustics. The approach taken here tries to capture some major features of the oceanographic dynamics relevent to acoustic propagation by

including spatially distributed (though sparse) environmental information obtained from experimental data. The hope is that such a simplified approach will capture the dominant characteristics of the acoustic field without the need to incorporate a complete oceanographic description—which is not achievable with existing measurement resources.

ACKNOWLEDGMENTS

The authors would like thank Dr. Roger Oba for his help with the bathymetry and bottom parameter file structures. This work was supported with funds from the Office of Naval Research.

[1] S. Flatte, R. Dashen, W. Munk, K. Watson, and F. Zachariasen, *Sound Transmission Through a Fluctuating Ocean* (Cambridge University Press, ,Cambridge, 1979).
[2] D. B. Creamer, ''Scintillating shallow-water waveguides,'' J. Acoust. Soc. Am. **99**, 2825–2838 (1996).
[3] D. Tielbuerger, S. Finette, and S. Wolf, ''Acoustic propagation through an internal wave field in a shallow water waveguide,'' J. Acoust. Soc. Am. **101**, 789–808 (1997).
[4] J. R. Apel, M. Badiey, C. S. Chiu, S. Finette, R. Headrick, J. Kemp, J. F. Lynch, A. Newhall, M. H. Orr, B. H. Pasewark, D. Tielbuerger, A. Turgut, K. von der Heydt, and S. Wolf, ''An overview of the 1995 SWARM shallow-water internal wave acoustic scattering experiment,'' IEEE J. Ocean Eng. **22**, 465–500 (1997).
[5] B. G. Katsnel'son and S. A. Pereselkov, ''Resonance effects in sound scattering by internal wave packets in a shallow sea'' Acoust. Phys. **44**, 684–689 (1998).
[6] A. K. Liu, ''Analysis of nonlinear internal waves in the New York Bight,'' J. Geophys. Res. **93**(C10), 12317–12329 (1988).
[7] S. A. Kitaigordskii, V. P. Krasitskii, and M. M. Zaslavskii, ''On Phillip's theory of equilibrium range in the spectra of wind-generated gravity waves'' J. Phys. Oceanogr. **5**, 410–420 (1975).
[8] J. R. Apel, L. A. Ostrovsky, and Y. A. Stepanyants, ''Internal solitons in the ocean,'' The Johns Hopkins University Applied Physics Laboratory, Laurel, MD, Report MERCJRA0695 (1995).
[9] J. R. Apel, S. Finette, M. H. Orr, and J. F. Lynch, ''The 'dnoidal' model for internal tides and solitons on the continental shelf,'' submitted to J. Geophys. Res.
[10] J. C. Preisig and T. F. Duda, ''Coupled acoustic mode propagation through continental-shelf internal solitary waves,'' IEEE J. Ocean Eng. **22**, 256–269 (1997).
[11] T. F. Duda and J. C. Preisig, ''A modeling study of acoustic propagation through moving shallow water solitary wave packets,'' IEEE J. Ocean Eng. **24**, 16–32 (1999).
[12] K. V. Konyaev, D. E. Leikin, K. D. Sabinin, Y. I. Tuzhilkin, and V. R. D'yachenko, ''Correlations between internal solitons and amplitude variations of a sound signal on a fixed path,'' Acoust. Phys. **44**, 476–485 (1998).
[13] A. V. Gurevich and L. P. Pitaevskii, ''Decay of initail discontinuity in the Kortweg-de Vries equation,'' JETP Lett. **17**, 193–195 (1973).
[14] A. V. Gurevich and L. P. Pitaevskii, ''Nonstationary structure of a collisionless shock wave,'' Sov. Phys. JETP **38**, 291–297 (1974).
[15] F. S. Henyey and A. Hoering, ''Energetics of borelike internal waves,'' J. Geophys. Res. **102**, 3323–3330 (1997).
[16] S. Pond and G. L. Pickard, *Introductory Dynamic Oceanography*, 2nd ed. (Pergamon, New York, 1983).
[17] S. Finette, J. A. Apel, M. H. Orr, and the SWARM Group, ''Acoustic field propagation through a shallow water waveguide dominated by internal waves,'' J. Acoust. Soc. Am. **102**, 3062(A) (1997).
[18] J. R. Apel, *Principles of Ocean Physics* (Academic, New York, 1990).
[19] T. C. Yang and K. Yoo, ''Internal wave spectrum in shallow water: Measurement and comparison with the Garrett-Munk model,'' accepted for publication in IEEE J. Ocean Eng. **24**, 335–345 (1999).
[20] M. D. Collins, ''A higher-order parabolic equation for wave propagation in an ocean overlying an elastic bottom,'' J. Acoust. Soc. Am. **86**, 1459–1464 (1989).
[21] D. Draper, ''Assessment and propagation of model uncertainty'' J. R. Stat. Soc. B **57**(1), 45–97 (1995).

Tomographic inversion for sediment parameters in shallow water

Gopu R. Potty[a] and James H. Miller
Department of Ocean Engineering, University of Rhode Island, Narragansett, Rhode Island 02882

James F. Lynch
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Kevin B. Smith
Naval Postgraduate School, Monterey, California 93943

(Received 3 June 1999; accepted for publication 10 April 2000)

This article discusses inversions for bottom geoacoustic properties using broadband acoustic signals obtained from explosive sources. The experimental data used for the inversions are SUS charge explosions acquired on a vertical hydrophone array during the Shelf Break Primer Experiment conducted south of New England in the Middle Atlantic Bight in August 1996. The SUS signals were analyzed for their time-frequency behavior using wavelets. The group speed dispersion curves were obtained from the wavelet scalogram of the SUS signals. A genetic algorithm (GA) was used for the inversion of sound speeds in the water column and compressional wave speeds in the sediment layers. The variations in the sound speeds in the water column were represented using empirical orthogonal functions (EOFs). A range-independent normal mode routine was used to construct the replica fields corresponding to the parameters. Comparison of group speeds for modes 1 to 9 and for a range of frequencies 8 to 200 Hz was used to arrive at the best parameter fit. An efficient hybrid optimization scheme using the GA and a Levenberg–Marquardt algorithm is presented. Linear perturbation methods were also used to ''fine tune'' the inversions and to obtain resolution and variance estimates. Analysis was also done to compute the degree of convergence of each of the parameters by explicitly calculating the Hessian matrices numerically. *A posteriori* estimation of mean and covariance was also done to obtain error estimates. Group speeds for the inverted sound speed fields provide an excellent match to the experimental data. The inverted sediment compressional speed profile compares well with *in situ* measurements. © *2000 Acoustical Society of America.* [S0001-4966(00)04907-9]

PACS numbers: 43.30.Ma, 43.30.Qd, 43.30.Pc, 41.30.Bp [DLB]

INTRODUCTION

Acoustic propagation in shallow water is greatly influenced by the properties of the bottom. Calculations of acoustic propagation characteristics based on geotechnical data obtained from cores generally give reliable results. But core surveys are time consuming and may be applicable to only a small area. Hence indirect bottom survey methods which are fast and cover large areas have been developed. Inversions for the properties of the bottom using acoustic data have been given much attention in underwater acoustics. The *in situ* measurements used as data in these inverse methods are the acoustic field or quantities derived from it. These approaches differ mainly according to the characteristics of the acoustic sources and measurements (travel time, phase, etc.) they utilize for the inversion. Tolstoy *et al.*[1] used low-frequency (20 Hz) data obtained from air-deployed explosive sources to tomographically estimate environmental parameters in a simulated deep water environment. Ratilal *et al.*[2] and Smith *et al.*[3] used a method based on ambient noise field directivity to invert for the sediment compressional speed. This method estimates a single compressional speed value as an average for the top layer of sediments. Rapids, Nye, and Yamamoto[4] designed and tested a 3-D small-scale high-resolution cross-well acoustic tomography system in shallow water. A damped least-squared inversion technique was used to construct compressional speed images from measured travel time data.

Tomographic inversions of sediment properties have been performed using both narrow and broadband sources. When a broadband acoustic source is used in a shallow water waveguide, the acoustic propagation exhibits dispersion effects. Group velocity dispersion characteristics have been successfully utilized for the inversion of geoacoustic properties by Lynch, Rajan, and Frisk[5] using a linear perturbation approach. They have applied this method with success to the data collected during the GEMINI experiment performed in the Gulf of Mexico, using a towed narrow-band acoustic source which output pure tones at 50 and 140 Hz. The perturbation approach breaks down the nonlinear problem into a linear one in the vicinity of the final solution. Hence, an accurate a priori model of the environment is required to achieve good estimates.

Explosives can be used as broadband sources as they have large power output, large bandwidth, and considerable energy at low frequencies. This ensures penetration of acous-

[a]Electronic mail: potty@oce.uri.edu

tic energy deep into the ocean bottom, enabling the determination of ocean bottom parameters to greater depths. The group velocities, i.e., the speeds at which energy is transported, differ for different frequencies and modes. In a shallow water waveguide, high frequencies generally arrive earlier whereas the low frequencies, which interact with the bottom and hence are more important in geoacoustic inversions, arrive later. The dispersion behavior can be used for the inversion of geoacoustic properties of the bottom. It should be noted that the lower frequency data and the higher modes in the dispersion curve should give a better estimate of the sediment features whereas the high-frequency data and lower modes will improve the water column sound speed estimate.

Dispersion effects can be observed by time-frequency analysis of the acoustic pressure signal recorded at sufficiently large distance away from the source. The group speed values are directly extracted from these time-frequency energy distributions. Time-frequency analysis is performed using wavelets. Wavelet analysis examines the frequency distribution of a nonstationary time series using a set of windows that have compact support in time (i.e., decay to zero quickly) and are band limited in the frequency domain. The wavelet transform is a localized transform in both time and frequency and this property can be advantageously used to extract information from a signal that is not possible to unravel with a Fourier or even windowed Fourier transform. The advantage of analyzing a signal with wavelets as the analyzing kernels is that it enables us to study features of the signal locally with a detail matched to their scale, i.e., broad features on a large scale and fine features on a small scale. A detailed description of wavelet analysis can be found in Ref. 6.

In this paper we will emphasize tomographic inversion technique based on these broadband dispersion curves. Nonlinear methods based on exhaustive searches are not constrained in performance by the availability of background information and, thanks to the large increase in computational abilities, are being widely used for inversion in recent times. An inversion scheme for sediment compressional speeds based on a genetic algorithm is used to invert the group speed curves.

The article is organized as follows. The genetic algorithm specifics are introduced in Sec. I. The hybrid inversion scheme is discussed in Sec. II followed by perturbation methods in Sec. II A. Section III covers the procedure for error estimation. The Shelf Break Primer Experiment is described in Sec. IV, followed by data analysis and sensitivity studies in Sec. V, and results and discussions in Sec. VI. The conclusions of the present study and future work planned are presented in Sec. VII.

I. GLOBAL OPTIMIZATION AND GENETIC ALGORITHMS

Global optimization schemes such as simulated annealing (SA) and genetic algorithms (GA) have been used increasingly in recent times for the inversion of underwater acoustic signals for bottom properties. Representative references would be Collins and Kuperman,[7] Hermand and Gerstoft,[8] and Gerstoft.[9] These methods rely on exhaustive searches, and the time required for the search is often very high. Collins and Kuperman[7] demonstrated that it is possible to accurately estimate the source location with limited *a priori* environmental information by expanding the parameter search space of matched field processing (MFP) to include environmental parameters, a method they called ''focalization.'' They used simulated annealing (SA) for searching the large focalization parameter space for optimal parameter values. This scheme was based on matching measured and modeled modal phases.

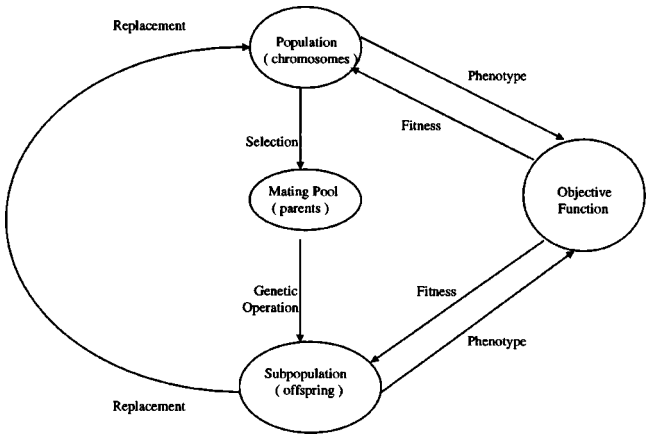


FIG. 1. A typical genetic algorithm cycle (Ref. 9).

Genetic algorithms (GAs) are nonlinear optimization schemes, highly efficient in optimizing discontinuous, noisy, highly dimensional and multimodal objective functions. It can be thought of as a ''smart'' Monte Carlo Search. A GA is not biased by an initial starting model, uses no gradient or curvature information, and has the ability to avoid local minima. Instead of selecting points in model space from a uniform distribution, the points are selected by repeated application of mathematical operators. The search is thereby channeled toward good solutions. Whereas SA is based on a single member meandering in a search space, the GA is based on a population which intercommunicates while meandering in a search space. A simple GA starts with a population of samples randomly generated from the model subspace which is defined by *a priori* bounds on the model parameters. These search limits are specified initially. The fitness of each member is computed based on the value of the objective function for that member. Then through a set of evolutionary steps the initial population evolves in order to become more fit. An evolutionary step consists of selecting a parental distribution from initial population based on the individual's fitness. The parents are then combined in pairs and operators are applied to them to form a set of children. The operators are traditionally crossover (recombination) and mutation (random bit change) operators. Finally the children replace part of the initial distribution to get a more fit population. The process of going from current to next population constitutes one generation in the execution of a GA. Tang *et al.*[10] give a detailed description of these operations. A typical GA cycle is shown in Fig. 1. In the study proposed here a GA will be used to optimize an objective function and

match measured and modeled modal dispersion characteristics of the broadband acoustic signal. The objective function for the GA inversions, based on minimization of group speed differences, will be modified to incorporate the nonuniform data quality. The objective function is of the form

$$E(\mathbf{m})=\sum_{i=1}^{N}\frac{[d_i-F_i(\mathbf{m})]^2}{\sigma_i^2};\quad i=1,2,...,N; \tag{1}$$

where $E(\mathbf{m})$ is the objective function for the mth parameter set and σ_i is the standard deviation associated with the ith data point. The numerator of Eq. (1) represents the mismatch between the observed data $(\mathbf{d},N\times 1)$ and the prediction $(F(\mathbf{m}),N\times 1)$ of the forward model. The observed data consists of group speed values calculated using the times of arrival for various modes and frequencies. The predictions are the theoretical group speed values calculated for the sound speed profile constructed using the model parameters.

II. HYBRID OPTIMIZATION SCHEME

Using a genetic algorithm does not guarantee that an exact global optimum will be found, even in an infinite amount of time. Because a GA makes no use of gradient or curvature information, it is not clear what type of point in error space the ''best model'' represents. It could be a local minimum, a global minimum, a saddle point, or none of these. We must also remember that a GA operates in a discrete subspace of the actual model space. For these reasons it is important to closely examine the region around the best model generated by a GA. To overcome these drawbacks many investigators have developed hybrid schemes combining GA with a local search method. Gerstoft[11] suggested a combination of global GA and a local Gauss–Newton method. Taroudakis and Markaki[12] proposed another hybrid scheme wherein the reference environment defined using matched field processing with a GA is subsequently used in connection with a modal phase inversion scheme. This linear modal inversion is meant to fine tune the results obtained through the matched field tomography. We use the optimum parameters obtained from the genetic algorithm inversion as the starting point for a local search using the Levenberg–Marquardt algorithm, which is more robust than the Gauss–Newton method. This algorithm uses a search direction which is a cross between the Gauss–Newton direction and the steepest descent direction. By applying this method at the end of the GA search, we can assess the quality of the GA solution locally and perhaps find a better solution.

A. Broadband perturbative modal inversion

Linear inversion methods give good estimates of the environmental parameters when good *a priori* knowledge of the environment is available. Lynch, Rajan, and Frisk[5] have successfully used a linear perturbative approach for the inversion of group velocity data for the sediment sound speed. The expression which relates the group velocity dispersion curve to the bottom sound speed profile is

$$\begin{aligned}\frac{1}{v_m^g}-\frac{1}{v_{nm}^{g(0)}}&=\left(\frac{1}{k_m}\right)\int_0^\infty\left\{\left[2\omega_0^{-1}\rho^{-1}(z)|\psi(\omega_0,z)|^2k^2(z)\frac{\Delta c(z)}{c(z)}\right]\right.\\&+\left[\left(\frac{1}{k_m v_m^{g(0)}}\right)\rho^{-1}(z)|\psi(\omega_0,z)|^2k^2(z)\frac{\Delta c(z)}{c(z)}\right]\\&\left.+\left[\rho^{-1}(z)\frac{d}{d\omega}|\psi(\omega_0,z)^2|k^2(z)\frac{\Delta c(z)}{c(z)}\right]\right\}dz,\end{aligned} \tag{2}$$

where v_m^g is the experimentally measured group velocity for the mth mode at ω_0, $v_m^{g(0)}$ is the group velocity of the mth mode at ω_0 for the background model, k_m is the mth eigenvalue for the background model, $\rho(z)$ is the density profile for the background model, $\psi(\omega_0,z)$ is the normalized mode function for the background model for the mth mode at ω_0, $c(z)$ is the background sound speed profile, $\Delta c(z)$ is the perturbation in sound speed from the background profile, and $k(z)$ is $\omega/c(z)$, where ω is the angular frequency.

Equation (2) is a Fredholm integral equation of the first kind which can be written as a discrete sum

$$d_i=\sum_j G_{ij}q_j,\quad i=1,2,...,N;\quad j=1,2,...,M, \tag{3}$$

or, in matrix form,

$$\mathbf{d}=\mathbf{G}\mathbf{q}, \tag{4}$$

where $\mathbf{d}$ is an $N\times 1$ data vector, $\mathbf{q}$ is an $M\times 1$ vector of unknown perturbations, and $\mathbf{G}$ is an $N\times M$ kernel matrix. The data vector consists of the differences between the measured group slowness and the group slowness for the background profile [left-hand side of Eq. (2)] and the unknowns are the perturbations in sound speed $[\Delta c(z)]$. Thus this linear tomographic system is characterized by an $N\times M$ Frechet matrix ($\mathbf{G}$) which is a linear mapping from the M-dimensional parameter space to the N-dimensional data space. The above linear problem can be solved by singular value decomposition. This is discussed in detail by Lynch *et al.*[5] To incorporate the nonuniform data quality Parker[13] suggested weighting the data with standard errors. Weighted data with unit variance is obtained by scaling the data by the standard errors

$$\mathbf{e}=\mathbf{\Sigma}^{-1}\mathbf{d}, \tag{5}$$

where $\mathbf{d}$ is the data vector, $\mathbf{\Sigma}$ is a diagonal matrix with $\sigma_1,\sigma_2,\ldots,\sigma_N$ as the diagonal elements, and σ's are the standard deviations associated with data. In our application these σ's can be approximately estimated from the time–frequency distributions based on the spread of the spectral lines. Multiplying both sides of Eq. (4) by $\mathbf{\Sigma}^{-1}$ and using Eq. (5),

$$\mathbf{e}=\mathbf{\Sigma}^{-1}\mathbf{G}\mathbf{W}\mathbf{y}, \tag{6}$$

where

$$\mathbf{y}=\hat{\mathbf{R}}\mathbf{q} \tag{7}$$

and $\mathbf{G}$ is the kernel matrix, $\mathbf{W}$ is the diagonal matrix of weights of a numerical quadrature approximation (trapezoi-

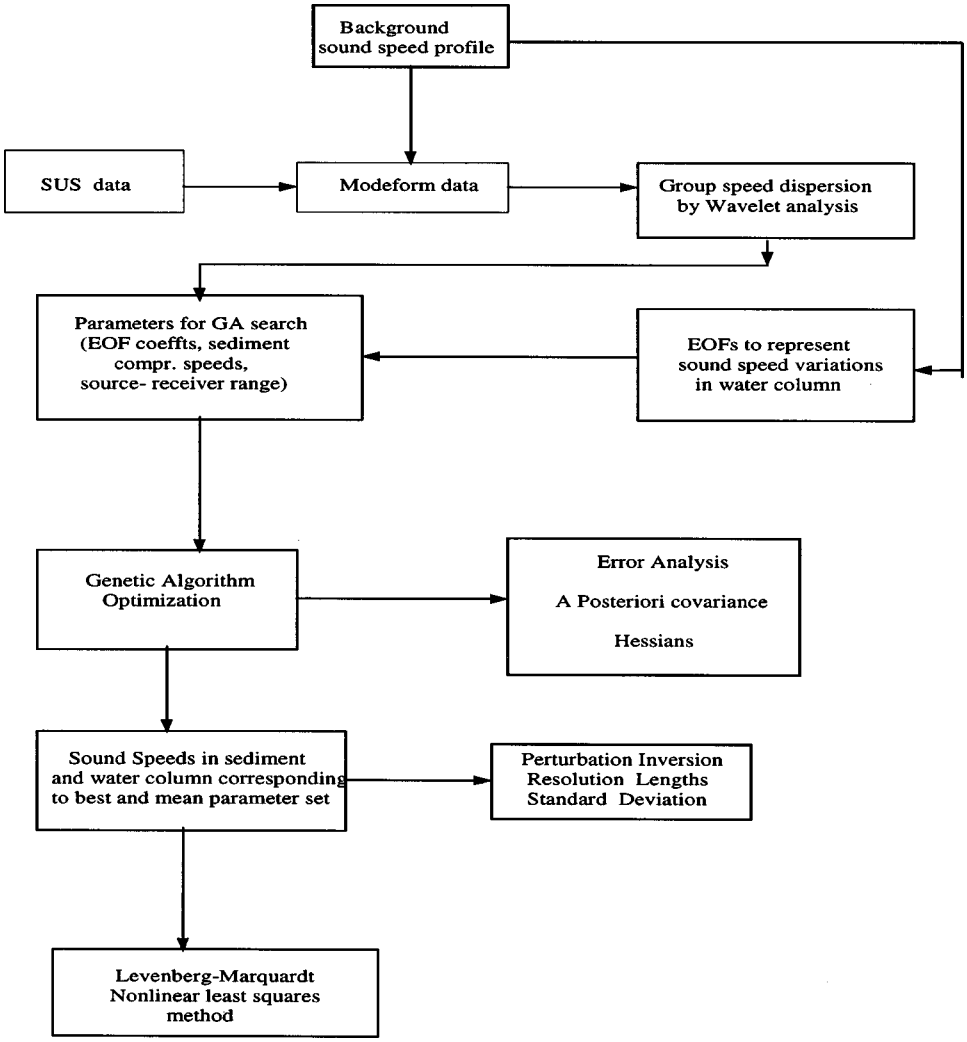


FIG. 2. Schematic representation of the steps involved in the inversion scheme.

dal rule in the present case), $\hat{\mathbf{R}}$ is the regularizing matrix, and $\mathbf{q}$ are the unknowns of the original problem $[\Delta c(z)]$. $\mathbf{W}$ is introduced here to stabilize the numerical implementation. For L_2 norm minimization $\hat{\mathbf{R}}$ can be taken as $\mathbf{W}^{1/2}$. These steps are discussed in detail by Parker.[13]

III. ERROR ESTIMATES AND VALIDATION OF THE INVERSION RESULTS

The goals of most inversions include not only finding a model which best fits the data, but also estimating error bounds on the parameters. Even though these error bounds cannot be considered as validation of the inversion, they do provide useful indication about the quality of the inversions. We have quantified the errors in our current work using different approaches based on different criteria. *A posteriori* covariance is estimated by defining an *a posteriori* model probability. Local error bounds of the model parameters are estimated using numerically evaluated Hessian matrices. Elements of the Hessian matrix are the second partial derivatives of the objective function with respect to the model parameters. They were evaluated numerically in the neighborhood of the best solution. In addition, resolution and variance estimates are also obtained using the linear perturbation approach.

In the present study the compressional speeds generated using the GA inversion are used as a background profile for the perturbation inversion. Thus the GA inversion is further refined using linear perturbation methods, as well as the Levenberg–Marquardt optimization method. The entire inversion process is schematically represented in the block diagram shown in Fig. 2.

A. *A posteriori* probability density and model covariance

During the GA optimization procedure, all the samples of the model space are stored and then later used to estimate *a posteriori* probabilities. From a Bayesian point of view, the solution of the inverse problem can be characterized by an *a posteriori* probability distribution of the model parameters. Hence, in addition to the best possible estimate, moments of the *a posteriori* distributions such as mean and covariance can also be calculated. Based on an analogy with SA, Sen and Stoffa[14] and Gerstoft[9] have used the Gibbs probability distribution to define the *a posteriori* probability density in the model space. The Gibbs probability $\sigma(\mathbf{m})$ is given by

$$\sigma(\mathbf{m}) = \frac{\exp[E(\mathbf{m})/T]}{\Sigma \exp[E(\mathbf{m})/T]}, \tag{8}$$

where $E(\mathbf{m})$ is the fitness function or equivalently the value of the objective function for the model vector $\mathbf{m}$. Here T is a control parameter similar to temperature used in SA and the sum is taken over all derived models. Choosing the energy of the fittest in the sample as the value of T will favor the fittest part of the population. Instead if the energy of the least fit sample is chosen as the value of T, a more even weighting of the population can be obtained. Experience has shown that a good temperature is the average of the best 50 samples.[15] We can estimate the mean and model covariance matrix by defining

$$\overline{\mathbf{m}} = \int \mathbf{m}\sigma(\mathbf{m})\, d\mathbf{m}, \tag{9}$$

which can be approximated by the sum

$$\overline{\mathbf{m}} = \Sigma \mathbf{m}\sigma(\mathbf{m}). \tag{10}$$

The covariance matrix can be obtained as

$$\mathbf{C_m} = \int (\mathbf{m}-\overline{\mathbf{m}})(\mathbf{m}-\overline{\mathbf{m}})^T \sigma(\mathbf{m})\, d\mathbf{m}. \tag{11}$$

This can also be written as

$$\mathbf{C_m} = \int \mathbf{m}\,\mathbf{m}^T \sigma(\mathbf{m})\, dm - \overline{\mathbf{m}}\,\overline{\mathbf{m}}^T \tag{12}$$

or as the sum

$$\mathbf{C_m} = \Sigma \mathbf{m}\,\mathbf{m}^T \sigma(\mathbf{m}) - \overline{\mathbf{m}}\,\overline{\mathbf{m}}^T. \tag{13}$$

The square roots of the diagonal terms of $\mathbf{C_m}$ are the standard deviations or the error bars of the model parameters from the mean. Similarly, the off-diagonal terms show how the different model parameters influence each other.

The model parameter values and their probabilities are stored for all the generations in the GA run. At the end of the GA run all the probability values are summed and each value is divided by the sum to derive the normalized *a posteriori* probability densities. Using this, the *a posteriori* mean and covariance matrix are evaluated.

We can also examine the locally defined error bounds on the model parameters by computing the covariance matrix using the Hessian matrix. When the data are uncorrelated and Gaussian, the covariance matrix can be computed using the following equation:[16]

$$[\mathbf{C_m}]_{\text{local}} = \sigma_{\mathbf{d}}^2 \left[\frac{1}{2} \frac{\partial^2 E}{\partial \mathbf{m}^2} \right]^{-1}_{\mathbf{m}=\mathbf{m}_{\text{est}}}, \tag{14}$$

where $\sigma_{\mathbf{d}}$ is the standard deviation of the error in the data. The variance of the model error is the diagonal of $\mathbf{C_m}$. Thus the uncertainty in a set of model parameters is the product of the uncertainty in the data and the second-order curvature of the error space about the point $\mathbf{m}_{\text{est}}$. It should be noted that by sampling a much larger model space, the GA gives a more realistic estimate of error bounds than this local method. We have evaluated error estimates using both methods and they seem to agree with each other for our particular case.

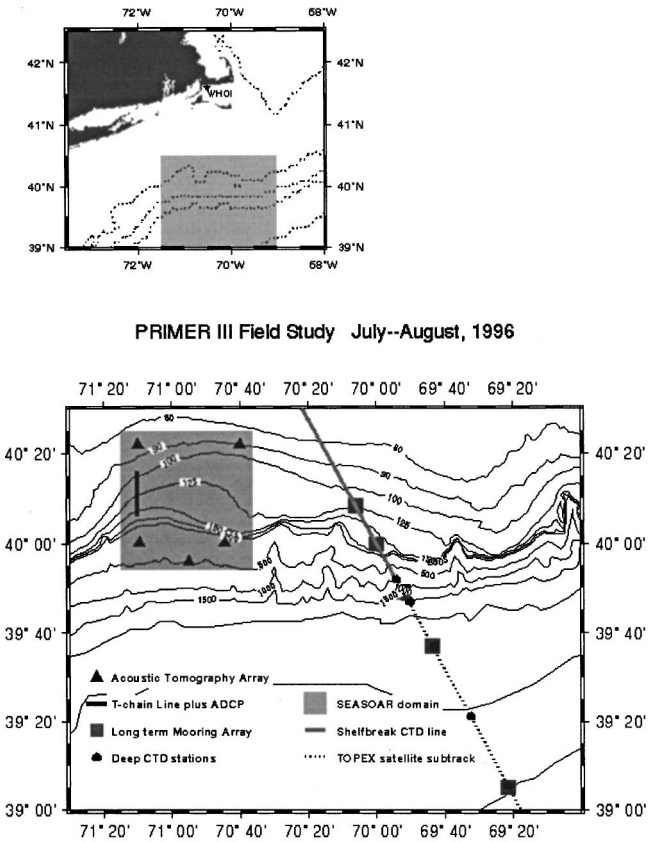


FIG. 3. Location of Shelf Break Primer Experiment. Locations of VLAs and the tomographic sources are also shown.

The model resolution length and standard deviation can be calculated based on the linear perturbation inversion using the singular value decomposition approach. This is described in detail by Lynch *et al.*[5]

IV. SHELF BREAK PRIMER EXPERIMENT

In the 1996–1997 Shelf Break Primer Experiment a number of oceanographic and acoustic measurements on the shelf break south of New England in the Middle Atlantic Bight during summer and winter conditions were conducted. Oceanographic observations mainly consisted of SeaSoar[17] hydrography, shipboard acoustic Doppler current profile (ADCP) measurements, and air-deployed expendable bathy thermographic (AXBT) drops. Some of these measurements were used in this study as background environmental information. Figure 3 shows the location of the experimental site and the positions of the vertical line arrays (VLA's), acoustic tomographic sources, and bathymetry. The acoustic component involved transmissions from moored tomographic sources and explosive SUS charges. The SUS component of the experiment involved the acquisition of broadband acoustic data on two vertical line arrays located on the continental shelf on the northwest and northeast corners of the experimental area, in water depths of approximately 90 m. A P-3 aircraft from the Naval Air Warfare Center in Patuxent Naval Air Station dropped over 80 MK61 explosive charges in an inverted F-pattern during both summer and winter cruises. Figure 4 shows the SUS drops along the slope and the loca-

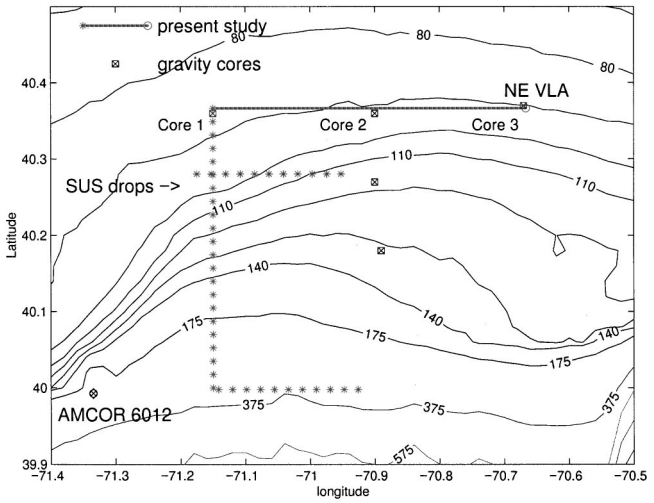


FIG. 4. The SUS drop locations at the experimental site. The AMCOR site is shown in the lower left corner of the figure. The propagation path corresponding to the present study is also shown. The gravity cores on this path (cores 1–3) are used in this study for the comparison and validation of the inversion.

tion of the Atlantic Margin Coring (AMCOR) program site 6012. Sediment compressional sound speeds at various depths were calculated at this AMCOR location using the Biot–Stoll model and then compared with the inverted values. The SUS charges were dropped at a spacing of one nautical mile. The SUS drop effort provides a temporal ''snapshot'' over a larger spatial area and a larger acoustic bandwidth than conventional ocean tomography measurements. Thus these shots complement the electronic tomography sources deployed in the experiment. MK61 SUS charges are 1.8 lbs of TNT and were set to detonate at a depth of 18 m in water depths varying from 80 to 300 m.

A. Geoacoustic data at the PRIMER site

The nature of the seafloor sediments in the continental shelf and upper slope regions of the Middle Atlantic Bight have been studied by various investigators (Knebel and Spiker,[18] Robb *et al.*,[19] Poag,[20] and Hatheway *et al.*).[21] Analyses of cores taken from this area reveals nearly horizontal layers of Pleistocene and early Holocene silty clays covered by various thicknesses of Holocene sands from 1 to 20 m.[18] In a location southeast of Primer site (73 °W, 39 °N) close to shelf break they estimated the average thickness of the surface sand layer to be 5 m using seismic reflection profiles and core samples. The sediment below the surface layer was texturally diverse and consisted of silty and clayey sands, sandy and silty clays, sand-silt-clays and clays. The clay layers are poorly sorted and very stiff. Layers containing more than 75% sand are also present, but they are intercalated with muddy zones. Gravel is also found in some of the cores. In another location at the upper continental slope (72 °50′ W, 38 °50′ N), Robb *et al.*[19] observed a thin surface layer (generally less than 2 m thick) of medium-grained Holocene sediments underlain by texturally diverse Pleistocene sediments composed of silty clay, silty sand, clayey sand, or sandy clay. The thickness of the surface Holocene sediment

TABLE I. Geoacoustic data at the AMCOR-6012 site (Ref. 24). These parameters were used for estimating sediment compressional speeds using Biot–Stoll theory.

Parameter	Unit	Value
Gain density	kg/m^3	2600
Fluid density	kg/m^3	1000
Grain bulk modulus	Pa	3.6×10^{10}
Fluid bulk modulus	Pa	2.25×10^{9}
Fluid viscosity	kg/m.s	1×10^{-3}
Shear specific loss		0.02
Volumetric specific loss		0.015
Added mass coefficient		0.2
Mean grain diameter–silty clay	μm	2.7

layer consisting of medium-to coarse-grained sand reaches 20-m thickness along the outer New Jersey shelf.[20] Poag also estimated the maximum thickness of the Pleistocene layer, consisting of clayey and silty sands and silty clays, to be 170 m in the same location (approximately 39 °N, 73 °W). Hatheway *et al.*[21] also reports Pleistocene layers consisting of silty clay and fine sand with occasional layers of pea gravel in the same location. On the upper continental slope, thick Pleistocene sequences consisting of silty and sandy clays,[20] silty clay units, alternating with silty sands and fine sands at some locations[21] were found. Over 50% of the clay sediments in the region have an effective stress friction angle of approximately 30 degrees,[21] which corresponds to silty sands or inorganic silt. The Pleistocene clays were mostly inorganic and have medium plasticity.

Broacher and Ewing[22] used high-resolution seismic reflection data to estimate seismic velocities in an area on the continental shelf south of Long Island, NY, in approximately 64 m of water. Sediment samples collected near this experimental site recovered coarse to fine sand. The top 30 m of sediments consisted of Pleistocene and younger sediments. They estimated the thickness of the Pleistocene sediments as 31 m and a velocity of 1750 m/s. It should be noted that the uncertainty associated with these measurements is high (450 m/s). McGinnis and Otis[23] obtained similar results using refracted arrivals in shelf areas of George's Bank and Long Island. They reported a mean velocity in the upper few tens of meters of sediments at locations close to 70 °W, 40 °N of the order of 1650–1675 m/s. They also estimated the vertical velocity gradients at this location to be 1.2 m/s/m.

Even though the areas adjacent to the Primer site were well investigated, very little geoacoustic data is available about the near surface sediments (top 1 to 100 m) within the present experimental location. The only published core data reported is from the AMCOR-6012 project. During 1976 the U.S Geological Survey conducted the Atlantic Margin Coring Project (AMCOR) to obtain information on the geotechnical and other properties at sites widely distributed along the continental shelf and slope of the Eastern United States.[21] The analysis of the cores obtained during this project provided information on the porosity, bulk density, and other geotechnical parameters down to a depth of about 300 m below seafloor.[24] The AMCOR-6012 site, at 39 °59.57′N and 71 °20.09′W, is near the southwest corner of the experimental area at a water depth of 263 m. The location of the drill

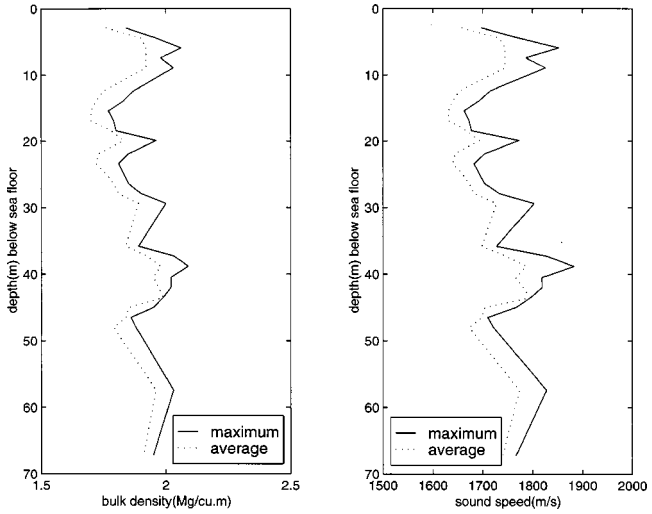


FIG. 5. Compressional sound speed at AMCOR 6012 site. The two curves show the maximum and average bulk densities and the compressional sound speeds calculated using them.

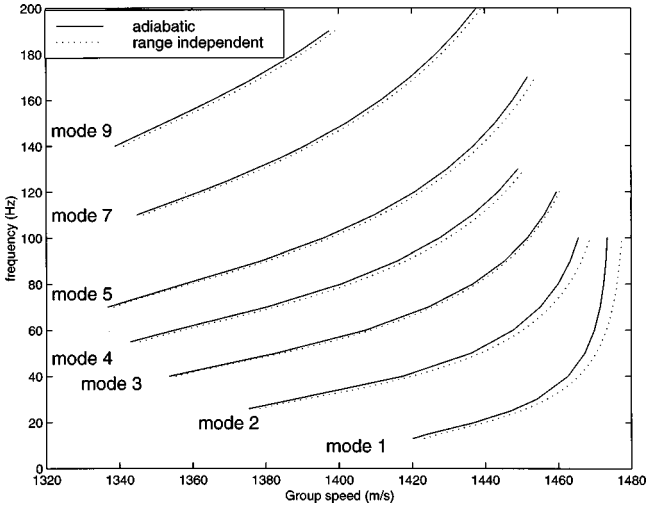


FIG. 6. Effect of range dependence of ocean sound speeds on group speeds.

site in relation to the SUS drop locations is shown in Fig. 4. Bulk densities at various depths and the compressional wave speeds calculated using the geotechnical parameters (Table I) based on the Biot–Stoll model[25] are shown in Fig. 5. The two profiles represent the average and maximum values of bulk density and the corresponding compressional speeds calculated based on them. In the absence of any error estimates these two curves can be used to gauge the spread of the values at each depth. The deviation from the mean is of the order of 30 m/s except at layers with high compressional speeds where it is nearly 75 m/s. Trevorrow and Yamamoto[26] obtained the compressional wave speed profile at the AMCOR 6010 site using gravity wave inversion technique. AMCOR 6010 is located south west of the Primer site in the outer continental shelf in approximately 69 m of water depth. It is interesting to note that the highest speed close to the surface (approximately 1800 m/s) is at 5-m depth for AMCOR 6012 whereas it is at 30 m for AMCOR 6010.

Since direct measurements of geoacoustic parameters were scarce in the experimental area, five gravity cores were taken to compare with the inversions. These five cores were taken at locations indicated in Fig. 4. The maximum core penetration in the shelf locations was approximately 1.5 m whereas it was 1.0 m at deeper slope locations. At these depths, core penetration was limited by the presence of sand. The core samples were analyzed at the Marine Geomechanics Laboratory in University of Rhode Island using a multisensor core logger. This analysis provided profiles of compressional speed, bulk density, and porosity. Compressional speeds obtained from cores which lie on the propagation path (cores 1–3) were used for comparison and validation of inversion. The average compressional speed for the top 1.5 m of sediments is of the order of 1575–1600 m/s.

V. ANALYSIS OF PRIMER SUS DATA

Data from the SUS explosions were collected on two VLAs at the northwest (NW) and northeast (NE) corners of the experimental area during the PRIMER experiment. The positions of these VLAs and the SUS drop locations are shown in Fig. 3. Data were collected at a sampling frequency of 1395.09 samples/second at the NE VLA, whereas the sampling frequency was 3906.25 samples/second at the NW array. The NE VLA consisted of 16 hydrophones in water depths varying from 45.4 to 93 m. The propagation path from the northernmost three SUS explosions to the NE array is over a fairly uniform bathymetry and was chosen for the present study. Longitudinal peak-to-peak variability in the sound speeds in the water column is of the order of 5 m/s and bathymetric variations are of the order of 3 m. Influences of range dependence of bathymetry and ocean sound speed are not significant. Hence range-independent conditions were assumed for the propagation. This assumption was verified by comparing the group speed values calculated assuming range independence and also using adiabatic range dependence (Fig. 6).

Figure 7 shows the time series from the SUS explosion at the NW corner received at the top hydrophone of the NE VLA. The signal is broadband in nature and dispersion effects are evident in the time series, e.g., separation of the

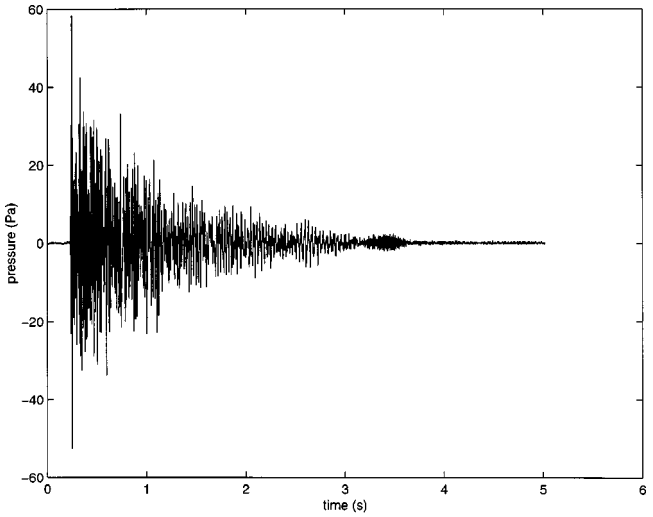


FIG. 7. Time series of SUS signal received at 40 km on one of the hydrophones. The arrival times are arbitrary. Note the arrival of a mode at high frequency at 3.3 s.

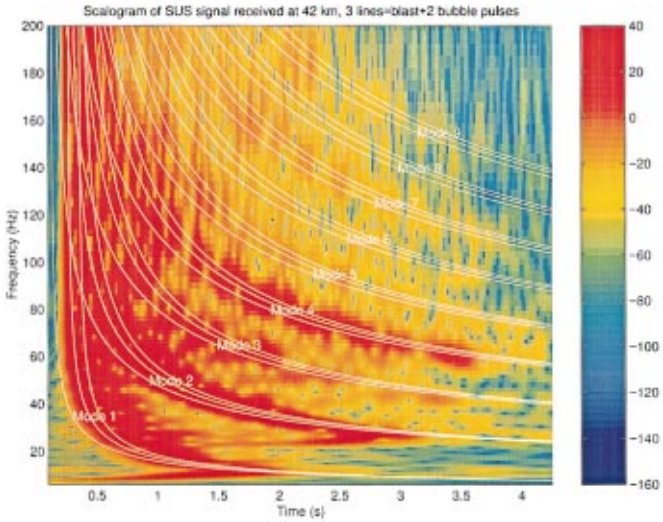


FIG. 8. Time-frequency scalogram of SUS signal shown in Fig. 7 using Morlet wavelet. The color scale is arbitrary and corresponds to intensity in decibels. The three lines correspond to the blast and two bubble pulses.

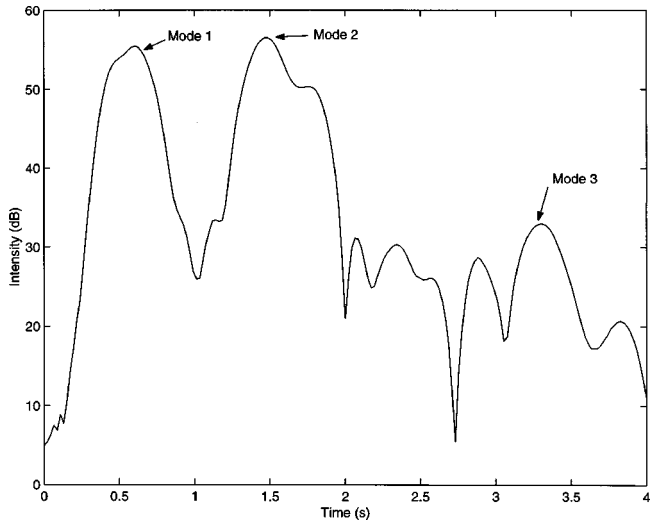


FIG. 9. The acoustic intensity at 40 Hz showing the time of arrivals corresponding to modes 1–3. Mode 3 intensity is very small compared to modes 1 and 2.

times of arrival of various frequencies is apparent. This signal is then further processed using a Morlet wavelet. A standard wavelet package[27] was used for this analysis. Following the shock wave due to the initial blast, a series of positive pressure (''bubble'') pulses are emitted by the pulsating gas globe at the successive instants of minimum volume. The time of arrival between the direct blast and the first bubble pulse is approximately 0.12 s for 1.8 lbs of TNT exploding at 18-m depth.[28] Because of the presence of these bubble pulses, it is difficult to identify and separate early individual mode arrivals. Late arrivals are better separated, and are easily identified as can be seen in Fig. 8. The continuous lines in Fig. 8 represents the theoretical curves for the direct blast and the two bubble pulses. These theoretical lines are generated assuming a simple sound speed model based on the available historical data. These theoretical group speed curves help to identify the individual mode arrivals especially when multiple peaks are present corresponding to a mode arrival. The arrival times are calculated based on Fig. 8 and the arrival pattern for individual frequencies as shown in Fig. 9 for 40 Hz. Another difficulty encountered in obtaining mode arrival times is the poor resolution at higher frequencies which is inherent in our wavelet analysis. Although the wavelet analysis gives better time–frequency resolution than the short time Fourier transform (STFT), the resolution is not uniform across the entire time–frequency plane. Specifically, it does not give good frequency resolution at higher frequencies when compared to low frequencies. Hence it is very difficult to extract group speed values for higher modes at higher frequencies.

The peak intensities are located for representative frequencies and the corresponding arrival times are used for calculating group speeds. The standard errors of the data at various frequencies and modes are estimated from the width of the peaks. The intensity of the acoustic signal at 40 Hz is shown in Fig. 9. The peaks corresponding to modes 1–3 are at 0.6, 1.5, and 3.4 s, respectively. The distance between points corresponding to 95% of peak intensity is taken as an approximate measure of the spread of the peaks.

A. Sensitivity study

In order to quantify the effects of various inverted parameters on the group speed values, a sensitivity analysis was performed, the results of which are tabulated in Table II. Parameters whose effects were studied include water depth, sound speed variations in the water column, and changes in sediment compressional speed in the top 30 m of the sediment. Values shown in the Table II are percentage increases in the value of the fitness function or objective function [Eq. (1)] due to changes in these parameters. The changes were compared with the fitness of the baseline model which corresponds to the best parameter set. Variations in the sound speeds in the water column produce considerable changes in the group speeds. The variations in the ocean sound speed along the propagation path are of the order of 5 m/s (Fig. 10). Hence changes in fitness due to ocean sound speed variations of magnitude 5 and 3 m/s are considered. A uniform increase of 5 m/s in sound speed in the water column changes the fitness by 80% while the change is 92% when it is decreased. The changes are respectively 32% and 42% when the magnitude of the variation in sound speed was decreased to 3 m/s. The effect of the absence of the warm surface layer is simulated by changing the sound speed to 1512.5 m/s from 1502.5 m/s in the top 15m. This produces only 5% change in the fitness. Water depth variations have considerable impact on the group speeds as seen from the changes in the fitness. An increase in water depth by 4 m changes the fitness by 54% while a decrease by the same amount produces 60% change in fitness. Variations in the bottom compressional speed by 50 and 30 m/s in the top 30 m of the sediment produce changes in fitness function whose magnitudes are comparable to those produced by the ocean sound speed variations (5 and 3 m/s) discussed earlier. But the natures of these changes are considerably different. The variations in the sediment speeds tend to affect the low fre-

TABLE II. Results of the sensitivity study. Changes in the fitness are listed as percentage changes from the fitness of the baseline model.

Parameter	Amount	Change in fitness (percentage)
Sound speed in water column	+5 m/s	80
	+3 m/s	32
	−5 m/s	92
	−3 m/s	42
	constant sound speed (1502.5 m/s) in top 15 m	5
Range dependent (adiabatic)		12
Water depth	+4 m	54
	−4 m	60
Sound speed in top 30 m of sediment	−30 m	23
	−50 m	81
	+30 m	42
	+50 m	82

quencies of higher modes to a greater extent whereas changes in water column sound speeds affect the higher frequencies of lower modes (Figs. 11 and 12).

Figure 6 shows the changes in group speeds when range-dependent analysis using adiabatic theory was applied. Sound speed profiles at eight locations along the propagation path were used to calculate the group speeds. These sound speed profiles were calculated using data collected by SeaSoar. Sediment properties were assumed range independent. The eigenvalues and eigenfunctions were assumed to vary linearly within the range sections. Group speed values were calculated at fixed intervals along the propagation path using these eigenvalues and eigenvectors and range averaged to get the mean group speed value. Differences between range-independent and adiabatic calculations are notable for modes 1 and 2 at higher frequencies. For higher modes, changes are minimal. Since higher modes contain more information for sediment inversion, use of range-independent analysis can be justified. This sensitivity analysis helps to arrive at the following conclusions which enabled the proper formulation of the inverse problem:

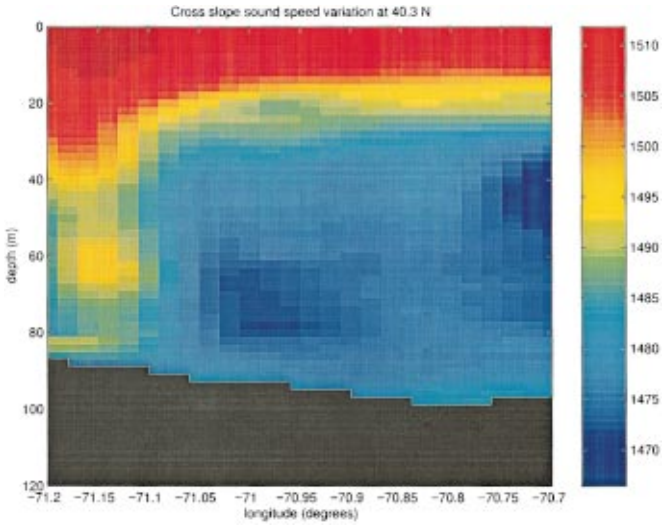


FIG. 10. Variation of sound speed along the propagation path.

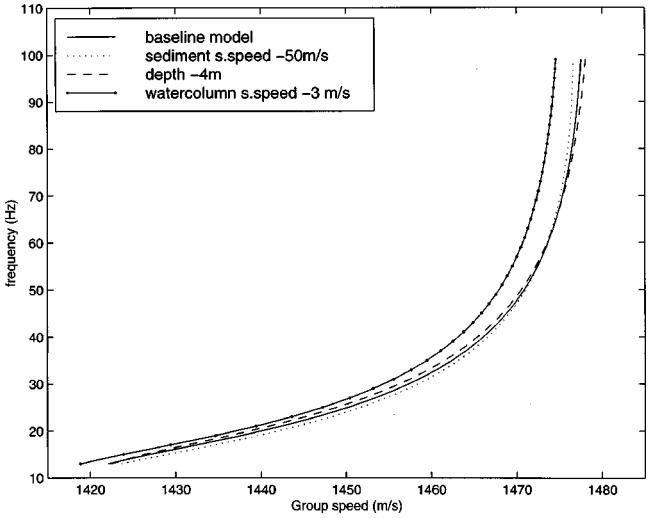


FIG. 11. Effect of changes in water depth, ocean sound speed, and sediment compressional speeds on group speeds corresponding to mode 1.

(i) For the given environmental conditions existing at the present location of study, compressional sound speeds can be estimated with an accuracy of approximately 30 m/s.
(ii) Inclusion of water depth, sound speed variation in the water column, and source–receiver range will improve the quality of inversion.
(iii) Quality of data corresponding to the higher modes at lower frequencies will have greater impact on the accuracy of the sediment compressional speed values.
(iv) Range-independent inversion is appropriate for the present study.

B. Genetic algorithm inversions

Our inversion scheme discussed is based on matched field processing (MFP) concepts which encompass the following components: an environmental model, an acoustic propagation model to predict the group speed values corresponding to the unknown parameter set, an objective function which is minimized, and an efficient algorithm for searching the parameter space. The variability in the water column sound speed is modeled using empirical orthogonal functions (EOFs). The background sound speeds required to generate the EOFs were obtained from the SeaSoar measurements made at the location on the day of the SUS experiment. Figure 13 shows the EOFs used to represent the sound speed variations in the water column. The sediment is modeled as layers with unknown compressional speeds. Shear and attenuation effects are neglected for the inversion. The number of layers was fixed based on the extent of acoustic penetration into the sediment and the total number of unknown parameters that can be handled computationally. Layers are provided thin in the top 30 m of the sediment to take advantage of the acoustic penetration down to those depths. In addition to sound speed, water depth and source–receiver range are also treated as unknowns. Hence the unknown pa-

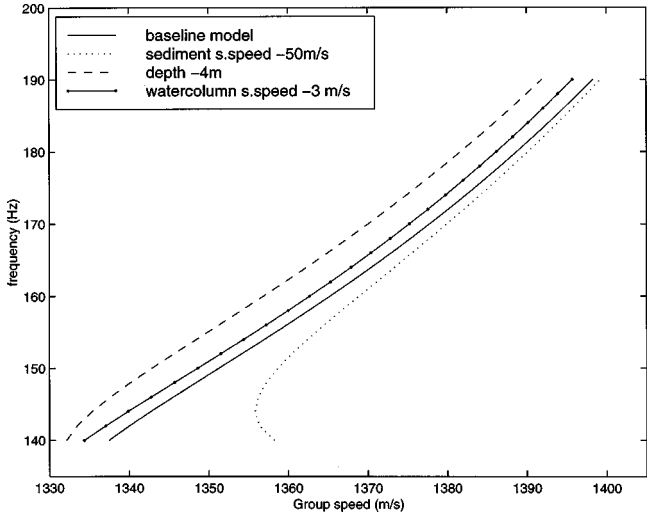


FIG. 12. Effect of changes in water depth, ocean sound speed, and sediment compressional speeds on group speeds corresponding to mode 9.

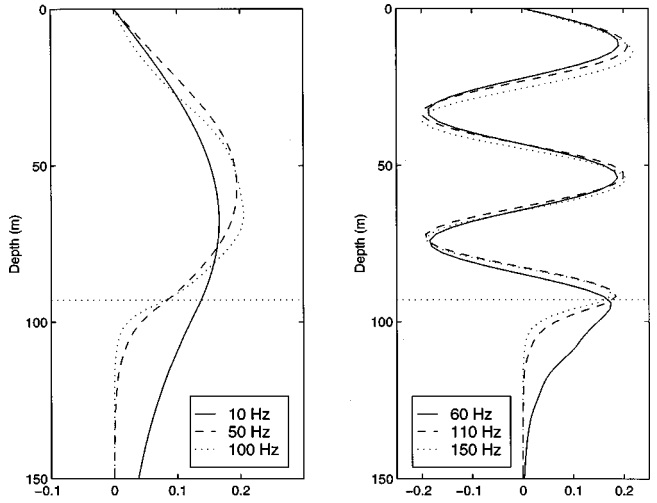


FIG. 14. Mode shapes for modes 1 and 5. The average water depth is about 93 m.

rameter set for the present study consisted of six EOF coefficients in the water column, compressional speeds in 13 sediment layers, water depth, and source–receiver range. Any *a priori* information about the sound speed values can be incorporated into the inversion by fixing the limits of the unknown parameter space. In the present case background information about sound speeds in the water column was available, whereas not much was known about the sediment. This necessitated a large parameter space for sediment compressional speeds. The limits of the model subspace for search varied at different depths in the present analysis. These were 1450 to 1650 m/s at the water–sediment interface, 1600 to 1800 m/s at 10-m depth, and 1650 to 1850 m/s at 30-m depth. It should also be noted that the results of the sensitivity study discussed in Sec. V A and shown in Table II directly influenced the choice of the parameters to be included as unknowns.

The fitness of each member of the population was measured based on the value of the objective function. The objective function for this analysis was based on minimizing the difference between group speed values calculated using the observed time of arrivals and the predicted group speeds [Eq. (1)]. Times of arrival corresponding to frequencies in the range of 20–200 Hz and modes 1–9 were utilized for the inversion. Mode-formed data from all the hydrophones in the VLA were used to generate the time–frequency distributions. A range-independent normal mode routine was used to calculate the group speed values corresponding to the model vectors.

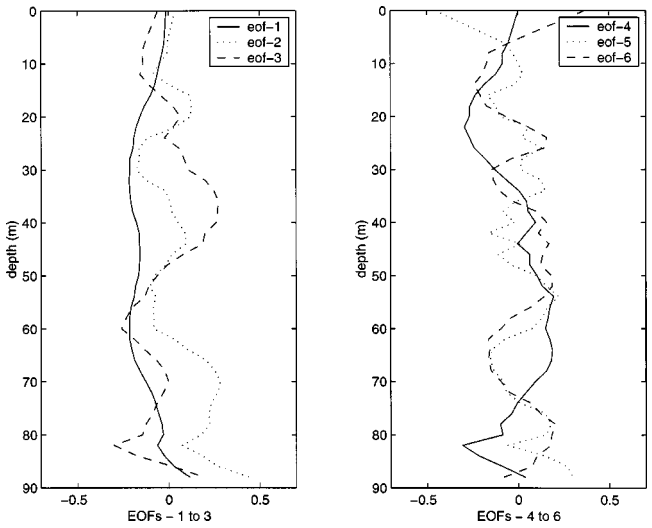


FIG. 13. The EOFs used to represent the sound speed variations in the water column. These were constructed from the background sound speed information obtained using SeaSoar observations made during the experiment.

Inversions were performed using the genetic algorithm Matlab toolbox[29] with a population size of 125 and with 100 generations. The stochastic universal sampling selection algorithm, real mutation and discrete recombination were adopted. Parallel GAs were run to make sure that the solution converged to the same minimum. Also, all the model parameters were stored in each generation and used to calculate the mean and covariance after assigning appropriate probability densities to them. Once the best parameter set was obtained using the GA inversion, further local optimization was done using a Levenberg–Marquardt algorithm and the linear perturbation approach. Both of these methods converged to the local minimum found near the GA inversion estimate. The perturbation inversion also gave estimates of resolution and variance.

VI. RESULTS AND DISCUSSIONS

Mode shapes for some representative frequencies (low, intermediate, and high) are plotted in Figs. 14 and 15 for modes 1, 5, 7, and 9. These figures show the extent of the acoustic penetration into the sediments for these modes at the corresponding frequencies. These figures also help us infer the relative importance of different modes and frequencies for the inversion of sediment speeds, and the depth to which these can be estimated with reasonable accuracy. From the figures it is clear that higher modes at lower frequencies, which penetrate deep into the bottom, are relatively important for the sediment inversion. This inference is in agreement with sensitivity analysis results discussed earlier.

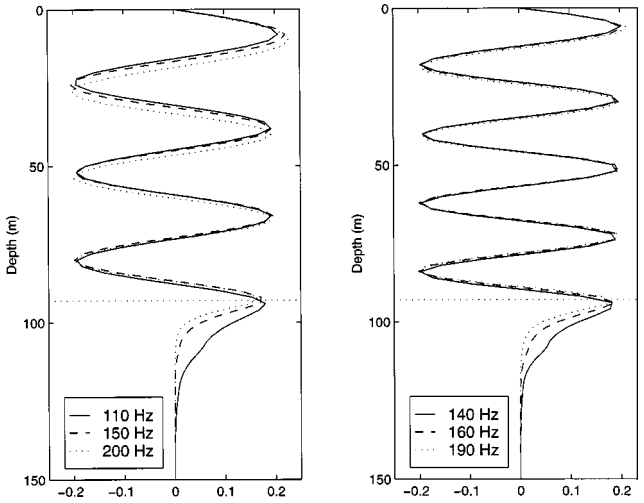


FIG. 15. Mode shapes for modes 7 and 9. The average water depth is about 93 m.

The compressional sound speed profile obtained using GA optimization is shown in Fig. 16 for the top 40 m of the sediment. The three profiles shown in that figure correspond to the sound speeds calculated using the best and the mean GA parameter set and the AMCOR-6012 data. Three parallel GAs were run and the entire population from these runs was utilized to calculate the mean after weighting with appropriate probabilities based on their fitness. Also shown in the figure are the gravity core data which extend down to a depth of 1.4 m only.

The inversion compares reasonably well with the AMCOR sound speeds for the top 20 m. The mean compressional speed in the upper 15 m of the sediment is approximately 1660 m/s, which compares favorably with the sediment compressional speeds reported by McGinnis and Otis[23] and Brocher and Ewing.[22] This average compressional speed corresponds to silty sands or sandy silts,[30] the presence of which in the upper sediment layers has been reported by almost all the investigaters as discussed in Sec. IV A. The

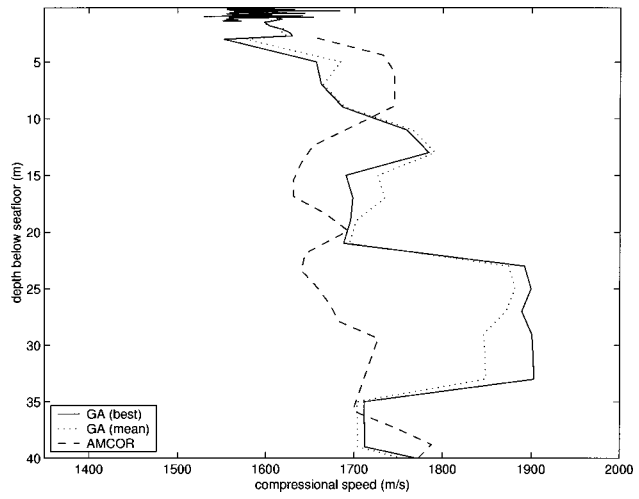


FIG. 16. Sediment compressional speeds obtained by genetic algorithm inversion. The compressional speeds obtained from the gravity cores are also shown in the top 1.4 m of the sediment. Compressional speeds calculated using AMCOR data are also shown. Note the difference between inversion and AMCOR at 3–7-m depth.

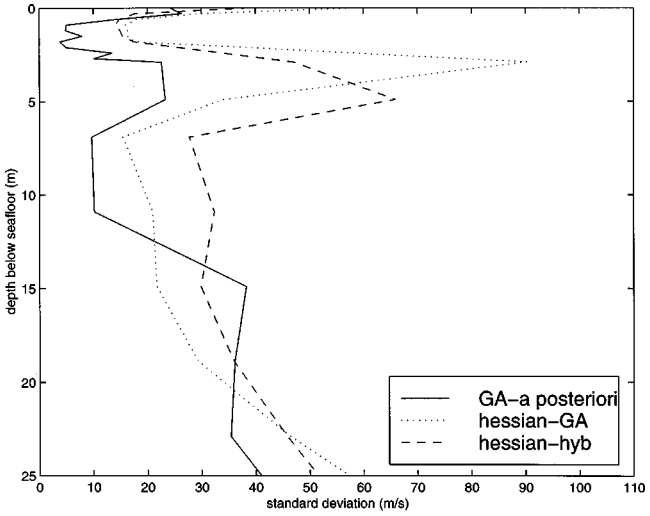


FIG. 17. Standard deviation estimates for inversions. Standard deviations shown are computed using *a posteriori* analysis of GA results and Hessians. Note the reduction in standard deviation due to local optimization in the top 5 m.

high compressional speeds at 10 m may be due to the presence of fine or very fine sands at these depths. Both AMCOR (6010 and 6012) profiles and inversion show this high-speed layer, but at different depths. This layer with compressional speeds of 1750 m/s is 5 m deeper in the case of inversion compared to the AMCOR-6012 profile. This may be due to an increase in thickness of the surface layer in the shelf region due to the deposition of sediments. It may be noted that in the case of AMCOR-6010 this high-speed layer is still deeper compared to AMCOR-6012. The compressional speed profile corresponding to the best GA parameters follow the inversion corresponding to the mean GA parameters very closely, indicating higher convergence except at depths between 5–7 m and 20–30 m. It can also be seen from Fig. 17 that the standard deviation is comparatively higher at these depths. At greater depths the fitness function becomes less sensitive to sound speed variations in the sediment layers, leading to a diverse population with slightly differing fitnesses resulting in higher standard errors and a larger difference in mean and best profiles. Penetration of acoustic energy is very low below 25 m into the sediment.

Figure 18 shows the comparison of inversion with gravity core data. They agree very well especially below 0.6 m. The differences between gravity core data and inversion are generally within 25 to 30 m/s. This is reasonable since the core logger has an accuracy of 5 to 10 m/s and standard deviations of the inversion are of the order of 15 m/s (Fig. 19).

Figure 20 shows the comparison of the group speeds calculated based on the inversions and experimentally observed values. The group speeds corresponding to the AMCOR data are also shown in the same figure. The inversion and AMCOR data differ mainly for the higher modes at low frequencies. At these locations the inversion matches the experimental data better than the AMCOR data. Interestingly for modes 1 and 2, at higher frequencies, both AMCOR and inversion differ from the experimental data. This is more likely due to the errors in sound speeds in the water column

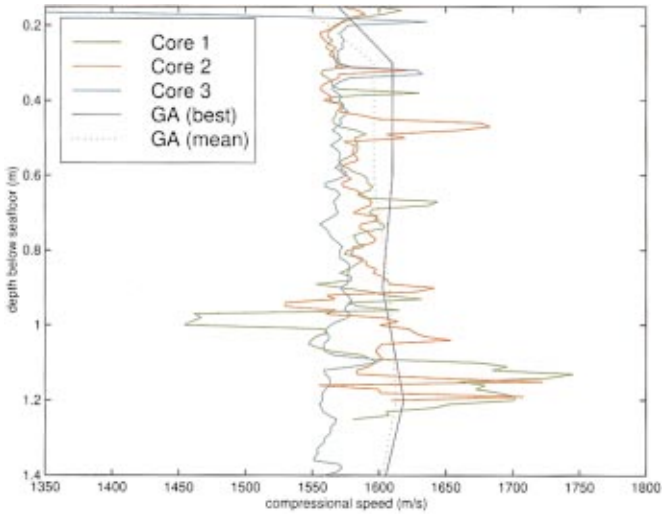


FIG. 18. Sediment compressional speeds for the top 1.4 m of the sediments estimated by genetic algorithm inversion. The compressional speeds obtained from the gravity cores 1–3 are also shown.

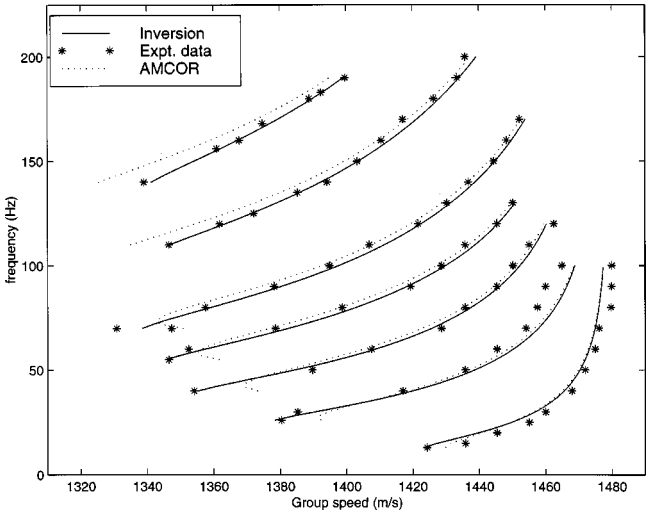


FIG. 20. Group speed dispersion curves calculated for sound speed profiles corresponding to inversion and AMCOR data. The experimental data are the group speeds estimated from mode arrival times.

and water depth than to errors in sediment compressional speeds.

Table III shows the EOF coefficients obtained using the inversion together with the specified search bounds. Figure 21 shows the ocean sound speed profile obtained using these EOF coefficients. Also shown in the figure are the sound speed profiles at 5-km intervals along the propagation path. Profiles with low values at 40-m depth correspond to the eastern side whereas those with high values at 60 m correspond to the western side. This cross-slope sound speed variation is shown in Fig. 10 also. Apart from the top 4 m of the water column, the inversion, which is a range-averaged profile, seems satisfactory. In the top 4 m, the inversion shows a warm layer with increased sound speed. As seen previously in the sensitivity study the presence (or absence) of the warm layer does not affect the group speed values significantly. The error in the top 4 m might have been produced by this insensitivity. The value for the water depth was also obtained by the inversion as 92 m (Table III). This seems to be the average water depth along the propagation path from the SUS shot location to the VLA.

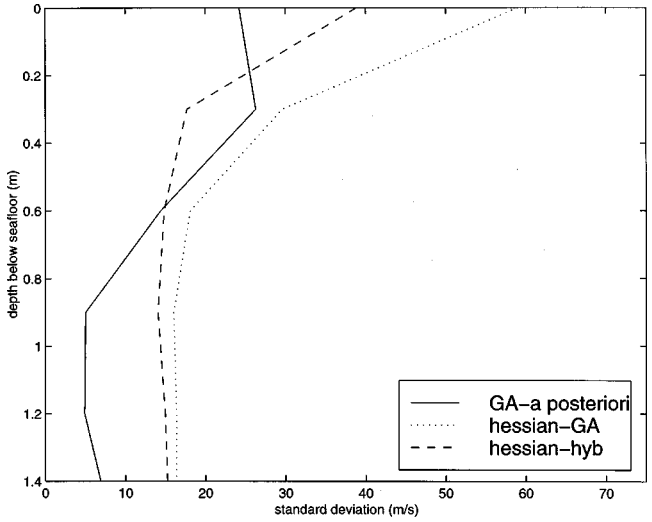


FIG. 19. Standard deviation estimates for inversions for the top 1.4 m of the sediments.

Figure 17 shows the standard deviations computed from the entire population samples for the three GA runs. Also shown in the same figure are the standard errors calculated using the Hessian matrix. Hessians were calculated at the location of the best parameter values by calculating the curvature of the objective function numerically. These two representations of the errors show disagreement mainly at depths 2–7 m. The higher values of standard deviations (lower values of Hessians) at these depths indicate that the GA was not able to converge to a local minimum. This was verified using local optimization methods and linear perturbation methods. The application of these methods may, in addition to this verification, lead to further improvement in the solution. The results of the hybrid inversions consisting of the application of the Levenberg–Marquardt scheme and perturbation methods are discussed in the following sections. The magnitudes of the standard errors are approximately 30 m/s within the top 30 m of the sediment, which can be considered reasonable. Standard deviations in the top 1.4 m of the sediment are shown in Fig. 19. The average standard deviation is approximately 15 m/s in the top 1.4 m.

A. Hybrid inversion

A local optimization technique, which uses Hessians and/or Jacobians to guide the search from the starting point to the nearest local minimum, was employed to ''fine tune''

TABLE III. Search bounds and inversion results for EOF coefficients and water depth.

Parameter	Lower bound	Upper bound	Inversion result
EOF 1	−80	80	7.66
EOF 2	−20	20	1.25
EOF 3	−20	20	−9.85
EOF 4	−20	20	13.27
EOF 5	−20	20	−15.1
EOF 6	−20	20	14.8
Water depth (m)	88	94	92

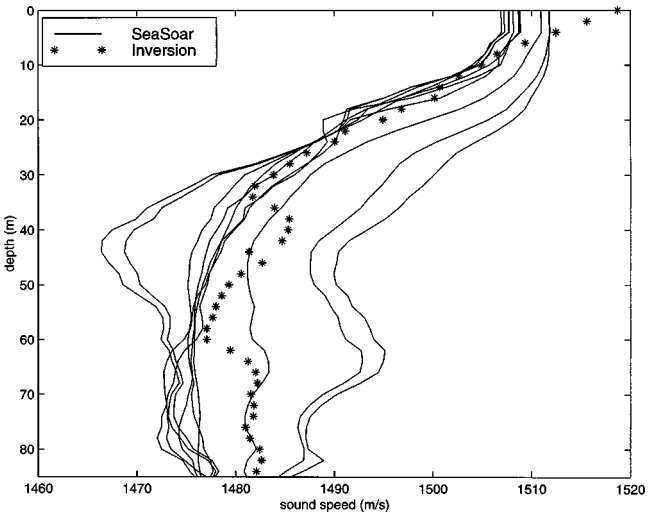


FIG. 21. Ocean sound speed profiles along the propagation path of present study. These were computed using the SeaSoar data. The sound speeds computed using the inversion are also shown.

the GA inversion results. Results of the Levenberg–Marquardt inversion, using the global optimum parameters obtained from the GA inversion as starting points, are shown in Fig. 22. This hybrid inversion improves the inversion at depths 3–7 m and below 20 m. This was expected as the standard deviations calculated using Hessians were comparatively large at these depths. This hybrid inversion is able to decrease the standard deviations at these depths as shown in Fig. 17. It can be noted that the velocity gradients at 5-m depth for hybrid inversion compares well with the AMCOR data. The results of the perturbation inversion using GA results as the background are shown in Fig. 23. The result of the hybrid inversion is also shown in this figure for comparison as both linear perturbation approach and the Levenberg–Marquardt scheme are used to ''fine tune'' the GA results. The changes in the compressional speeds predicted by both these methods show a similar trend. Both these methods predict similar compressional speeds between 3–7-m depths

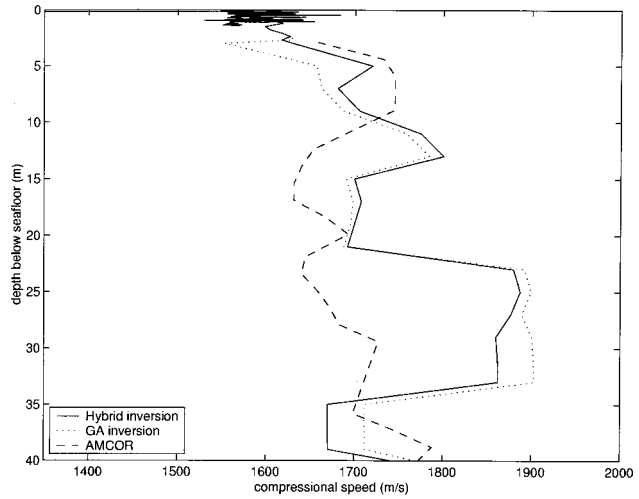


FIG. 22. Sediment compressional speeds obtained using Hybrid inversion. The Levenberg–Marquadt scheme was applied to the best parameter set produced by GA. Note the agreement between the Hybrid inversion and AMCOR data between depths 3 to 7 m. The compressional speed obtained from the gravity cores are also shown in the top 1.4 m of the sediment.

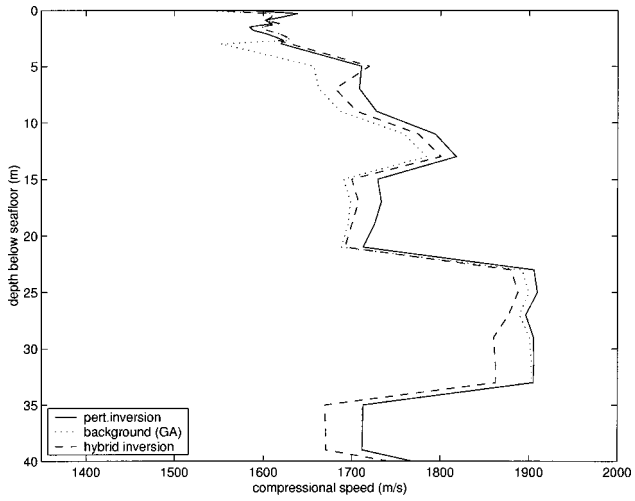


FIG. 23. Sediment compressional sound speeds obtained by linear perturbation inversion. The compressional speeds obtained by GA were used as the background profile for this inversion.

where the GA inversion is most uncertain. We can extract the standard deviation and resolution lengths from perturbation theory as shown in Figs. 24 and 25. The resolution length varies approximately from 5 m/s at the top to 30 m/s at 30-m depth. The standard error is low at 6 m/s. The low values obtained may be due to the good quality of the background profile used for inversion. It should be noted that the linear perturbation inversion was carried out using weighted data as discussed in Sec. II A. The detailed variations in the values of standard deviation depthwise are likely to be caused by different modes turning at different depths.

VII. CONCLUSIONS

Sediment compressional speeds were evaluated using hybrid optimization schemes based on the group speed dispersion properties of the shallow water waveguide. Data for the inversion were obtained from the signals produced by explosive sources. Time-frequency analysis was done using wavelets which provided better frequency-time resolution than usual Fourier-based methods. Results of the inversion provided compressional speeds of sediments down to 30 m

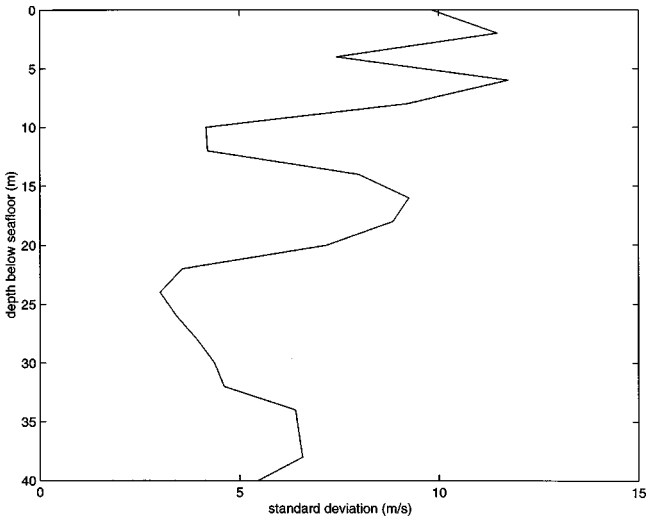


FIG. 24. Standard deviation estimates for linear perturbation inversion.

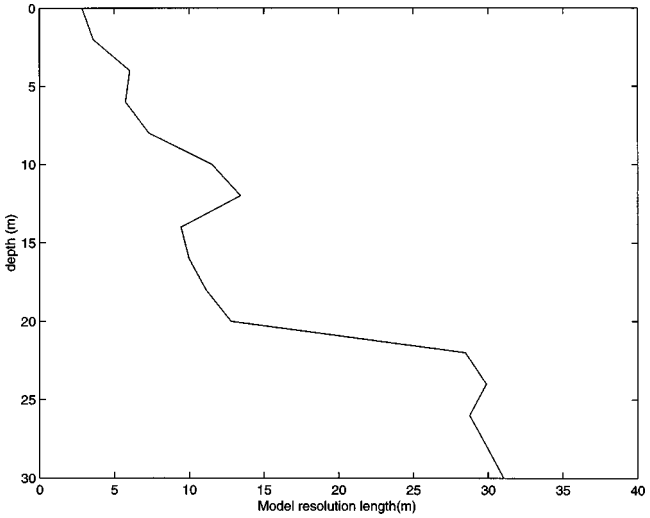


FIG. 25. Resolution length for linear perturbation inversion.

in depth. The inversion compared well to gravity core data taken at the location on the track and also to AMCOR drill site data in the top 20 m. The calculated standard deviations for these depths were of the order of 20 m/s. The results of the sensitivity study quantified the effects of parameters of influence (sediment compressional speeds, sound speed in the water column, water depth, etc.) on the group speed values. Higher modes seem to have greater influence on the inversion of sediment properties than lower modes. This factor emphasizes the importance of obtaining good time-frequency resolution for higher modes. Efforts are currently underway to address this aspect. Also, these inversion results correspond to a range-independent environment. Extension of this scheme to range-dependent environments needs further study.

ACKNOWLEDGMENTS

The authors wish to thank the scientific team of the PRIMER experiment including Robert Beardsley, Ken Brink, Ching-Sang Chiu, Glen Gawarkiewicz, Robert Pickart, Alan Robinson, and Brian Sperry, captain and crew of R/V ENDEAVOR, and pilots and crew from Naval Air Warfare Center at Patuxent River NAS for their significant contributions to data collection. In particular, the authors recognize David Volak and Thomas Gabrielson for their contribution to experimental design. This work was supported by Office of Naval Research, Jeffrey Simmen, Code 321 OA Program Manager, and Steven Ramp, Code 322 PO Program Manager.

[1] A. Tolstoy, O. Diachok, and L. N. Frazer, ''Acoustic tomography via matched field processing,'' J. Acoust. Soc. Am. **89**, 1119–1127 (1991).
[2] P. Ratilal, P. Gerstoft, J. T. Goh, and K. P. Yeo, ''Inversion of pressure data on a vertical array for seafloor geoacoustics properties,'' J. Comput. Acoust. **6**, 269–289 (1998).
[3] K. B. Smith, J. G. Rojas, J. H. Miller, and G. Potty, ''Geoacoustic inversions in shallow water using direct methods and genetic algorithm techniques,'' in Proc. Pacific Ocean Remote Sensing Conf. (PORSEC98), edited by M.-X. He and G. Chen (invited paper) (1998), pp. 703–707.
[4] B. Rapids, T. Nye, and T. Yamamoto, ''Pilot experiment for the acquisition of marine sediment properties via small scale tomography system,'' J. Acoust. Soc. Am. **103**, 212–224 (1998).
[5] J. F. Lynch, S. D. Rajan, and G. V. Frisk, ''A comparison of broad band and narrow band modal inversions for bottom geoacoustic properties at a site near Corpus Christi, Texas,'' J. Acoust. Soc. Am. **89**, 648–665 (1991).
[6] C. J. Pike, ''Analysis of high resolution marine seismic data using wavelet transform,'' in *Wavelets in Geophysics*, edited by E. F. Georgiou and P. Kumar (Academic, New York, 1994), pp. 1–43.
[7] M. D. Collins and W. A. Kuperman, ''Focalization: Environmental focusing and source localization,'' J. Acoust. Soc. Am. **90**, 1410–1422 (1991).
[8] J.-P. Hermand and P. Gerstoft, ''Inversion of Broadband Multitone Acoustic Data from the YELLOW SHARK Summer Experiments,'' IEEE J. Ocean Eng. **21**(4), 324–346 (1996).
[9] P. Gerstoft, ''Inversion of seismoacoustic data using genetic algorithms and a posteriori probability distributions,'' J. Acoust. Soc. Am. **95**, 770–781 (1994).
[10] K. S. Tang, K. F. Man, S. Kwong, and Q. He, ''Genetic Algorithms and Their Applications,'' IEEE Signal Process. Mag. **13**(6), 22–37 (1996).
[11] P. Gerstoft, ''Inversion of acoustic data using a combination of Genetic Algorithms and the Gauss-Newton approach,'' J. Acoust. Soc. Am. **97**, 2181–2190 (1995).
[12] M. I. Taroudakis and M. G. Markaki, ''On the use of matched-field processing and hybrid algorithms for vertical slice tomography,'' J. Acoust. Soc. Am. **102**, 885–895 (1997).
[13] R. L. Parker, *Geophysical Inverse Theory* (Princeton U.P., Princeton, NJ, 1994).
[14] M. K. Sen and P. L. Stoffa, ''Rapid sampling of model space using genetic algorithms: examples from seismic waveform inversion,'' Geophys. J. Int. **108**, 281–292 (1992).
[15] P. Gerstoft and D. F. Gingras, ''Parameter estimation using multifrequency range-dependent acoustic data in shallow water,'' J. Acoust. Soc. Am. **99**, 2839–2850 (1996).
[16] K. Koper and M. Wysession, ''Modelling the Earth's Core and Lowermost Mantle with a Genetic Algorithm,'' http://levee.wustl.edu/seismology/koper/Papers/JGR96/pkp-ga.html, 1996.
[17] A. Newhall, K. Von der Heydt, B. Sperry, G. Gawarkiewicz, and J. Lynch, ''Preliminary Acoustic and Oceanographic Observations from the Winter Primer Experiment,'' Woods Hole Oceanog. Inst. Tech. Rep., WHOI-98-19 (1998).
[18] H. J. Knebel and E. Spiker, ''Thickness and age of Surficial Sand Sheet, Baltimore Canyon Trough Area,'' Bull. Am. Assoc. Pet. Geol. **61**, 861–871 (1977).
[19] J. M. Robb, J. C. Hampson, Jr., and D. C. Twichell, ''Geomorphology and Sediment Stability of a Segment of the U.S. Continental Slope off New Jersey,'' Science **211**, 935–937 (1981).
[20] C. W. Poag, ''Stratigraphy and Depositional Environments of Baltimore Canyon Trough,'' Bull. Am. Assoc. Pet. Geol. **63**, 1452–1466 (1979).
[21] J. C. Hathaway, C. W. Poag, P. C. Valentine, R. E. Miller, D. M. Schultz, F. T. Manheim, F. A. Kohout, M. H. Bothner, and D. A. Sangrey, ''U.S. Geological Survey Core Drilling on the Atlantic Shelf,'' Science **206**, 515–527 (1979).
[22] T. M. Brocher and J. I. Ewing, ''A comparison of high-resolution seismic methods for determining seabed velocities in shallow water,'' J. Acoust. Soc. Am. **79**, 286–298 (1986).
[23] L. D. McGinnis and R. M. Ottis, ''Compressional velocities from multichannel refraction arrivals on George's Bank-northwest Atlantic Ocean,'' Geophysics **44**, 1022–1032 (1979).
[24] A. F. Richards, ''Atlantic Margin Coring Project 1976-Preliminary report on shipboard and some laboratory geotechnical data,'' U.S. Geol. Surv. Open File Rep., No. 78-123 (1977).
[25] M. Badiey, A. H-D. Cheng, and Y. Mu, ''From geology to geoacoustics-Evaluation of Biot-Stoll sound speed and attenuation for shallow water acoustics,'' J. Acoust. Soc. Am. **95**, 309–320 (1994).
[26] M. V. Trevorrow and T. Yamamoto, ''Summary of marine sedimentary shear modulus and acoustic speed profile results using gravity wave inversion technique,'' J. Acoust. Soc. Am. **90**, 441–456 (1991).
[27] F. Auger, P. Flandrin, O. Lemoine, and P. Gonsalves, ''Time-Frequency Toolbox for Matlab,'' http://www.physique.ens-lyon.fr/ts/tftb.html/, 1997.
[28] R. J. Urick, *Principles of Underwater Sound for Engineers* (McGraw-Hill, New York, 1983).
[29] H. Pohlheim, ''Genetic and Evolutionary Algorithm for use with MATLAB-version 1.83,'' http://www.systemtechnik.tu-ilmenau.de/pohlheim/GA-Toolbox/, 1996.
[30] E. L. Hamilton, ''Geoacoustic modeling of the sea floor,'' J. Acoust. Soc. Am. **68**, 1313–1340 (1980).

In situ estimation of sediment sound speed and critical angle

Alain Maguer[a] and Edoardo Bovio
NATO SACLANT Undersea Research Centre, Viale San Bartolomeo 400, 19138 La Spezia, Italy

Warren L. J. Fox
Applied Physics Laboratory, University of Washington, Seattle, Washington 98105

Henrik Schmidt
Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 14 June 1999; accepted for publication 1 March 2000)

Understanding the basic physics of sound penetration into ocean sediments is essential for the design of sonar systems that can detect, localize, classify, and identify buried objects. In this regard the sound speed of the sediment is a crucial parameter as the ratio of sound speed at the water-sediment interface determines the critical angle. Sediment sound speed is typically measured from core samples using high frequency (100's of kHz) pulsed travel time measurements. Earlier experimental work on subcritical penetration into sandy sediments has suggested that the effective sound speed in the 2–20 kHz range is significantly lower than the core measurement results. Simulations using Biot theory for propagation in porous media confirmed that sandy sediments may be highly dispersive in the range 1–100 kHz for the type of sand in which the experiments were performed. Here it is shown that a direct and robust estimate of the critical angle, and therefore the sediment sound speed, at the lower frequencies can be achieved by analyzing the grazing angle dependence of the phase delays observed on a buried array. A parametric source with secondary frequencies in the 2–16 kHz range was directed toward a sandy bottom similar to the one investigated in the earlier study. An array of 14 hydrophones was used to measure penetrated field. The critical angle was estimated by analyzing the variations of signal arrival times versus frequency, burial depth, and grazing angle. Matching the results with classical transmission theory yielded a sound speed estimate in the sand of 1626 m/s in the frequency range 2–5 kHz, again significantly lower the 1720 m/s estimated from the cores at 200 kHz. However, as described here, this dispersion is consistent with the predictions of the Biot theory for this type of sand.
[S0001-4966(00)03206-9]

PACS numbers: 43.30.Bp, 43.30.Gv [DLB]

INTRODUCTION

In certain types of ocean sediments, completely buried mines are considered undetectable by conventional high-frequency mine hunting sonars. This is due to the very low level of energy transmitted into the sediment for these frequencies at the low grazing angles used for mine hunting operations. As the attenuation effect of the sediment is less at lower frequencies, some recent work at SACLANTCEN has investigated the possibility of using much lower frequency sonars, at low grazing angles, in order to enhance the ability to hunt buried mines.

The authors have previously reported an investigation of the mechanisms by which ''anomalously'' high levels of energy are transmitted into the sediment[1] in the 2–16 kHz regime. The conclusion was that classical sound propagation theory can explain the received levels for frequencies below 5–7 kHz, and that scattering due to seabed roughness is the dominant mechanism above these frequencies. However, it was found that to match model predictions with the penetration data it was necessary to assume a sound speed which was significantly lower than the one measured from cores at higher frequencies.

Since the sediment sound speed determines the critical angle, this dispersion is critical to the performance of sonars for buried mine detection. Moreover, the estimation of the sound speed in the sediment is essential for new buried mine classification concepts based on physical scattering models. Thus a sound speed mismatch may obviously lead to false classification by such advanced systems. Finally, since the dispersion is particularly strong in the midfrequency regime of relevance to buried mine sonars, the results presented in the previous study and here suggest very strongly that the sound speed be estimated *in situ* at the relevant sonar frequencies, preferably through the sonar itself, rather than indirectly through core measurements.

The frequency variation of sediment sound speed and critical angle is investigated through the analysis of *in situ* acoustic measurements, and compared to core measurements and theory. As will be demonstrated, simulation using Biot theory,[2] taking into account sediment porosity and permeability, suggests significant sediment sound speed dispersion in the range 1–100 kHz for the type of sand found near Elba Island. On the other hand, the study shows that for frequencies in the 2–20 kHz regime, the *in situ* critical angle estimation, together with classical transmission theory, is ad-

[a]Currently with Thomson Marconi Sonars PTY., 274 Victoria Road, Rydalmere, NSW 2116, Australia.

equately describing the sediment penetration and therefore eliminating the need for estimating the 13 sediment parameters necessary for describing the sediment using Biot theory.

Section I describes the experimental setup and the equipment, including the characteristics of the parametric sonar used as the transmitting source, the rail facility specially designed for this kind of experiment[3] and the geometry of the buried hydrophone array. Also, the results of the sediment core analysis are presented, followed by estimates of the associated frequency dispersion as predicted by Biot theory. Section II describes the experimental results and the analysis. The temporal characteristics of the buried hydrophone data are first discussed. Then the results of the penetration analysis are presented. The analysis confirms the earlier result,[1] that the subcritical penetration is dominated by the evanescent field below 5–7 kHz, while scattering due to surface roughness is the dominant mechanism at higher frequencies. The low sediment speed required to theoretically explain the penetration data is then confirmed by using array delay processing to achieve an independent *in situ* estimate. Finally, in Sec. IV the OASES wave number integration code is used to directly model the time response of the hydrophone array. It is shown that the predictions using the *in situ* sound speed estimate in combination with classical acoustic theory is close to identical to the results achieved using a full Biot theory representation of the sediment. Also, it is demonstrated that the new OASES-3D model of three-dimensional seabed scattering provides excellent prediction of the contribution above 5 kHz of the rough seabed scattering.

I. EXPERIMENTAL SETUP

A. Equipment and geometry

The parametric sonar used for this experiment was the SIMRAD TOPAS (TOpographic PArametric Sonar). Its primary frequency range is 35–45 kHz, and it generates secondary frequencies in the range 2–16 kHz.[1] The TOPAS transducer consists of 24 staves, electronically controlled to form a beam in a selected direction. A short single pulse is

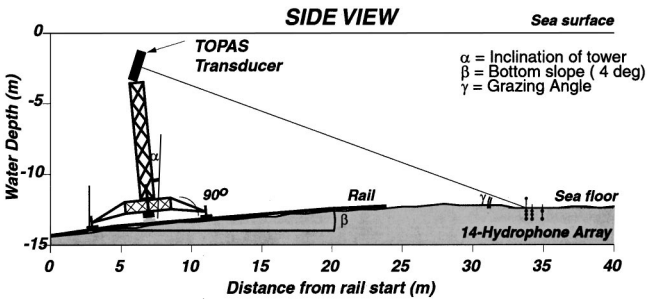


FIG. 1. Experimental setup for penetration measurements in Marciana Marina, Elba Island. A parametric source is mounted on a 10-m-tall tower which can be repositioned on a rail on the seabed for varying the incident angles.

obtained by transmitting a weighted HF-burst at the primary frequency. The transmitting source level is approximately 238 dB for the primary frequency, while the source levels obtained at different frequencies vary from about 190 to 207 dB *re*: 1 μPa at 1 m.

The experimental configuration for the penetration measurements is shown in Fig. 1. To allow insonification of the bottom at a wide spectrum of sub- and supercritical incident angles, the transmitter was mounted on a 10-m tower which could be accurately positioned along a 24-m linear rail on the bottom. In addition, the TOPAS transmitter was mounted in a pan-and-tilt assembly with a motion reference unit (MRU) so that the transmission direction could be accurately controlled and measured.

As indicated in Fig. 1, the seabed, and subsequently the rail, was not horizontal. The slope, which is critical in the sense that it will change the grazing angles, was measured at the end of the cruise by divers who measured the water depth along the rail, as well as over the buried hydrophone field every 10 cm. A detailed description of the experimental setup may be found in Refs. 1 and 3.

In order to validate the results obtained in Ref. 1 on subcritical penetration into the sediment, an array of 14 buried hydrophones was deployed in the seabed. The hydrophones were mounted on five vertical poles as shown in Fig. 2. The burial depths of the hydrophones varied from 5 to 52

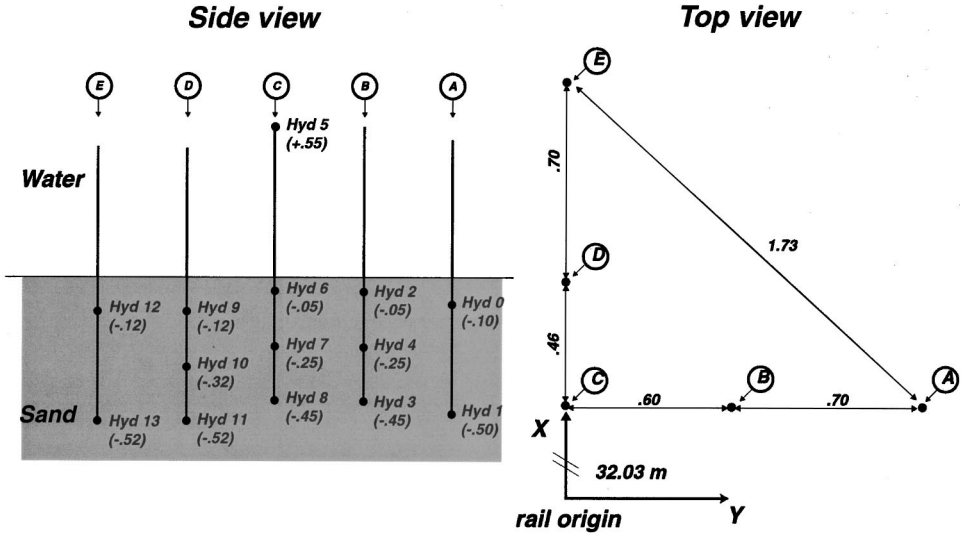


FIG. 2. 14-hydrophone Reson TC4034 array geometry. The array is deployed on five vertical poles A, B, C, D and E. The *x*-axis coincides with the seabed rail, with the central pole being deployed at a distance of 32.02 m from the far end of the rail, as indicated in Fig. 1. The depth of the buried hydrophones in cm is indicated.

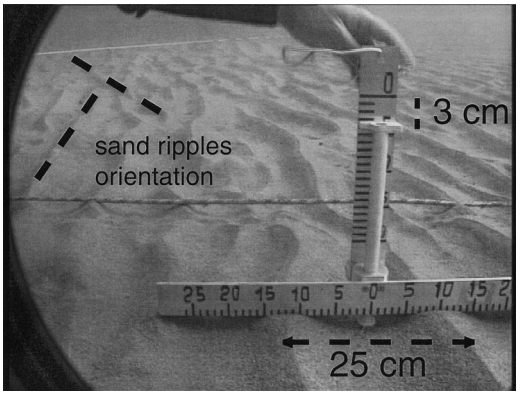


FIG. 3. Bottom roughness measured by divers with underwater digital video camera. The rms height of the ripples is measured on the order of 1.5 cm. The wavelength of the ripples is measured on the order of 25 cm. The directions of the sand ripples are not uniform everywhere, as indicated.

cm. The central pole had a hydrophone mounted above the seabed for measuring the incident field.

For all positions of the source tower, the range to the hydrophone array was larger than the nonlinear interaction length for the parametric sonar, such that nonlinear penetration effects can be ignored in the analysis.[4] For each selected position of the source tower, the pan and tilt of the sonar was remotely adjusted to yield maximum signal at the uppermost buried hydrophone on the central pole (hydrophone number 6 in Fig. 2). A series of 100 pings were transmitted in each configuration, allowing for robustly averaged penetration measurements.

B. Bottom properties

The experiment was performed on a sandy bottom in 13–14-m water depth near Marciana Marina on the north side of Elba Island, Italy. The experiment site was selected because a sub-bottom profiling (''boomer'') survey indicated the presence of a 9-m-thick layer of homogeneous sediment. The seabed was characterized by a distinct ripple pattern with root-mean-square (rms) roughness 1.5 cm, cross-ripple correlation length 25 cm, and skew angle 40° relative to the source-receiver axis.

These ripple parameters were estimated by divers, but the associated roughness spectral characteristics were estimated by stereo photography. An example of an underwater video image of the seabed is shown in Fig. 3. The digital stereo photogrammetry system uses two spatially separated cameras mounted in a rigid frame to take digital photographs of a 650×450 mm patch of the seabed. A height field is then produced via a stereo-correlation procedure,[5] yielding horizontal and vertical resolutions of 1 and 2 mm, respectively. Figure 4 shows an example of the two digital images and the associated roughness estimates.

The sediment properties were determined through core analysis. Grain size estimates classified the sediment as medium sand, with average density 2.67 g/cm^3 and porosity 45.8%. The average sound speed in the upper 25 cm of sediment was estimated to 1720 m/s by pulsed travel time measurements centered at 200 kHz. CTD measurements during the experiment indicated a well mixed water column with a

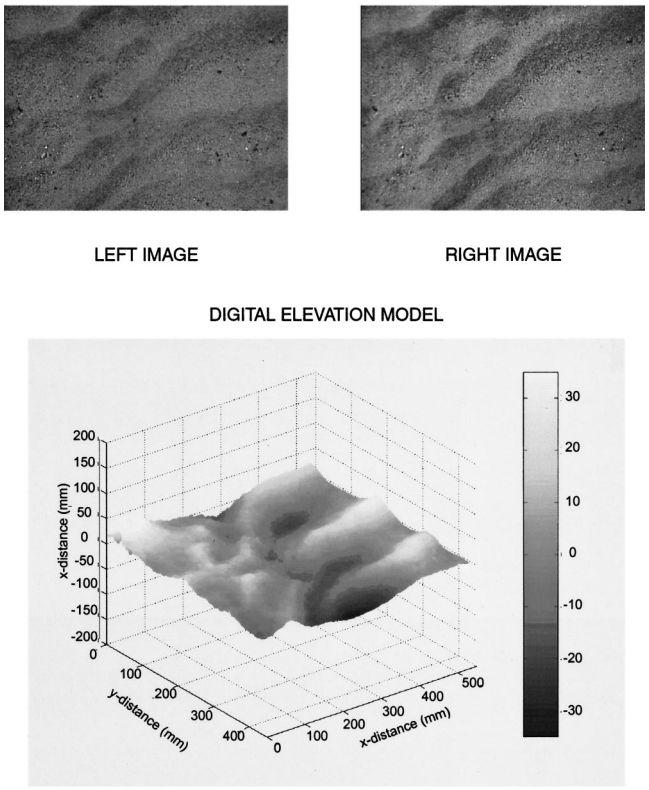


FIG. 4. Digital close-range stereo photogrammetry of the seabed. The left and right images of the bottom are shown on top, while the lower graph shows the estimated roughness elevation, with the gray scale representing the elevation in mm.

sound speed of 1515 m/s. The nominal critical angle corresponding to these sound speed estimates is approximately 28°.

The permeability of the sand was estimated to be 1.7 $\times 10^{-11}$ m^2. The technique used was the standard method for measurement of hydraulic conductivity of saturated porous materials using a flexible wall permeameter. This technique is described in the method D5084-90 from the American Society for Testing and Materials (ASTM). The measurements were performed at the Geology Department of the Engineering University in Rome, Italy.

The frequency variation of the sound speed predicted by Biot's theory using the measured sediment parameters is shown in Fig. 5, suggesting a sediment sound speed at low frequencies which is significantly lower than the 1720 m/s measured from the cores. For example, Fig. 5 suggests a sediment sound speed 1620 to 1660 m/s in the 2–16-kHz frequency bandwidth of the TOPAS sonar. The Biot parameters which were not directly measured, such as the frame bulk modulus, were chosen to be consistent with historical data for sand,[2] but also such that the predicted sound speed at 200 kHz was close (1710 m/s) to 1720 m/s. All the parameters used for the Biot model are given in Table I.

The frequency dispersion for porous sand is most dramatic in the midfrequency regime where the sound speed transitions from its low-frequency to its high-frequency limits. The characteristics of this transition regime are particularly sensitive to the sediment permeability, as discussed by Stoll[2] and illustrated in Fig. 5. The dashed-dotted curve

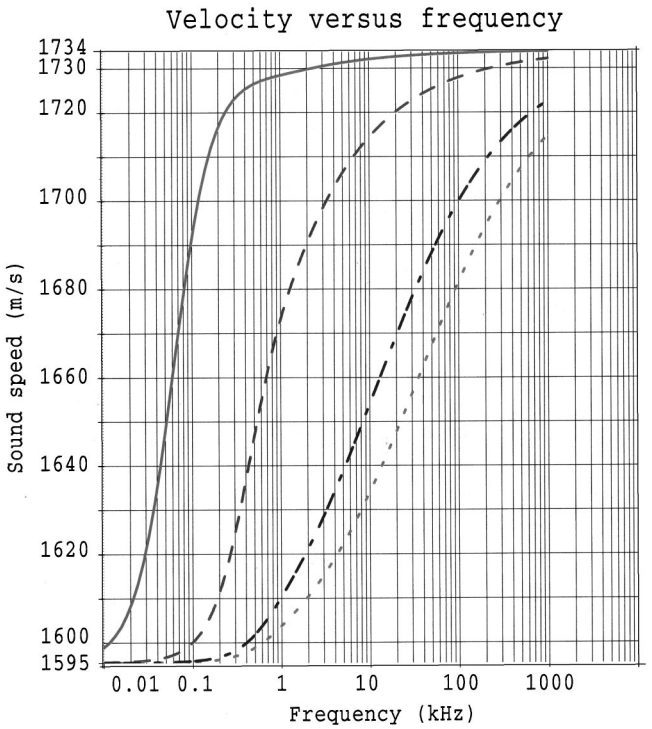


FIG. 5. Frequency dispersion of sound speed in sand as predicted by Biot theory. Four different values of permeability are considered: Solid line: 10^{-9} m^2, dashed line: 10^{-10} m^2, dot-dashed line: 1.7×10^{-11} m^2, dotted line: 10^{-11} m^2.

shows the dispersion predicted using the measured permeability of 1.7×10^{-11} m^2, while the other curves show the predictions for permeabilities of 10^{-9}, 10^{-10}, and 10^{-11} m^2, with all other parameters fixed. The center frequency as well as the width of the transition region is clearly extremely sensitive to the permeability.

It should be stressed that the sediment parameters, both the measured and the historical estimates, are associated with significant uncertainty, and the sound speed dispersion estimates in Fig. 5 should therefore not be directly used quantitatively. On the other hand, they clearly show that within the Biot theory framework, dispersion on the order of 100 m/s is realistic and consistent with the *in situ* critical angle estimates described in the following.

TABLE I. Biot parameters.

Fluid density (kg/m^3)	1000
Bulk modulus of pore fluid (Pa)	2.27E09
Viscosity of pore fluid [kg/(m/s)]	1E-03
Grain density (kg/m^3)	2650
Bulk modulus of grain (Pa)	3.6E10
Porosity of sand	0.43
Permeability of sand (m^2)	1.7E-11
Pore size coefficient (m)	5E-05
Shear modulus of sediment frame (Pa)	2.61E07
Bulk modulus of sediment frame (Pa)	4.36E07
Shear attenuation of sediment frame (dB/lambda)	1.3
Bulk attenuation of sediment frame (dB/lambda)	1.3
Virtual mass parameter	1.25

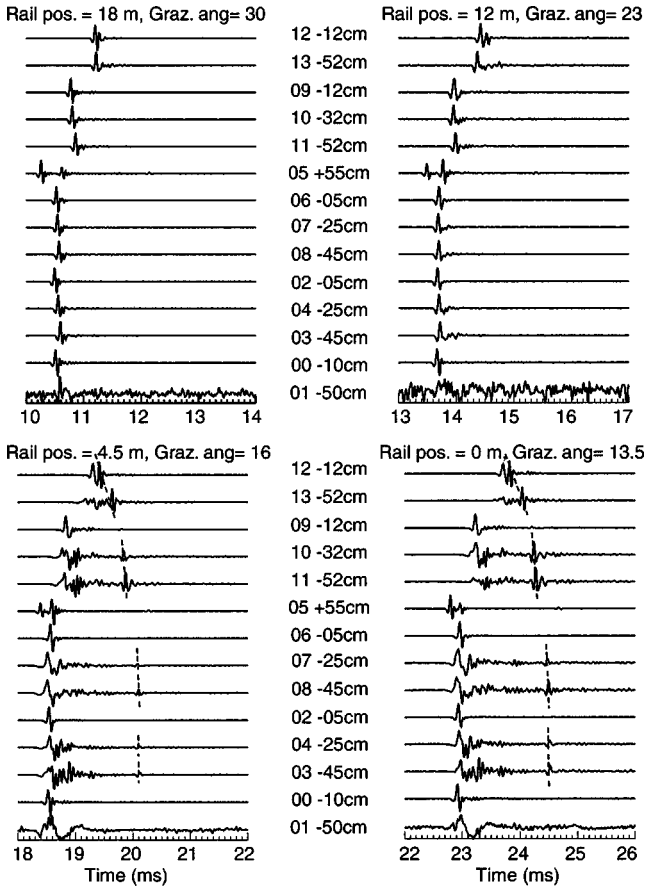


FIG. 6. Examples of signals received on buried hydrophone array for four different incident grazing angles, 30, 23, 16, and 13.5°.

II. RESULTS AND ANALYSIS

A. Hydrophone signals

The signals were recorded on the array for 16 different positions of the TOPAS transmitter on the rail covering grazing angles from 13.5 to 31.5° in steps of one degree. Figure 6 shows examples of the signals received on the 14 hydrophones for four different grazing angles: 30, 23, 16, and 13.5°. The nominal critical angle is 28° if the sediment is assumed to have a sound speed of 1720 m/s as measured from the cores. Consequently only the first angle is nominally supercritical, while the remaining three angles are subcritical. The results in Fig. 6 do not reflect amplitude variations as automatic scaling was applied to each hydrophone. However, the results demonstrate clearly the qualitative differences in the penetration above and below critical, with the subcritical penetration showing a significant decrease in correlation with depth of the hydrophone, suggesting the importance of seabed scattering mechanisms in this angular regime, as discussed in detail in Ref. 1. It should be noted that hydrophone number 1, which was buried at a depth of 50 cm, was not working correctly. Hydrophone number 5 was located 55 cm above the seabed and shows two distinct arrivals, one corresponding to the incident field, and one corresponding to the coherent reflection from the seabed. At all angles the reflected signal was sufficiently separated in time to allow windowing the incident pulse.

It is important to notice that very high signal to noise

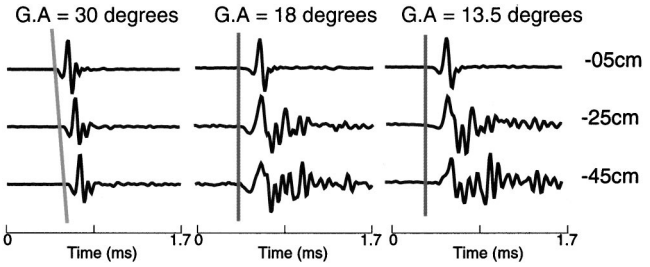


FIG. 7. Signals received on central pole (C) of the buried hydrophone array for incident grazing angles 30, 18, and 13.5°. Arrival move-out indicated by the lines suggests that the highest grazing angle is supercritical, while the two lower grazing angles are subcritical.

ratios were achieved on all the buried hydrophones, even at the subcritical grazing angles. The dotted lines in the two lower frames of Fig. 6 indicate a ''ghost'' arrival which was associated with a strong reflection from the hydrophone cable lying on the bottom behind the array. However, this ''ghost'' arrival is a source of noise which does not interfere with the data we are interested in.

Figure 7 shows the signals received by the three buried hydrophones of the central stick C (hydrophones 6–8), at three incident grazing angles, 30, 18, and 13.5°. Above critical the three hydrophone signals are highly correlated with a well defined move-out characteristic of supercritical transmission. Also, the signals are close replicas of the incident field measured above the interface.

At the other two subcritical angles, the received signals have two distinctly different components. The first part of the signal appears relatively coherent, but with an obvious decrease in high-frequency content for increased burial depth. Also, there is no apparent move-out of this arrival with depth. As discussed later this behavior is consistent with the prediction of classical transmission theory in the evanescent regime.

The second part of these signals appears uncorrelated between receivers, and becomes increasingly dominant with depth, consistent with this signal being associated with a seabed scattering mechanism. The dominance of higher frequencies in this part of the signal as compared to the initial pulse is also consistent with this interpretation.[1]

As emphasized by the lines in Fig. 7, the dominant feature of the signals is the differences in relative arrival time above and below critical angle. Thus at supercritical angles an increase in delay—or move-out—is observed with depth. This is consistent with the classical theoretical prediction of the transmitted field being refracted into the bottom at these angles.[6]

Below critical incidence, the classical theory predicts the transmitted field to be a horizontally propagating *lateral wave* with evanescent amplitude distribution in depth.[6] It will therefore manifest itself as having no move-out with depth, a behavior clearly reflected in the initial coherent component at the two subcritical angles in Fig. 7.

As will be demonstrated in the following, this fundamental difference in the temporal characteristics of the penetrated field is a robust indicator of the effective *critical angle* for the sediment, and may therefore be used for estimating sediment sound speed. It will be demonstrated that the sound speed estimates obtained using this approach are consistent with the effective sound speeds required to theoretically match the frequency- and grazing-angle dependence of the observed amplitude distribution in the sediment.

B. Bottom penetration analysis

To quantify the acoustic penetration into the sediment a measure called the ''penetration ratio'' was devised.[1] For a given frequency, the penetration ratio is defined as the magnitude squared of the ratio of the pressure at a point in the sediment to a reference pressure,

$$PR(f,z)=\left|\frac{P(f,z)}{P_{\mathrm{ref}}(f)}\right|^2=\frac{|P(f,z)|^2}{|P_{\mathrm{ref}}(f)|^2}. \tag{2.1}$$

The reference pressure is defined as the pressure amplitude of the incident field at the seabed, i.e., pressure that would exist at the seabed in the absence of the sediment. Thus the penetration ratio directly corresponds to the pressure produced by a unit amplitude incident wave field. To achieve this, the reference pressure is here chosen as the frequency spectrum of the time-gated reference signal on the hydrophone in the water. Note that the definition allows the penetration ratio to be larger than unity, or positive in terms of dB, because of the effect of the reflected wave.

The conversion of the buried hydrophone signals was performed using the calibration constants determined in water. As shown by Wingham,[7] the change in calibration constants for buried hydrophones is measurable, but insignificant for practical purposes.

As shown above, the Biot theory predicts that the sediment at the experimental site will be strongly dispersive in the 1–10-kHz regime, suggesting that the penetration physics be modeled using a full implementation of the Biot theory. In a later section, the OASES code which implements this formulation will be used to directly model the time series response, but for interpretation of the significance of the frequency dispersion it is extremely useful to analyze the penetration physics within the context of the classical, single-frequency reflection-transmission problem at an interface separating two ideal fluid halfspaces. In particular, the apparent *critical angle* is useful since it may be directly associated with an effective sediment sound speed in the classical sense. The classical critical angle concept, of course, applies rigorously only to an interface between two lossless media, and as shown by Stoll and Kan,[8] it may not be straightforwardly defined for reflection from a porous sediment at the transition frequencies where the equivalent viscoelastic attenuation is large.

Thus for subcritical grazing incidence onto a lossy sediment, viscoelastic theory predicts that the lateral, evanescent wave in the seabed be accompanied by a small, vertically propagating component, accounting for the associated reflection loss. Similarly, above critical, the penetrated wave is an inhomogeneous plane wave with decaying amplitude due to the viscous attenuation. The dispersion predicted for the sediment is closely tied to the attenuation,[9] which could therefore suggest that the classical interpretation be inadequate for strongly dispersive sandy sediments in the transi-

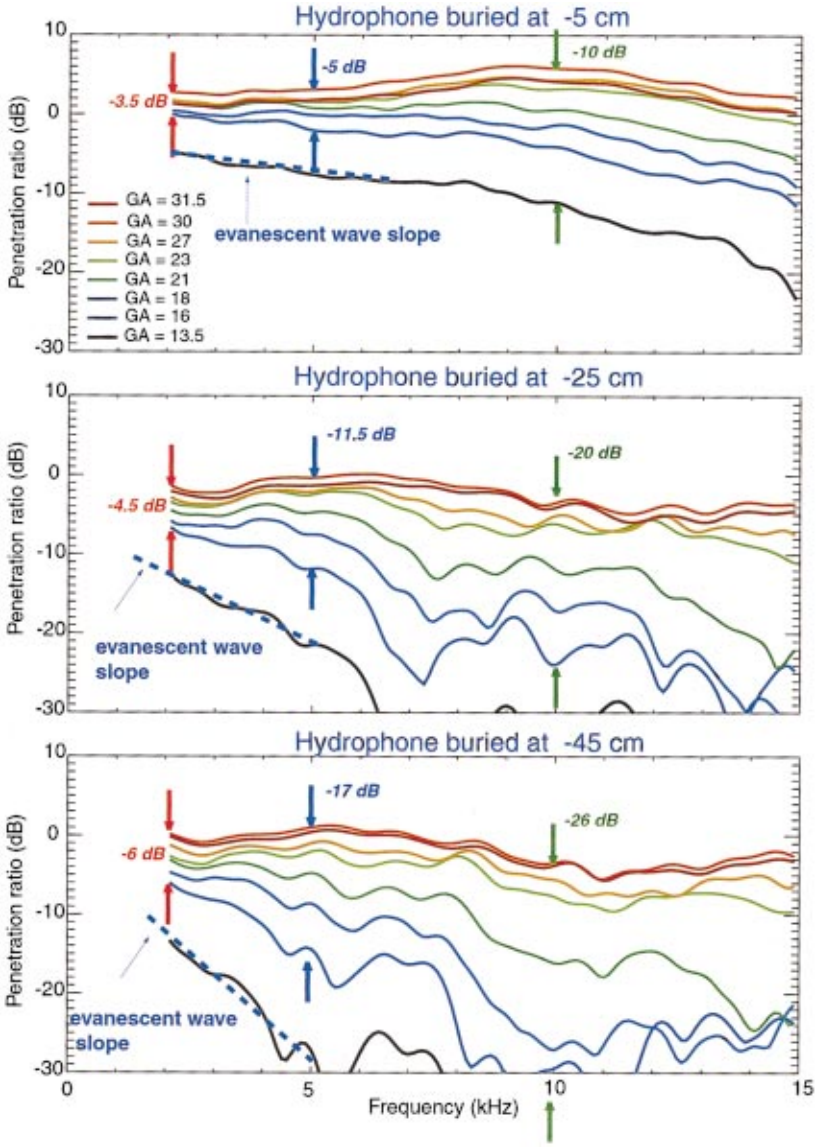


FIG. 8. Penetration ratio in dB versus grazing angles for hydrophones on central pole (C), buried at 5-, 25-, and 45-cm depth. The results for eight different grazing angles in the range 31.5° (red curves) to 13.5° (black curves) are shown.

tion frequency regime. However, as will be demonstrated by the following analysis, even though the attenuation of the porous sand has its maximum at the transition frequencies, the effective attenuation is still small enough for a distinct critical angle in the classical sense to be evident in the penetrated field, in turn providing a robust estimate of effective sediment sound speed.

For an incident plane wave of time dependence $\exp(-j2\pi ft)$ and unit pressure amplitude, the penetrated signal is, according to classical theory,[6,10]

$$P(f,z)=\frac{2m\sin\theta}{m\sin\theta+\sqrt{n^2-\cos^2\theta}}\times\exp[jkx\cos\theta+jk\sqrt{n^2-\cos^2\theta}z]. \quad (2.2)$$

Here, with the water being the reference, $m=\rho_1/\rho$ is the sediment density ratio, $n=c/c_1$ is the normalized index of refraction of the sediment, and θ is the incident grazing angle. Attenuation in the bottom is taken into account by making the index of refraction complex, $n=n_r(1+j\delta)$, where δ is the *loss tangent*, which relates to the *attenuation coefficient* α in dB/λ, as $\alpha=54.58\delta$.[6]

At subcritical incidence, $\cos\theta>n$, the bottom field becomes evanescent, decreasing exponentially (i.e., linearly in dB) with frequency, burial depth and grazing angle,

$$P(f,z)=\frac{2m\sin\theta}{m\sin\theta+j\sqrt{\cos^2\theta-n^2}}\times\exp[jkx\cos\theta-k\sqrt{\cos^2\theta-n^2}z]. \quad (2.3)$$

Figure 8 shows the estimated penetration ratio in dB at the buried hydrophones on the central pole C, for a series of grazing angles between 13.5 and 30°. For all three burial depths the highest grazing angle results show a quasilinear decrease with frequency which is in agreement with Eq. (2.2) assuming a bottom attenuation of $\alpha=0.5$ dB/λ.

At the lowest grazing angles, in particular, the penetration ratio is strongly dependent on frequency. For frequencies up to 5–7 kHz, the penetration ratio (expressed in dB) is decreasing quasilinearly with frequency, with the slope in-

creasing with the burial depth and grazing angle. As indicated by the dashed lines, this behavior is consistent with the subcritical prediction of classical penetration theory. The lines correspond to the result of Eq. (2.3) assuming a sound speed of 1626 m/s in the sediment.

As expected, the results in Fig. 8 have a distinct interference pattern where seabed scattering is the dominant mechanism, i.e., above approximately 5 kHz, and at low grazing angles. However, it is interesting to note that the interference patterns at different grazing angles are clearly correlated, even though the results are obtained following rearrangement of the source tower. In fact, as the grazing angle decreases, the interference pattern shifts down in frequency consistent with the Bragg scattering condition. Thus the Bragg theory is predicting interference minima at a frequency which is depending on the water sound speed c, the roughness correlation length L in the insonification direction, and the incident grazing angle,

$$f \simeq \frac{c}{L \cos \theta}. \tag{2.4}$$

Detailed analysis confirms that the frequency and angle dependency of the interference patterns are consistent with Eq. (2.4). This behavior strongly supports the hypothesis of the seabed roughness being the dominant subcritical penetration mechanism at high frequencies.[11,12]

In summary, the present analysis of the penetration ratio strongly supports the conclusion of earlier work,[1] using a different data set, that the subcritical penetration is dominated by the evanescent field below 5–7 kHz, while scattering due to surface roughness is the dominant mechanism for higher frequencies. Also, these results confirm the earlier finding that the significant dispersion predicted by Biot theory for porous sand must be accounted for to match the data theoretically. In the following section an independent estimate of the sediment sound speed is achieved, using array processing on the buried hydrophone array.

C. Sound speed estimation

To confirm the hypothesis that frequency dispersion of sandy sediments must be taken into account when interpreting bottom penetration data in the 2–20 kHz regime, a more direct measurement of sound speed was devised.

The bandwidth of the source signal allows for time of arrival estimates for the coherent component of the signal on the order of 0.05 ms, which, however, is only slightly less than the 0.1-ms order of the travel time between pairs of hydrophones in the array. This, together with the uncertainty of the array element positions of order cm, makes it impossible to directly use correlation analysis to estimate the sediment sound speed. However, as it will be demonstrated in the following, the fundamental grazing angle dependence of the arrival time structure in depth can be used to robustly estimate the effective critical angle, which can then in turn be translated into an effective sound speed for the sediment. This *in situ* estimate clearly has the advantage of being representative for the bottom at the actual sonar frequencies, which is particularly important for highly dispersive sediments such as porous sand.

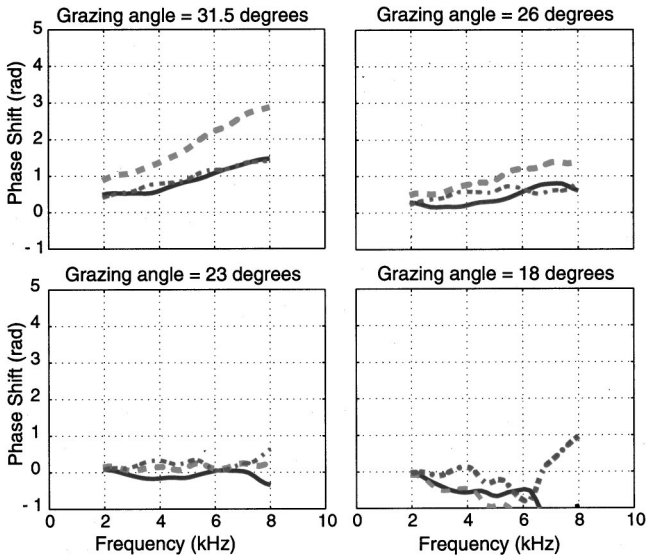


FIG. 9. Phase shift in radians *versus* frequency measured between two hydrophones on pole C. Solid line: Hydrophones at 5- and 25-cm depth. Dashed-dotted line: Hydrophones at 25- and 45-cm depth. Dashed line: Hydrophones at 5- and 45-cm depth.

In the following section, it will be shown that the sound speed estimates are consistent with the Biot theory predictions of the full time domain array response, using measured sediment parameters such as porosity, grain size and permeability, in combination with historical parameter values for this type of sand.

The variation of the time of arrival between hydrophones at different burial depths on the same pole is first estimated as a function of grazing angle. These delays are then compared to the theoretical delays derived from the standard reflection theory in discretely layered fluids described above in Eqs. (2.2) and (2.3).

The analysis will be performed using the hydrophones on the central pole C. To evaluate the frequency variation of the depth-dependent arrival delay, the analysis is performed in the frequency domain in terms of phase difference.

Figure 9 shows the phase shift variations *versus* frequency obtained for four different grazing angles. At 31.5° grazing angle, the results are consistent with the theory for a plane interface. The phase shifts (solid curve) measured between the hydrophones buried at 5 and 25 cm are identical to those (dash-dotted curve) measured between the hydrophones buried at 25 and 45 cm. Moreover, the phase shifts (dashed line) measured between the hydrophones buried at 5 and 45 cm are twice the delay between the close neighbors, as expected. The close to linear frequency dependence is evidence of the high coherence of the penetrated signal above critical incidence.

The phase shifts clearly decrease with decreasing grazing angle, and continue to be nearly linear in frequency even at the lowest grazing angles, at least below 5 kHz where the evanescent coupling is dominant. At the higher frequencies, the phase shifts fluctuate, again suggesting a scattering mechanism being responsible for the penetration. The sound

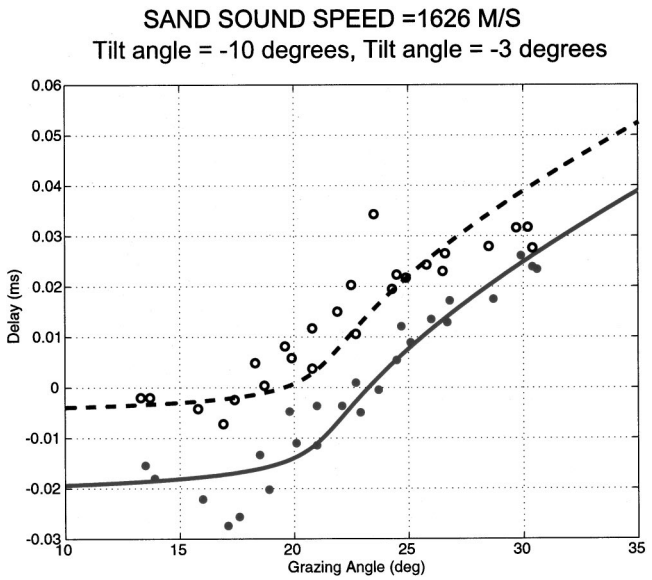


FIG. 10. *In situ* critical angle and sound speed estimation. Comparison of measured and predicted phase delays as a function of grazing angle. Filled and empty circles indicate the measured delays between the hydrophones at 5 and 25 cm, and the hydrophones at 25 and 45 cm, respectively. The solid and dashed lines represent theoretical predictions providing the best match to the data, both assuming a sound speed of 1626 m/s, but different tilt angles of -10 and $-3°$, respectively.

speed estimation analysis was therefore restricted to the 2–5 kHz band. Within this band, it is assumed that there is no dispersion (i.e., sound speed is constant versus frequency), and the phase shift ϕ can therefore be expressed as $2\pi f\tau$, where τ is the temporal delay. The slope of the phase shifts in Fig. 9 therefore provides an estimate of the delays.

Using this procedure the delays between hydrophones were estimated from the *in situ* data for all grazing angles, ranging from 13.5 to 31°, in steps of one degree. Figure 10 shows the results as a function of grazing angle for the hydrophones on pole C. The estimated phase delays between the hydrophones at 5 and 25 cm depth are shown as filled circles, and the delays between 25- and 45-cm depth are shown as empty circles.

The theoretical phase delays are given by the imaginary part of the argument to the exponential function in Eq. (2.2). However, the poles with the buried hydrophones were not perfectly vertical as illustrated in Fig. 11, and the projected tilt angle θ_t must be taken into account. Thus for two hydrophones separated a distance z along the pole, the associated coordinate separations are $(\Delta x,\Delta z)=(z\sin\theta_t, z\cos\theta_t)$. Consequently, the theoretical delay between the two hydrophones is

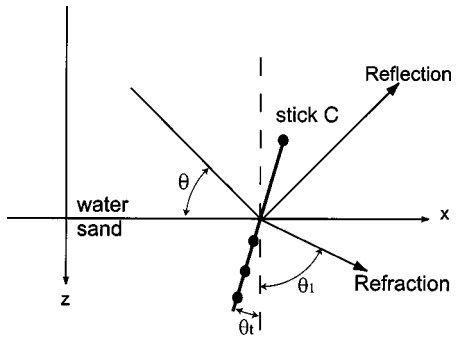


FIG. 11. Reflection and transmission geometry for tilted hydrophone pole.

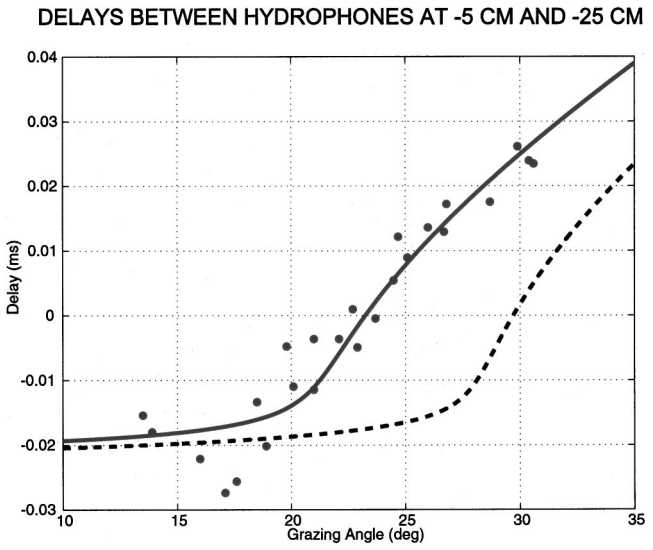


FIG. 12. Comparison of measured and predicted delays with the prediction obtained assuming the sand to be nondispersive with a sound speed of 1720 m/s as measured at 200 kHz.

$$\tau_{ab}=\frac{1}{2\pi f}\mathrm{Re}[kz(\sin\theta_t\cos\theta+\cos\theta_t\sqrt{n^2-\cos^2\theta})]. \tag{2.5}$$

Note that the index of refraction n is complex due to the attenuation, which was assumed to be 0.5 dB/λ, as estimated above in connection with the penetration ratio analysis. This value is typical for a sandy sediment.[6]

The theoretical predictions that best match the data are indicated by the solid and dashed curves, respectively, in Fig. 10. The delays for both hydrophone pairs are best matched assuming a sediment sound speed of 1626 m/s. However, the tilt angle giving the best match was -10 deg for the two upper hydrophones and -3 deg for the two lower hydrophones. This difference is not unreasonable given that the pole was somewhat flexible and was deployed by divers.

The experimental estimates and the theoretical match indicate a distinct change in slope, or ''elbow,'' of the delay curves at the critical angle, separating a region with constant delay corresponding to the evanescent penetration regime and the supercritical regime where the penetrated field is propagating vertically and therefore is associated with an increasing phase delay for increasing incident grazing angles. Based on the scattering of the experimental results around the theoretical prediction, the standard deviation of the position of the ''elbow,'' and therefore the critical angle estimate, is approximately 1 deg. In terms of sound speed estimate, this corresponds to a resolution of order 10 m/s, which is significantly better than what can be achieved by simple correlation analysis between array elements.

The robustness of using the grazing angle dependence of the phase delays to estimate sediment speed is illustrated in Fig. 12. Here the data and theoretical match for one of the hydrophone pairs are compared to the theoretical prediction achieved using a sediment sound speed of 1720 m/s as estimated from the core measurements. This comparison makes

it evident that the high-frequency core measurements of sediment speed lead to a heavily overestimated critical angle at lower frequencies, suggesting very strongly that dispersion be accounted for in sandy sediments. As indicated in Fig. 5, the Biot theory predicts a sediment sound speed 1610–1650 m/s in the 1–10-kHz frequency regime if the sound speed at 200 kHz is 1710 m/s, and the permeability is 1.7×10^{-11} m^2 as measured.

Even though the present analysis suggests that Biot theory is needed to represent the strong frequency dispersion of sandy sediments, it is highly desirable to use a simpler acoustic-elastic representation of the sediment, with its more limited requirements in terms of environmental characterization. In the following section it will be shown that the experimental penetration results in the 2–16 kHz regime are consistently and accurately modeled by both full Biot theory and by classical acoustic theory using the *in situ* critical angle and sound speed estimates, thus eliminating the need for extremely involved laboratory procedures for sediment characterization.

III. OASES TIME SERIES MODELING

The OASES wave number integration code[13] is well established for modeling seismo-acoustic propagation in arbitrary stratifications involving fluid, elastic, and poroelastic layers or halfspaces, and is capable of directly modeling the experimental scenario of the previously described experiment.

The standard OASES model computes the field, assuming all interfaces to be plane, and is therefore applied to directly compute the coherent component of the bottom-penetrated field. Even though the model is capable of modeling arbitrary source beam patterns, the source field will here be assumed to be a plane wave, incident on the bottom at the nominal grazing angle. The impulse responses computed by OASES are then convolved with the incident field as measured by the hydrophone above the seabed.

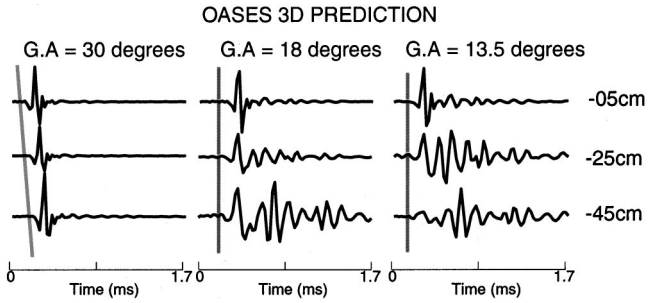


FIG. 13. OASES prediction of signals received on the central pole with hydrophones buried at 5-, 25-, and 45-cm depth, for grazing incident angles of 13.5, 18, and 30°. The direct signal measured by the hydrophone in the water is assumed as representing an incident plane wave field. The individually normalized traces may therefore be directly compared to the corresponding experimental results in Fig. 7. The sand is represented as a fluid medium with sound speed 1626 m/s, and the roughness is a random realization of the measured roughness statistics, with a 1.5-cm rms elevation.

As suggested by the analysis above, the scattering due to seabed roughness is a significant contributor to the penetration at higher frequencies. It was demonstrated in the earlier study[1] that the new OASES-3D code for modeling 3D scattering from rough interface patches in a stratified seabed[14] accurately predicts the energy levels of the scattering component. This code will be used here to directly model the temporal response of the buried hydrophones to the scattered field.

To isolate the seabed penetration effect, the sea surface and deeper sub-bottom interfaces are ignored in the modeling. Figure 13 shows the OASES predictions of the signals on the three buried hydrophones at 5-, 25-, and 45-cm depth for 13.5, 18 and 30° grazing incidence. The qualitative agree-

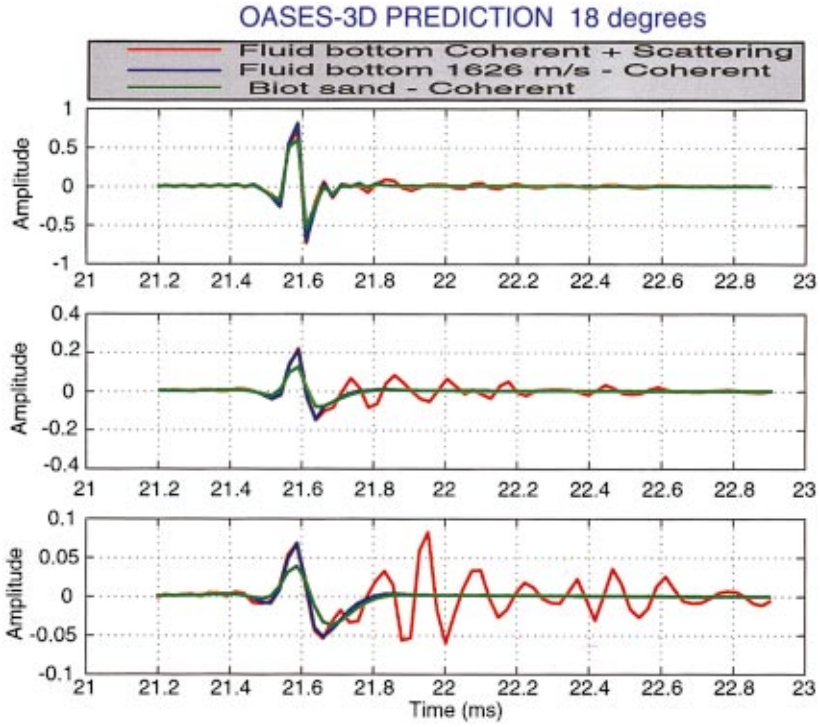


FIG. 14. OASES prediction of signals received on the central pole with hydrophones buried at 5-, 25-, and 45-cm depth, for grazing incident angle 18°. The red curve shows the full response including the scattered field, computed assuming the sand to be a fluid medium with sound speed 1626 m/s. The blue curve shows the coherent component, dominating the initial part of the penetrated signals. The green curve shows the coherent prediction when the sand is represented by Biot theory using the parameters in Table I.

ment between the coherent part of the simulated signals and the experimental results in Fig. 7 is obviously very good. The quantitative agreement in the depth dependence of the signal amplitudes for the coherent component is evident from Fig. 8, where the theoretical results are obtained using the same classical formulation for the penetration as used by OASES. The coherent component was calculated using a fluid representation of the sediment, with sound speed 1626 m/s, as suggested by the *in situ* estimate. The scattering field was computed deterministically using modeled realizations of a seabed with given estimated roughness statistics,[1] i.e., ripples with 25-cm correlation length, 40° aspect, and 1.5-cm rms height. The scattered field obviously cannot be expected to be modeled correctly deterministically, although the spatial and temporal characteristics of the experimental data are clearly captured accurately by the model.

The fact that OASES computes the coherent and scattered components separately allows for directly analyzing the significance of each component. Thus Fig. 14 shows the coherent component on the three hydrophones for 18° grazing incidence as a blue curve, while the total field is indicated by the red curve. This confirms the earlier interpretation of the experimental results, concluding that the initial part of the signal is dominated by the coherent component described by classical penetration theory, while the ''tail'' of the signals is associated with the seabed scattering which increasingly dominates with increase in depth and decrease in grazing angle.

Figure 14 also shows, as a green curve, the coherent prediction obtained using the full Biot representation of the sediment with the sediment parameters in Table I. Clearly, within the band of the parametric sonar, the dispersion effect is insignificant, and the classical theory produces accurate results, both in terms of amplitude and phase of the penetrated signals. The agreement is of equal quality for the other incident angles. This agreement in turn suggests that classical theory adequately represents the penetration of sound into sandy sediments, at least within a frequency decade, provided an *in situ* sediment sound speed is used.

IV. CONCLUSIONS

The strong frequency dispersion of sandy sediments predicted by Biot theory and confirmed by earlier analysis of acoustic bottom penetration data was investigated by devising an independent, *in situ* estimation procedure for sediment sound speed. A parametric source was used to insonify a sandy seabed in the frequency range 2–16 kHz, and a buried hydrophone array was used to measure the bottom penetration over a wide range of grazing angles. It was demonstrated that the grazing angle dependence of the phase delay in the seabed provides a robust and accurate measure of the sediment critical angle, and therefore the sound speed. Consistent with earlier analysis of penetration data from other geographical areas, the present estimates of bottom sound speed confirm the presence of a significant frequency dispersion in sandy sediments, with the estimated speed at 2–5 kHz being 1626 m/s while the sound speed measurements performed at 200 kHz on core samples yielded a speed of 1720 m/s. However, as demonstrated here, this dispersion is totally consistent with the predictions of Biot theory for this type of sand. This dispersion is significant to the prediction of the sediment critical angle. On the other hand, the study has shown that, within approximately a frequency decade, the experimental data are adequately modeled by classical theory provided the *in situ* sound speed estimate is used, eliminating the need for full Biot modeling with its elaborate sediment characterization. Finally, the very good agreement obtained between the experimental data and the model predictions for both the coherent and scattered contributions provides valuable validation of the OASES modeling framework, including the new 3D seabed scattering model.

ACKNOWLEDGMENTS

The authors would like to thank the SACLANTCEN personnel without whom this work would not have been possible, in particular the crew of the R/V MANNING, P. A. Sletner and R. Chiarabini of the Engineering Technology Department, and M. Mazzi and M. De Grandi of the Mine Countermeasures Group. The effort by Dr. T. Lyons in performing the stereo photogrammetry measurements and producing the bottom roughness statistics is highly appreciated. Also, the comments and suggestions provided by Dr. N. G. Pace contributed significantly to the quality of the paper and are highly appreciated. The modeling and analysis performed at MIT was funded by the Office of Naval Research, Ocean Acoustics Program.

[1] 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).
[2] R. B. Stoll, *Sediment Acoustics* (Springer-Verlag, New York, 1989).
[3] S. Fioravanti, A. Maguer, W. L. J. Fox, L. Gualdesi, and A. Tesei, ''Underwater rail facility for highly controlled experiments at sea,'' submitted to Third European MAST Conference, Lisbon, Portugal, 1998.
[4] D. J. Wingham, N. G. Pace, and R. V. Ceen, ''An experimental study of the penetration of a water-sediment interface by a parametric beam,'' J. Acoust. Soc. Am. **79**, 363–374 (1986).
[5] A. P. Lyons, T. Akal, and E. Pouliquen, ''Measurements of sea floor roughness with close range digital photogrammetry,'' *OCEANS 98*, Nice, September 1998.
[6] F. B. Jensen, W. A. Kuperman, M. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, New York, 1994).
[7] D. J. Wingham, ''The dispersion of sound in sediment,'' J. Acoust. Soc. Am. **78**, 1757–1760 (1985).
[8] R. D. Stoll and T. K. Klan, ''Reflection of acoustic waves at a water-sediment interface,'' J. Acoust. Soc. Am. **70**, 149–156 (1981).
[9] J. E. White, *Underground Sound. Application of Seismic Waves* (Elsevier, New York, 1983).
[10] L. M. Brekhovskikh, *Waves in Layered Media*, 22nd ed. (Academic, New York, 1980).
[11] E. I. Thorsos, D. R. Jackson, J. E. Moe, and K. L. Williams, ''Modeling of subcritical penetration into sediments due to interface roughness,'' in *High Frequency Acoustics in Shallow Water*, edited by N. G. Pace, E. Pouliquen, O. Bergem, and A. P. Lyons, SACLANTCEN Conference Proceedings Series, CP45, pp. 563–569, NATO SACLANT Undersea Research Centre, June 1997.
[12] E. Pouliquen, A. P. Lyons, and N. G. Pace, ''Penetration of acoustic waves into sandy seafloor at low grazing angles: The Helmholtz-Kirchoff approach,'' SACLANTCEN SM290, NATO SACLANT Undersea Research Centre report, La Spezia, Italy, 1998.
[13] H. Schmidt, *OASES: Version 2.1. User Guide and Reference Manual* (Massachusetts Institute of Technology, Cambridge, MA, 1997).
[14] H. Schmidt and J. Lee, ''Physics of 3D-scattering from rippled seabeds and buried targets in shallow water,'' J. Acoust. Soc. Am. **105**, 1605–1617 (1999).

Measurement technique for bottom scattering in shallow water

Charles W. Holland, Reginald Hollett, and Luigi Troiano
SACLANT Undersea Research Centre, Viale San Bartolomeo, 400, 19138 La Spezia, Italy

(Received 20 July 1999; revised 3 February 2000; accepted 17 May 2000)

Sonar performance predictions of reverberation in shallow water rely upon good estimates of the bottom-scattering strength. However, little is understood about bottom scattering in shallow water in the frequency range 400–4000 Hz, particularly its dependency upon frequency and its relationship to the physical properties of the seafloor. In order to address these issues, a new measurement technique has been developed to probe the frequency and angular dependency of bottom-scattering strength. The experimental technique is described which employs either coherent or incoherent sources (lightbulbs). In addition, measurement and modeling results for two diverse shallow water sites are presented. At one site, the scattering appears to arise at or near the water–sediment interface. At the other site, scattering from a 23-m sub-bottom horizon is clearly apparent in the data at and below 1800 Hz. The fact that our measurement technique can directly reveal the presence of sub-bottom scattering is a significant advance in the development of methods to explore the physical mechanisms that control bottom scattering. [S0001-4966(00)00109-0]

PACS numbers: 43.30.Gv, 43.30.Xm, 43.30.Ma [DLB]

INTRODUCTION

The traditional method for estimating bottom-scattering strength in shallow water is by measuring long-range reverberation, and extracting the scattering strength via modeling. A few examples in the frequency range of interest of this study include Urick (1970), Cole and Podeswa (1974), Blanc *et al.* (1978), Zhou *et al.* (1983), Ellis (1995), Scanlon *et al.* (1996), and Cable *et al.* (1997). The advantages of this traditional method are that the measurements are easy to conduct and that the scattering strength is averaged over a large spatial area. The disadvantage of the method is that the scattering strength estimate is subject to a rather large set of assumptions.[1] A significant assumption pertains to the incident field at the seafloor. Frequently it is not measured, e.g., Blanc *et al.* (1978) assume a perfectly reflecting bottom. Sometimes transmission loss (TL) is measured in the water column; however, this does not guarantee the correct modeling of the incident field at the seafloor, which depends upon knowing the geoacoustic properties. Since reverberation can be dominated much more by the propagation than by the scattering events, it is inherently difficult to extract scattering information without knowledge about the propagation. Moreover, it is almost universally assumed that the scattering occurs at the water–sediment interface. Other common assumptions include range and azimuthal independence in TL and/or scattering strength and assumptions about the absence or presence of air–sea interface and biologic scattering. It is worth noting that the reported frequency dependence of the scattering strength can be erroneous if any of these assumptions are incorrect. Thus, the ability to probe the frequency dependence of shallow-water bottom scattering using long-range reverberation data may be limited.

An alternative approach is to make direct-path measurements of bottom scattering. The disadvantage of this type of measurement in shallow water is the presence of multipaths, which makes it difficult to isolate specific paths of interest. Another disadvantage is that the measurement only samples a very small region, which may not be immediately useful for making reverberation predictions over large regions. However, the principal advantage of the direct path measurement technique is that many of the assumptions that plague interpretation of the reverberation data are eliminated, so that the frequency and angular dependence of bottom scatter can be studied in detail. Thus, the direct-path measurements might also serve as a basis from which to examine the validity of the assumptions commonly applied to extract scattering strength from reverberation measurements. Another important advantage of the direct-path technique is the potential for revealing the dominant scattering mechanisms in a particular environment, i.e., for discerning the relative contributions of water–sediment interface, sub-bottom volume, or sub-bottom horizon scattering.

The scope of this paper is to describe the direct-path bottom scatter measurement technique and show measurement results in two diverse sedimentary regimes. The Capraia Basin (see Fig. 1), was selected as the study area because of its diversity in bottom types including sedimented regions with complex layering and magmatic rock outcrops.

While the main of focus of the paper is to describe the experimental technique and present the resulting data, some modeling results are also presented. The object of the modeling is to reinforce and illuminate the capability of the measurements to distinguish between water–sediment interface scattering and sub-bottom scattering directly, i.e., without resorting to the usual deductions or conjecture. This is an important advance, because up to the present, bottom-scattering analyses have inferred the presence of sub-bottom scattering based on the frequency and/or angular dependence of the data, or deductions made about the environments [e.g., see Mourad and Jackson (1996), Novarini and Caruthers (1998), and Holland and Neumann (1998)], rather than on direct evidence in the data.

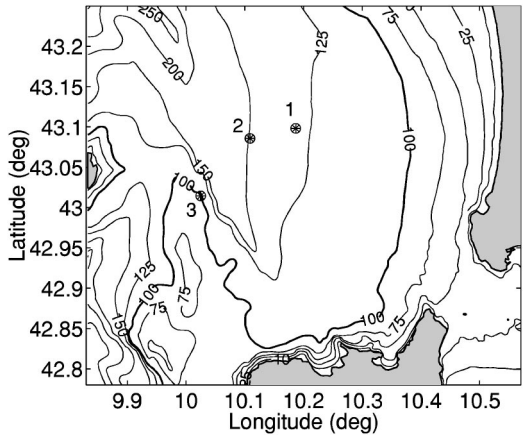


FIG. 1. Experiment area showing location (*) of bottom-scatter measurements. Measurement locations are designated from east to west as site 1, site 2, and site 3.

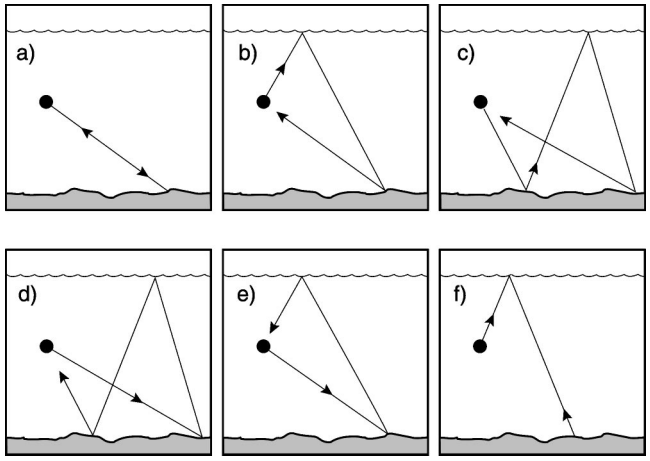


FIG. 2. Bottom-scattering multipaths. Only those paths that have less than two surface interactions are depicted.

I. EXPERIMENT DESIGN

The experimental design was dictated by the requirement to conduct the measurements in shallow water (water depths of 50–200 m), over a large bandwidth (400–4000 Hz) with greatest interest on low to intermediate grazing angles (0–45 deg) that control sonar system performance.

Perhaps the most significant challenge of such an experiment in shallow water is to avoid or control multipaths. One multipath problem is potential contamination by sub-bottom reflections. It is well-known that scattering may arise from not only the interface but also from sub-bottom inhomogeneities or horizons. However, sub-bottom reflections at normal incidence contaminate the scattering measurement, since the normal incidence reflections are often at higher amplitude than the scattering events.

A second multipath problem is contamination from hybrid paths. Hybrid paths are paths that belong to a different family of scattering event that arrive at the same time as the scattering path that is being measured. Figures 2 and 3 show the various paths, and their relationship in time and angle. For example, in this geometry, beyond about 0.4 s, the monostatic (path a) and bistatic[2] paths (b and d) cannot be separated in time or angle. A receive and/or source array with vertical aperture can be used to control both types of multipaths, i.e., by reducing the contribution of the normal incidence reflections and also providing discrimination against the various scattering paths. Short pulse lengths are required in order to resolve angular dependence of the scattering function.

Figure 4 depicts the system employed in this study. The vertical aperture of the receive array helps minimize effects of multipath, in addition it provides the potential for identification of scattering mechanisms (see Sec. III). Two kinds of sources were employed. Short pulses from sources located near the bottom of the array provided a repeatable and stable source. In addition, standard 100-W lightbulbs were employed as inexpensive and safe broadband pulses. The equipment is deployed from the forecastle and the ship is either set in a single-point moor or left to drift. The weight of the Mod-40s (90 kg in water) provides enough ballast to keep the array straight when the current shear is small. A small fin (not shown) on the Mod-40 frame stabilizes the array from rotational forces.

The source–receive array was placed at various heights above the seafloor in order to explore the potential for inferring sub-bottom scattering from diversity in source heights. In order to stay in the farfield of the seafloor, the experiments were generally conducted with source/receiver heights greater than 20 m.

In addition to the direct-path bottom-scattering measurements, associated acoustic and geophysical measurements were conducted. The purpose of these measurements was to provide the environmental basis to understand the relationship between the scattering and the physical properties of the sediment. Shallow-water bottom reflection loss measurements [Holland and Osler (2000)] were designed to provide high resolution (of the order of 50 cm in the vertical plane) sub-bottom deterministic properties. Core data provided additional sound speed and density data in the upper 5 m. Seismic reflection and SWATH bathymetry data provided regional understanding of sub-bottom layering structure and water–sediment interface morphology.

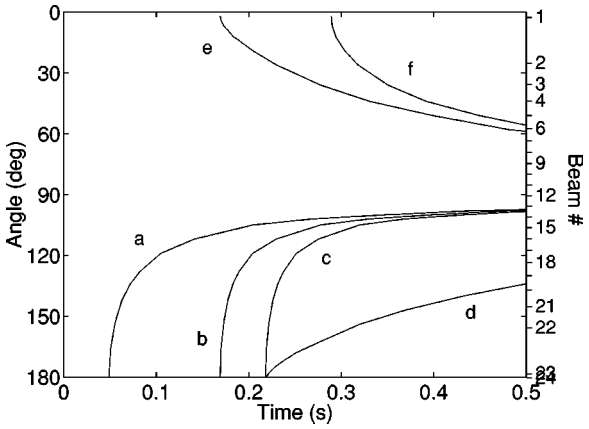


FIG. 3. Vertical arrival angles versus time for various bottom scattering paths. Source–receiver depth is 91 m and the water depth is 128 m. Time is referenced to source initiation. Angles are measured with 0 deg towards the sea surface. Path (a) is the monostatic backscattering path. See Fig. 2 for other path descriptions.

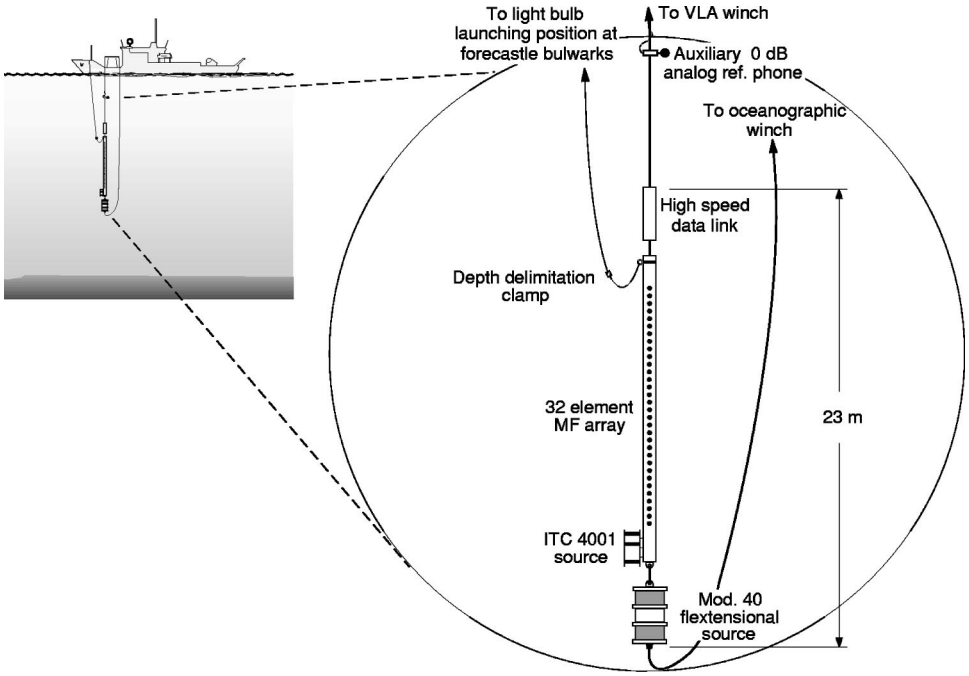


FIG. 4. Experiment geometry used for measuring bottom scattering. Details of the source and receive arrays are found in the text.

A. Coherent sources

Modeling indicated that for some environments (large reflection coefficients at normal incidence relative to the scattering) directivity in both the source and receiver is required in order to control bottom and sub-bottom reflections. Thus, a source array was constructed with pairs at $\lambda/2$ spacing in the vertical so that transmitting in-phase would yield a theoretical null on the vertical axis (see Fig. 5). The midfrequency source array was constructed of three ITC-4001 transducers spaced at 18 and 42 cm, to yield $\lambda/2$ spacing at 3600, 1800, and 1200 Hz. The top of this array was placed 1 m below the bottom phone of the receive array. The low-frequency array consisted of two Mod 40 flextensional transducers spaced at 1.25 m (or $\lambda/2$ spacing at 600 Hz). Source-array calibration techniques and results are provided in Appendix A.

Ping types included both continuous wave (cw) and LFM pulses of 15-ms pulse length. Repetition rates of 6 pings/min were used, and the acquisition system was triggered 1 s before transmit to acquire ambient noise for each ping. Ten to 15 pings were collected for each source/receiver depth.

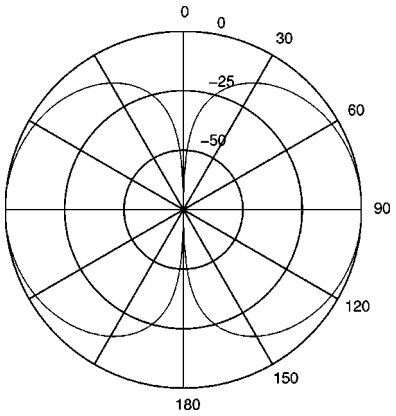


FIG. 5. Theoretical beam pattern from two sources spaced at $\lambda/2$, which is symmetric around the vertical axis.

B. Lightbulb sources

The acoustic signal from an ordinary lightbulb that is imploded can make a useful source for undersea experiments [see Heard *et al.* (1997) for a good review of the subject]. The signal is generated from oscillations of the gas bubble that is released after the implosion. The advantage of the lightbulbs as acoustic sources is their low cost and safety relative to explosives if a high source level is not required. Another advantage is their short time signature—on the order of 5 milliseconds at 100-m depth. While lightbulbs and laboratory glassware have been used by other researchers for various purposes, this is the first time to the author's knowledge that they have been used for bottom-scattering strength measurements.

The nominal crush depth for a typical lightbulb is about 270 m. The experiment concept required that implosion depth be easily controlled from depths of 20–150 m. This was accomplished by attaching the lightbulb to a 4-kg weight which was designed to slide freely down a kevlar rope attached to the top of the receive array. A depth delimitation ring placed on the wire 1.5 m above the top of the receive array served as the crush plate. Rubber padding on the bottom of the weight minimized a secondary acoustic signal arising from the weight striking the crush plate.

Figure 6 shows a lightbulb source signal at about 90-m depth. The acoustic pressure is about 180 deg out of phase with the bubble displacement. Thus, the pressure peaks occur when the bubble is at its smallest radius. Five or six acoustic peaks are visible before the potential energy of the gas bubble is dissipated [see McDonald and Holland (2000) for a model of bubble pulse decay characteristics].

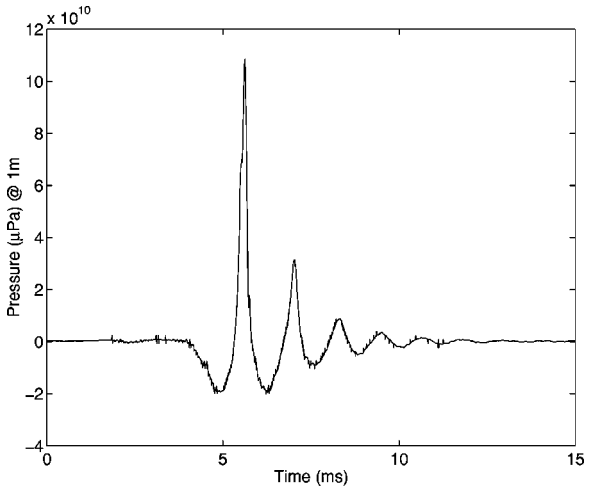


FIG. 6. Lightbulb implosion at about 90 m depth.

Figure 7 shows the source level for three consecutive shots at the same depth. The undulating behavior of the source level at the low frequencies arises from constructive and destructive interference between the bubble pulses. One of the shots was audibly different from the others and possibly was misaligned on the rope when it struck the crush plate. For this shot the second bubble pulse was much lower than the others. Thus, a disadvantage of lightbulb sources is the variability in source signature that requires measurement of each implosion.

Preparation time required for each lightbulb deployment gave an implosion rate of typically 1 implosion every 4 min. The triggering was done manually using the observed 2-m/s fall rate of the weight and lightbulb assembly.

C. Receiver

The receive array consisted of 32 Benthos AQ-4 hydrophones with 0.18-m spacing. The data were sampled at 12 kHz and low-pass filtered at 3.8 kHz with a seven-pole six-zero elliptic (70 dB per octave roll off) antialias filter. The RC high-pass filter (6 dB per octave roll off) was set at 150 Hz. A high-speed digital link within the array provided programmable signal conditioning, digitization and serialization of the signals. The array was hardwired directly to the NATO Research Vessel (NRV) ALLIANCE. Nonacoustic data (gains, filter settings, etc.) were interleaved in the serial stream.

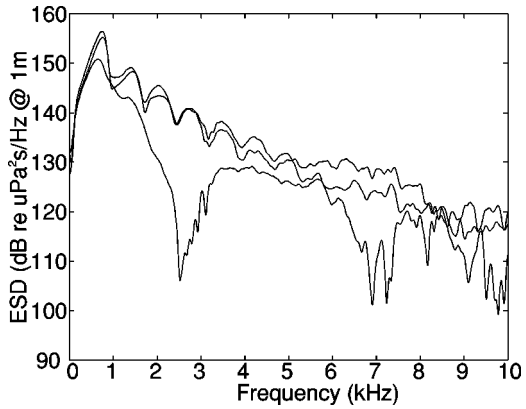


FIG. 7. Source level (energy spectral density) for three successive implosions at 90-m depth. The variability in source signature requires measurement of each bulb implosion.

A separate reference hydrophone, 14 m above the crush plate, was employed in order to acquire the data at a higher sampling rate (50 kHz) which was necessary for studying the details of the lightbulb source characteristics. A separate reference hydrophone was also required because the dynamic range of the vertical array was insufficient for acquiring both the direct blast and the scattered field.

II. DATA PROCESSING

Data processing details are presented for several types of processed outputs. The simplest processing is a display of the filtered, beamformed results. These data can then be further processed to produce bottom scattering strengths.

A. Beam-time series

Following signal conditioning, data were beamformed (Hanning shading) using a plane wave time-domain beamformer [see Sylva *et al.* (1986)] implemented in hardware. Beams were spaced to yield 3-dB down crossing points at the design frequency of the array (see Fig. 3). The data are filtered with a sixth-order low-pass digital elliptic filter with 0.5 decibels of ripple in the passband and a stop band 50 dB down. Bandwidths were chosen to be 150 Hz.

Experimental results from a single ITC source transmitting at 3600 Hz are shown in Fig. 8. The water depth and source depth are 128 and 91 m, respectively. Zero time in the figure corresponds to the direct blast, which overloads the array and is seen on all beams. The surface reflection is visible at 0.1 s. Also shown are the theoretical results of arrival time vs beam in the white lines which correspond to the various monostatic and bistatic scattering paths as shown in Figs. 2 and 3. These data are expected to be useful for exploring the relationship between the monostatic and bistatic scattering and the sediment geoacoustic properties.

However, the main point we wish to emphasize in this study is that these data can be used for directly determining (i.e., without any modeling) the scattering mechanism. Figure 9 shows a cartoon that depicts how the various scattering mechanisms are manifest in the measured beam time series. In Fig. 9(a), the pulse shape in the sediment at an arbitrary instant in time is depicted in black. In an isovelocity medium the pulse shape would be a spherical shell, but in general, its shape will be determined by the sediment sound-speed profile. Scattering may arise from the water–sediment interface as is shown in Fig. 9(b), which shows just the monostatic scattering branch (or branch ''a'' of Fig. 3). Scattering may also arise from a sub-bottom layer; Fig. 9(c) shows the monostatic scattering branch for scattering from the sediment interface and a sub-bottom layer. Figure 9(d) shows results for scattering from sediment volume inhomogeneities distributed in depth above the sub-bottom layer. Note that at any instant in time, sub-bottom scattering results in arrivals at grazing angles higher than that from the water–sediment interface.

Thus, by inspection of the beam-time series, the presence or absence of sub-bottom volume scattering and sub-

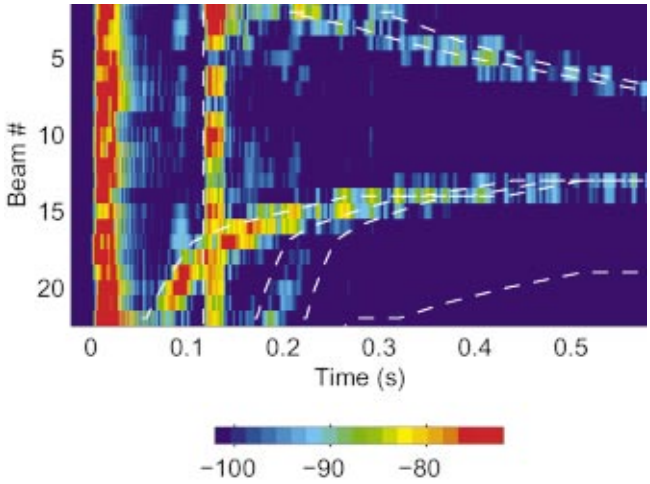


FIG. 8. Measured beam time series at site 1; the units are dB *re* 1 μPa^2/s/Hz. The water depth and source depth are 128 m and 91 m, respectively. Predictions of multipath structure are overlaid (white lines) on the data and correspond to the paths shown in Figs. 2 and 3.

bottom layer scattering can be observed. For sub-bottom layer scattering, the scattering will be observed in the data as an additional branch [see Fig. 9(c)]. Sub-bottom volume scattering will be observed in the measured data as a broadening in beam space (always toward steeper grazing angles) of a branch, relative to water–sediment interface scattering [see Fig. 9(d)].

In practice, the ability to observe these features from the measurements will be limited by the pulse length; the 15 ms used in this study means that interface and sub-bottom scattering in the upper 10 m cannot be distinguished. There is also an observational limit imposed by the receive array aperture. Sub-bottom scattering is expected to become more prominent at lower frequencies; however, the concomitant increase in beamwidth leads to a reduced discrimination capability. Despite these limitations, it will be shown in Sec. III that the technique is capable of directly discerning the presence of sub-bottom scattering.

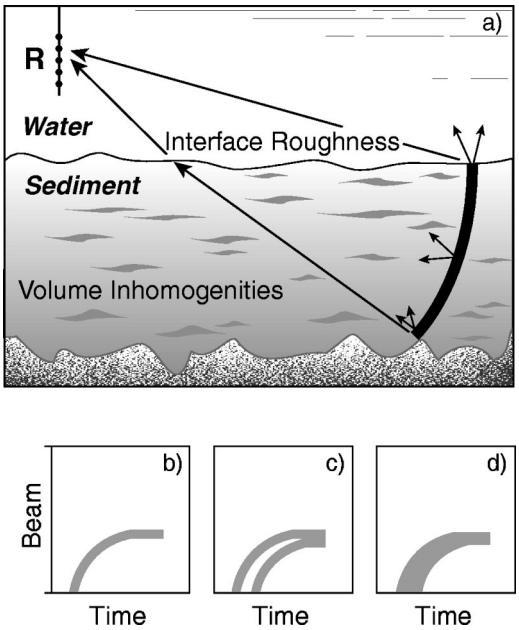


FIG. 9. Cartoon depicting how the various scattering mechanisms can be determined from the measured beam-time series. (a) The pulse (black arc) at an arbitrary instant in time is shown along with the various scattering mechanisms. The resulting beam-time series from each mechanism are shown as follows: (b) scattering from the water–sediment interface; (c) scattering from water-sediment interface and sub-bottom interface; and (d) scattering from sediment volume inhomogeneities.

B. Scattering strength from coherent pulses

The data of Fig. 8 can be processed for both monostatic and vertically bistatic scattering strength. In this paper the monostatic processing technique is demonstrated. For a lossless isovelocity medium, the instantaneous intensity i_o received from a scattering area can be written as

$$i_o = \int\int isbb' r^{-4} \rho \, d\rho \, d\vartheta , \tag{1}$$

where i is the transmit intensity, s is the scattering cross section, b and b' are the transmit and receive beam patterns, respectively, r is the distance from the array to the scattering patch, and ρ and θ are the radial and azimuthal polar coordinates (see Fig. 10). If the beam pattern is invariant over the insonified area, then

$$s = i_o \left(bb' 8\pi \int i \, dt \right)^{-1} c^2 t^3 (1-\tau/2t)^2 (1+\tau/2t)^2 . \tag{2}$$

When the pulse length is short ($\tau \ll 2t$), the last two factors of Eq. (2) can be ignored, yielding an expression identical to Chapman and Harris (1962).

The above assumptions are appropriate for the experimental results presented here; the frequencies are low enough and ranges short enough so that seawater absorption is negligible, the beam pattern is slowly varying with vertical angle and independent of azimuth, and the sound-speed profiles are roughly isovelocity[3] within 30 m of the seafloor (see Fig. 11). The geometry is nearly monostatic with vertical offset between the center of the receive array and the ITC and Mod40 sources of 4 and 8 m, respectively. These offsets were small enough so that in the processing, the monostatic source/receiver depth was referenced to the average depth of the source–receiver pair (see Appendix B).

The only assumption which may not be met is that pertaining to the scattering process. For some environments, scattering from the sub-bottom is expected to dominate the response rather than scattering from the interface. The effect of sub-bottom scattering can either be taken into account in the processing or the subsequent modeling. Since the frequencies and environments at which sub-bottom scattering dominates are not known *a priori*, we choose to process the data as if the scattering process were arising from the interface and account for any sub-bottom contributions in the modeling.

The potential presence of sub-bottom contributions presents a practical problem of how to select the beams that contribute to the scattering strength at a given time. This was done in the following way. First, the peak arrivals as a function of beam and time are found along the monostatic branch, aided by using the theoretical results of arrival time vs beam as a starting solution. Then, adjacent beams are

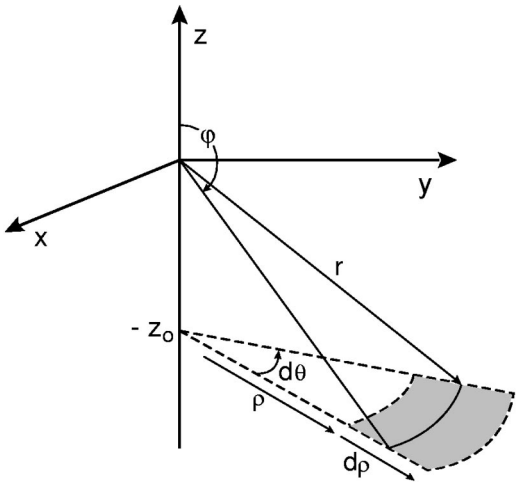


FIG. 10. Coordinate system for bottom-scattering experiment.

summed if they fall within a certain threshold. The threshold used in the following examples was 10 dB. This choice of threshold in practice provided a useful criterion to discriminate against sub-bottom reflections and bistatic paths at early times without discarding potential contributions from neighboring beams. At late times, the bistatic paths merge with the monostatic path in beam-time space, and a measurement of backscattering strength becomes impossible unless the bistatic contributions are weak relative to the monostatic path.

Averaging was performed over several (typically ten) pings. It is more convenient to perform the averaging on the received level data rather than the scattering strength. This is possible since the experiment geometry and the source function are constant from ping to ping. The signal-to-noise ratio (SNR) criterion was 6 dB for both the instantaneous received level and a smoothed (30-ms sliding window) average.

C. Scattering strength from lightbulb implosions

Scattering strength processing from the lightbulb sources was similar to that from the coherent pulses [i.e., Eq. (2) without the last two terms]. However, the sometimes significant variability in source spectrum necessitated processing source energy flux density for each implosion as well as requiring every implosion to be fully processed to a scattering strength before averaging. Another step required in the lightbulb processing was the manual picking of the direct and surface arrivals in order to measure the implosion time and depth. Figure 12 shows an example of the data (Hilbert transform from a single phone) employed in this step.

III. EXPERIMENT RESULTS

In this section, bottom scattering measurement results are provided for site 2 and site 3, showing first the filtered beam-time series and then the processed scattering strengths. It is demonstrated that the filtered beam-time series data can be used to directly determine the scattering mechanism. In addition, beam-time series model predictions are shown which serve to illustrate the limits of the measurement technique. Detailed modeling, including extraction of sediment statistical properties, will be undertaken in a later study.

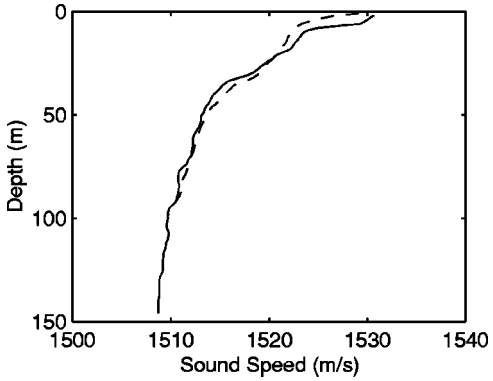


FIG. 11. Sound-speed profile from CTD casts at site 2 (solid line) and site 3 (dashed line).

For each site, the measurement results are preceded by a discussion on the environment. In particular, colocated bottom reflection data are exploited for high-resolution geoacoustic parameters. An important aspect of the bottom reflection technique used here [Holland and Osler (2000)] is that it samples the same spatial aperture (less than 1000 m) as our bottom-scattering technique. By contrast, common matched-field and other geoacoustic inversion methods generally sample spatial scales of the order of tens of kilometers and thus yield too much spatial smearing in the geoacoustics for the direct path bottom scatter analysis.

A. Site 2 environmental description

Site 2 lies on the Tuscan shelf in a water depth of 150 m (see Fig. 1). The bottom was flat in the north–south direction with a very small (approximately 0.1 deg) slope in the east–west direction. Sound speed data from a conductivity–temperature–pressure (CTD) cast are shown in Fig. 11.

One of the most crucial parameters required for understanding the dominant scattering mechanism is the sediment sound-speed profile. While some authors have indicated that an isospeed approximation is sufficient in the sediment [e.g., Yamamoto (1996) and Novarini and Caruthers (1998)], the sediment sound speed is crucial because it determines what features in the sub-bottom are illuminated and how they are

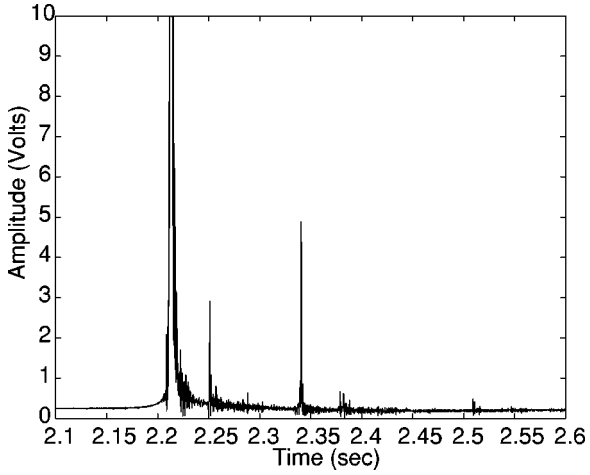


FIG. 12. Hilbert transform of data from top phone. Implosion is at about 2.21 s (clipped), bottom arrival is at 2.25 s, and surface arrival is at 2.34 s.

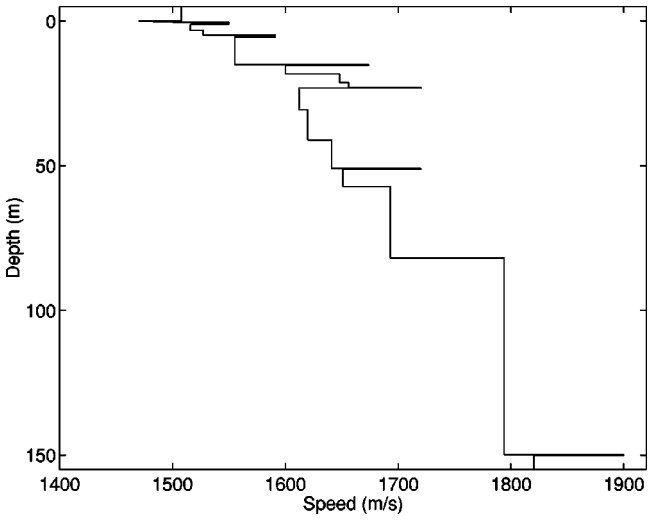


FIG. 13. Sediment sound-speed profile at site 2. Note the high-speed intercalating layers, which are hypothesized to give rise to the sub-bottom scattering.

illuminated [see Holland and Neumann (1998)]. Figure 13 shows a high resolution sediment sound-speed profile at this site [Holland and Osler (2000)]. As confirmed by core data, the upper 50 cm are a very soft silty clay (porosity of 80% at the water sediment interface) with very low sound speed and density, about 1470 m/s and 1.2 g/cc, respectively, at the interface and large positive gradients in both velocity and density. Below this there are randomly interspersed intercalating high-speed layers with velocity and density gradients between them. This kind of sediment deposition is caused by cyclic variations in sea level.

Seismic reflection data across the scattering area are shown in Fig. 14. The actual scattering experiment took

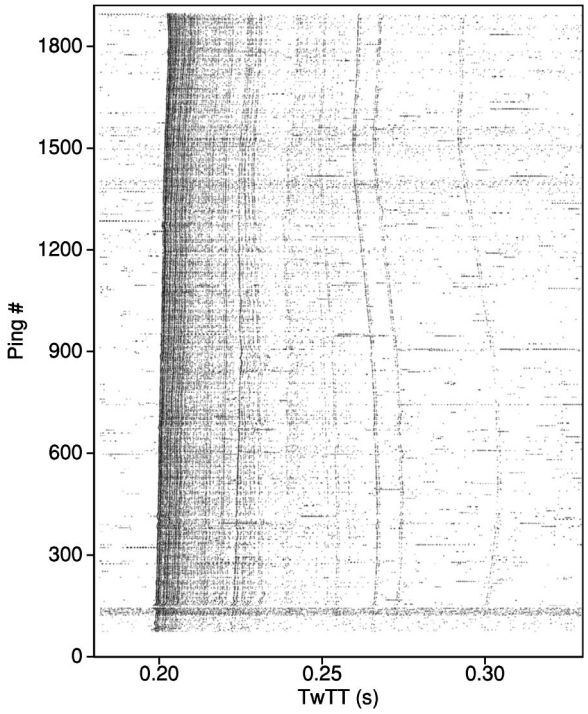


FIG. 14. Seismic data along a 7.5-km track centered at the bottom-scattering experiment location. Two-way travel time (TwTT) is in seconds.

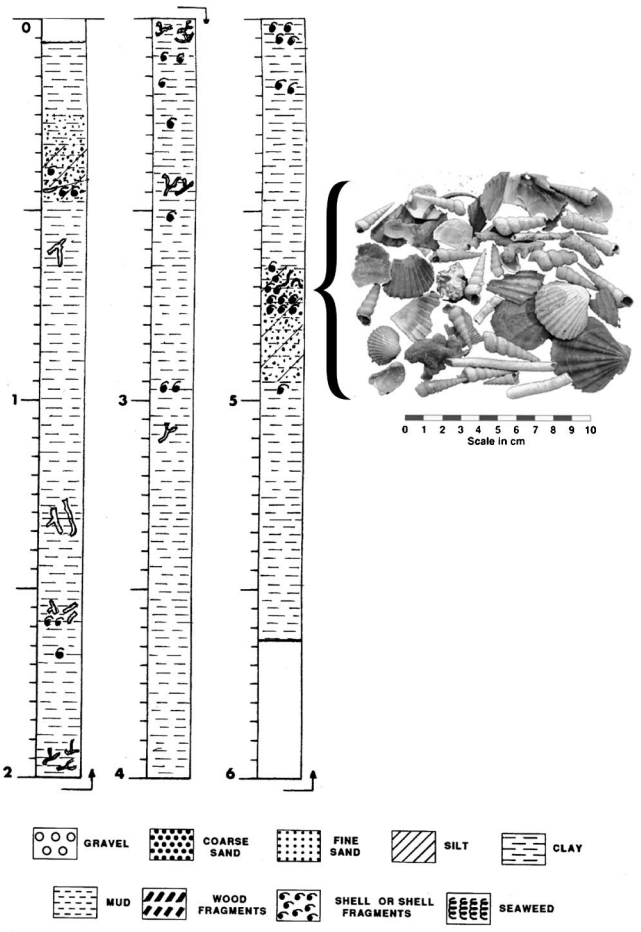


FIG. 15. Core lithology (uncorrected for compression due to the piston). Note major shell layers at 0.30 and 4.75 m, which correspond to the first two high-speed layers in Fig. 13. The inset shows a photograph of the large amount of shell material found in a 10-cm layer at 4.75 m. The number of shells per unit volume in the layer at 0.30 m was very similar. Besides these two layers, shell material per unit volume in the rest of the core is very small.

place in a small region, of the order of several hundred meters around the center of this track. These data serve to indicate that the layering structure is continuous across the region in a large-scale sense. The microstructure of the layering is of course discontinuous.

What feature(s) of the bottom is expected to give rise to the scattering? From the sediment sound-speed profile and bathymetry, one would expect that scattering from the water–sediment interface would be negligibly small. That is, both the interface impedance contrast and the roughness are small. This is also evident in side-scan data from this area which are featureless with occasional faint lines from bottom trawling. Scattering from a sub-bottom basement horizon can also be ruled out, since there was no basement observed within 150 m of the water–sediment interface.

Therefore, the most likely geologic feature to give rise to the scattering is the thin high-speed layers (i.e., the sound speed ''spikes'' in Fig. 13). Piston core data indicate (see Fig. 15) that the spikes are sandy silt layers with large shells and coral debris. They are manifest in the shown lithology at about 0.3- and 5-m sub-bottom.[4] The inset photograph shows the large amount of shell material in a 0.04-m^3 volume cen-

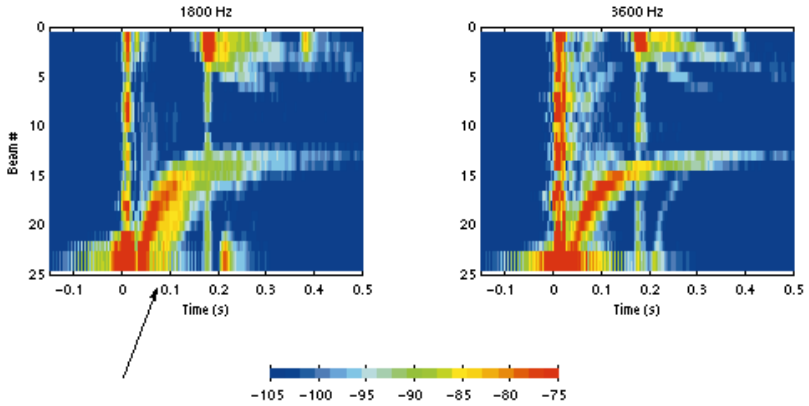


FIG. 16. Measured beam-time series (ten averages) at site 2 using two sources at (a) 1800 Hz and (b) 3600 Hz. The quantity plotted is the ratio of received level to source level in dB. The 1800-Hz data show clearly the sub-bottom scattering contributions. The arrow indicates the onset at normal incidence of the sub-bottom layer 23 m below the water–sediment interface.

tered at the deeper layer. The number of shells per unit volume in the layer at 0.30 m was very similar. Besides these two layers, the number of shells per unit volume in the rest of the core is very small. Core sampling was limited to about 6 m in depth, so the deeper spikes (at 15, 23, and 51 m) cannot be probed directly; however, it is highly probable given the depositional environment that they are also shelly sandy layers that will also give rise to scattering.

B. Site 2 measurements

Thus far, the conclusion about the probable scattering mechanism has been based solely on inference. We wish to contrast that approach with a novel aspect of our scattering measurement technique which allows the scattering mechanism to be directly observed.

Figure 16 shows the measured beam time series data at 1800 and 3600 Hz using the ITC array at a source receiver depth of 130 and 126 m, respectively. Note that the SNR is quite good and that there is no evidence of any scattering from biologics from within the water column. The surface reflection is observed through all beams at about 0.17 s. The hybrid paths (paths b, c, and d in Fig. 2) in the lower beams are substantially lower in level than the monostatic path, due the effect of the source and receiver beam patterns. The relationship between time and angle at the water–sediment interface is shown in Fig. B2.

The data in Fig. 16 can be used for directly determining the scattering mechanism simply by inspection, as discussed in Sec. II A. That is, sub-bottom scattering will be observed along a scattering branch as a broadening in beam space toward higher angles. The salient feature of Fig. 16 is that there does exist a broadening of the monostatic branch at 1800 Hz relative to the width of the branch at 3600 Hz. This must be due to sub-bottom scattering at 1800 Hz. In fact, at 1800 Hz a separate branch is apparent, as indicated by the arrow. From its travel time at normal incidence, this arrival corresponds to a scattering horizon at about 25 m below the water–sediment interface. The sound-speed inversion results in fact show a high-speed layer[5] at 23-m depth in the sediment (see Fig. 17). It is precisely this high-speed layer which gives rise to the observed sub-bottom scattering.

Thus, we may conclude directly from the beam-time series that scattering from a deep (23 m) sub-bottom layer occurs at 1800 Hz. It is somewhat remarkable that in shallow water such a deep scattering layer was observable at this frequency, i.e., at a depth of 25 wavelengths! It is more remarkable that this scattering persists from normal incidence down to fairly low angles (about 15° grazing angle). As another example of scattering from relatively deep in the sediment column, Holland and Neumann (1998) showed in a deep-water environment scattering dominated by a sub-bottom horizon at 250 wavelengths depth in the sediment (at angles from 30°–50°).

With model predictions, it is possible to explore the scattering mechanisms in even more detail. It was of interest to determine with the modeling whether or not sediment volume scattering occurs at 1800 Hz between the interface and sub-bottom and also whether there is any sub-bottom volume scattering at 3600 Hz. At this stage of the analysis the intent is to shed more light on the scattering mechanisms (rather than try to extract statistical parameters). The model treats the sediment as a single refracting layer. The modeled sound-speed profile in the upper 30 m of sediment is shown in Fig. 17.

Model predictions (a brief description of the model is found in Appendix C) were made of scattering from the water–sediment interface only and from the water–sediment interface plus a sub-bottom scattering horizon. Notwithstanding that water–sediment interface scattering is apparently negligible in this environment, the water–sediment interface

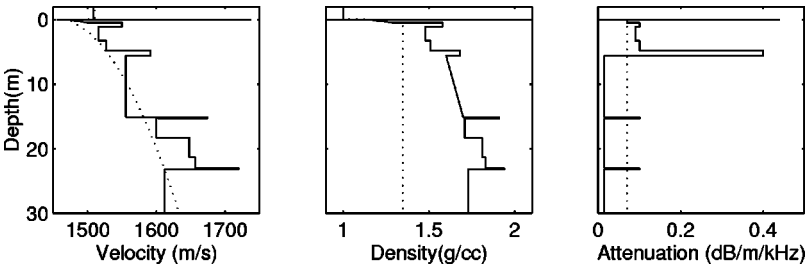


FIG. 17. Site 2 sediment parameters. Parameters from Holland and Osler (2000) are given in the solid line and those used for the Fig. 18 model predictions in the dotted line.

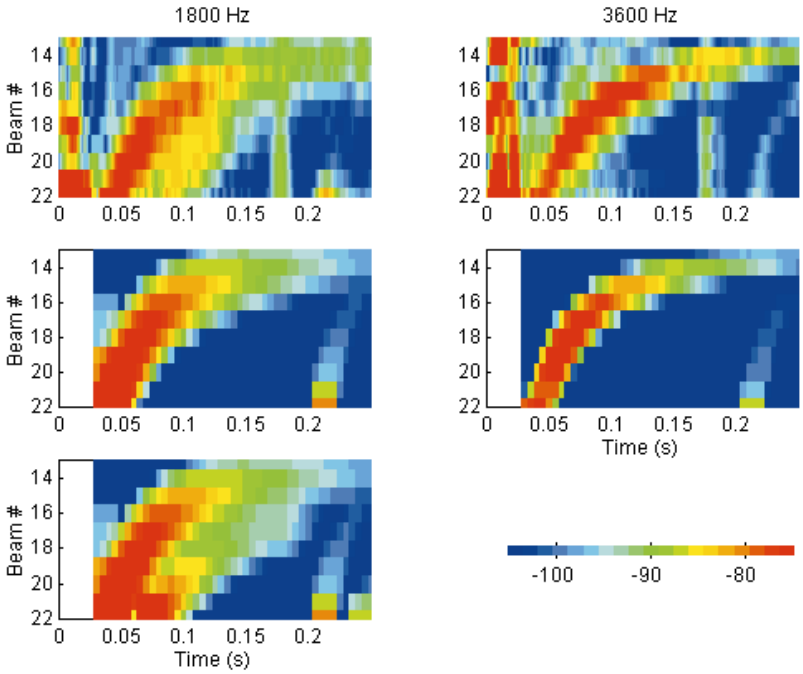


FIG. 18. Comparison of measured data at: (a) 1800 Hz and (b) 3600 Hz with modeled results for water–sediment interface scattering (c) 1800 Hz, (d) 3600 Hz, and (e) 1800 Hz modeled results with both water–sediment interface scattering and interface scattering at 25-m sub-bottom. The lettering sequence follows a left-to-right, top-to-bottom order.

model scattering results provide an indication of the *minimum* possible beam spread. Any broadening of this branch observed in the measurements can be interpreted as effects of deep (>10 m) sub-bottom volume scattering. The modeling results also give additional weight to the hypothesis of negligible interface scattering because such an unrealistically large interface roughness spectral strength parameter, 0.0025 m^4, is required to match the data. The other parameters used in the modeling were interface roughness spectral exponent of 3.25, layer (23 m sub-bottom) roughness spectral strength parameter[6] and exponent of 0.0075 m^2 and 3.25 respectively.

Measurement results at 1800 and 3600 Hz are shown in Figs. 18(a) and 18(b), respectively. Corresponding model results presented in Figs. 18(c) and 18(d) predict the scattered field for water–sediment interface scattering only. At 3600 Hz, the predicted beam spread [Fig. 18(d)], or width of the monostatic path *a*, is similar to that observed in the data [Fig. 18(b)]. This result indicates that the scattering at 3600 Hz occurs in the upper 10 m (essentially the pulse resolution) of the sediment column.

However, at 1800 Hz [Fig. 18(a)], the 23-m sub-bottom layer discussed previously is clearly visible in the data (starting at 0.08 s in the highest/steepest beam), but not in the interface scattering model result, Fig. 18(c). Figure 18(e) shows model predictions at 1800 Hz including this deeper layer (i.e., the layer at 23-m depth in Fig. 17), which in comparison with the measurements in Fig. 18(a) confirms its origins. There are differences between the model predictions and the observations that provide further clues about the nature of the scattering.

The main difference is that the modeled sub-bottom scattering arrival at 1800 Hz decays more rapidly with decreasing grazing angle (i.e., decreasing beam number) than does the measured result. This is probably due to inaccurate assumptions about the angular dependence of the scattering function but could also be due to the fact that the model predictions do not treat the detailed layering structure, e.g., layers at 0.5 and 5 m sub-bottom. Scattering from the sub-bottom layer at 23 m is not observed at 3600 Hz [Fig. 18(b)] presumably due to increased attenuation and/or reduced transmissivity through the overlying sedimentary layers.

C. Site 3 environmental description

Site 3 lies in 104 m of water on the Elba Ridge, a completely different physiographic province than site 2, being separated from the Tuscan shelf by the Elba Submarine Valley. The ridge/valley morphology effectively restricts any inflow of the fine-grained sediment deposition processes that shaped site 2. The Elba Ridge is a complex structure of magmatic origins, overlain by relatively coarse-grained material. Bathymetry data showed a small (approximately 0.5-degree) slope in the east–west direction. Sound-speed data from a CTD cast are shown in Fig. 11.

Wide-angle reflection measurements at this site are shown in Fig. 19. From these measurements, interval velocity and thickness for each layer can be extracted based on the separation and shape of each arrival [see Holland and Osler (2000)] for details on experiment geometry and processing. Analyses of these data indicated a 15-m layer with an interval velocity of 1700 and a 5-m layer with a 2100 m/s interval velocity. A 30 cm core was collected at this site. Granulometry analyses indicated a fine sand with shell fragments throughout the core. Core sound-speed measurements ranged from 1580–1600 m/s.

Figure 20 shows the bottom reflection loss data at this site. The salient feature of these data is the frequency dependence of the critical angle which is a function of the sound speed profile and attenuation in the bottom. At the highest frequency where a critical angle can be discerned (3150 Hz), the critical angle indicates an effective velocity of 1571 m/s, which compares very favorably to the core measurements in

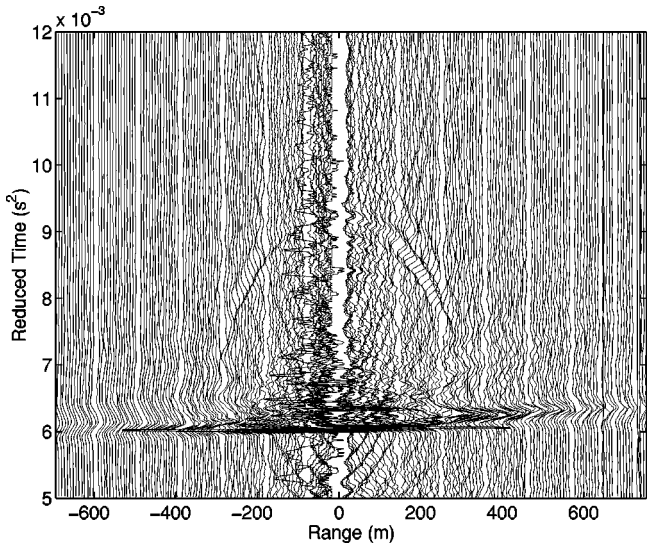


FIG. 19. Wide-angle reflection measurements at site 2. A reducing velocity has been applied to the data to flatten out the arrival structure. At zero range offset the water-sediment interface reflection is observed at 6 s^2 and the base of the 15-m sand layer at 9 s^2.

the upper 15 cm. At lower frequencies the observed critical angle increases, i.e., the effective velocity increases. At 1250 Hz the effective velocity is about 1630 m/s (critical angle of 22 deg) and at 600 Hz the effective velocity is about 1665 m/s (critical angle of 25 deg).

A single-layer geoacoustic model (see dotted line, Fig. 21) was constructed based on the core data, the interval velocity data, and the frequency dependence of the critical angle. The agreement with the measured data in Fig. 22 is reasonable, but the model does not predict the sharp rise in loss for example, at 2000 Hz near 22 deg. A two-layer model (see solid line, Fig. 21) was developed that matched the measured reflection loss in Fig. 22 more closely.

D. Site 3 measurements

Figure 23 shows the measured data at 1800 and 3600 Hz using the ITC array at a source receiver depth of 79 and 75 m, respectively. The surface reflection is observed through all beams at about 0.10 s. As with site 2, there is no evidence of scattering from biologics in the water column. The relationship between time and angle at the water–sediment interface is shown in Fig. B1.

The modeling for this site follows the same approach as for site 2; i.e., interface scattering is assumed as a means for determining if the beam-time series show direct evidence of deep (>10 m) sub-bottom scattering. The parameters used in the modeling were: interface roughness spectral strength parameter of 0.00125 m^4 and a spectral exponent of 3.25. The measured monostatic bottom scattering branch is shown in more detail in Figs. 24(a), 24(b) at 1800 and 3600 Hz. Model results presented in Figs. 24(c), 24(d) predict the scattered field for water–sediment interface scattering. Overall, the model results show similar trends as the measured data. At 1800 and 3600 Hz, the predicted beam spread (or width of the monostatic path a) is similar to that observed in the data. This indicates that the scattering arises from within approximately the first 10 m of sediment (see discussion in Sec. II A).

As a final comment on the importance of correctly discerning the scattering mechanism, our current hypothesis is that the statistical characteristics of reverberation are related to the scattering mechanism. For example, in regions where the scattering is dominated by sub-bottom volume contributions (typically characterized by a large number of small scattering events), the reverberation statistics should be Rayleigh distributed (because of the central limit theorem). In regions where the scattering is dominated by interface scattering (especially large-scale roughness), the reverberation may be non-Rayleigh, i.e., relatively few, larger scattering events. Abraham and Holland (1998) showed that the low-

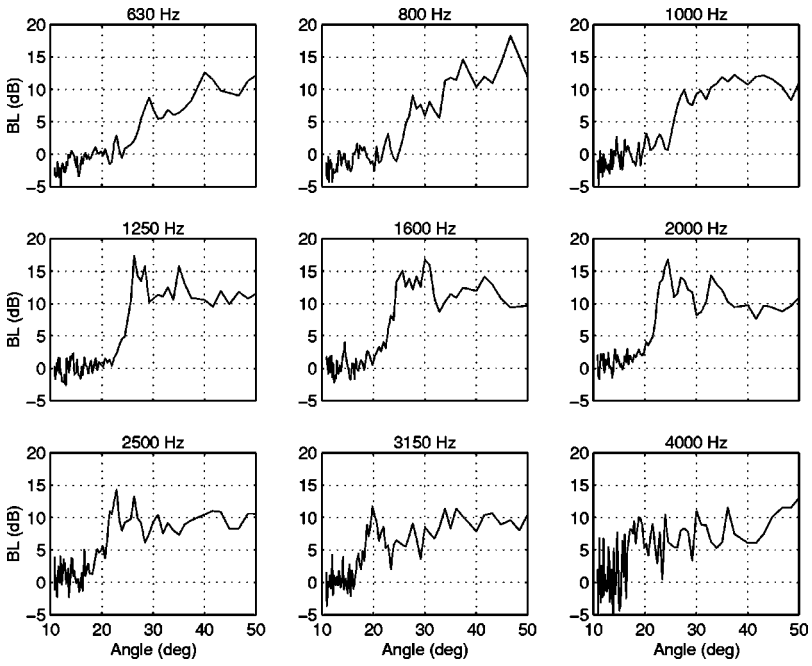


FIG. 20. Measured bottom reflection loss at site 3. Note that the critical angle is frequency dependent.

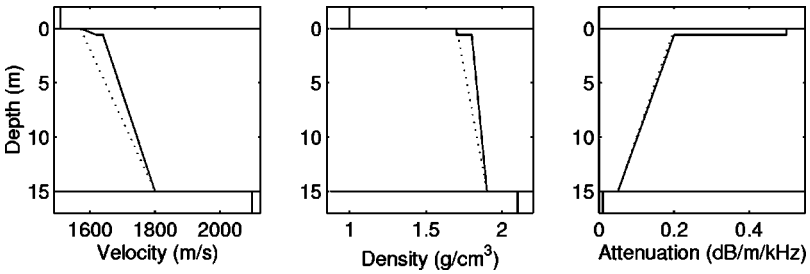


FIG. 21. Geoacoustic models at site 3: one-layer model (dotted line), two-layer model (solid line).

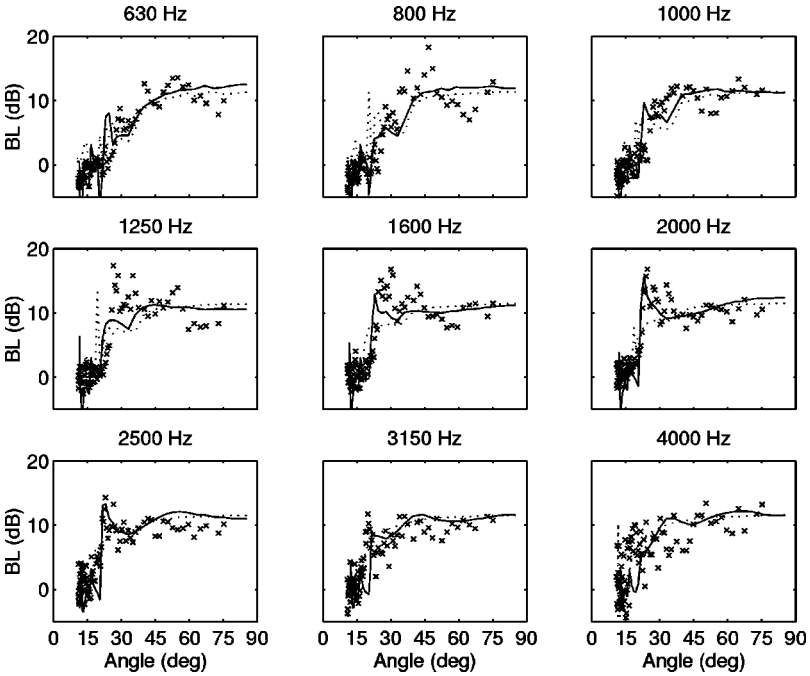


FIG. 22. Measured (*) bottom reflection loss at site 3 with model predictions from one-layer model (dotted) and two-layer sediment model (solid). The two-layer model provides a better fit with the data.

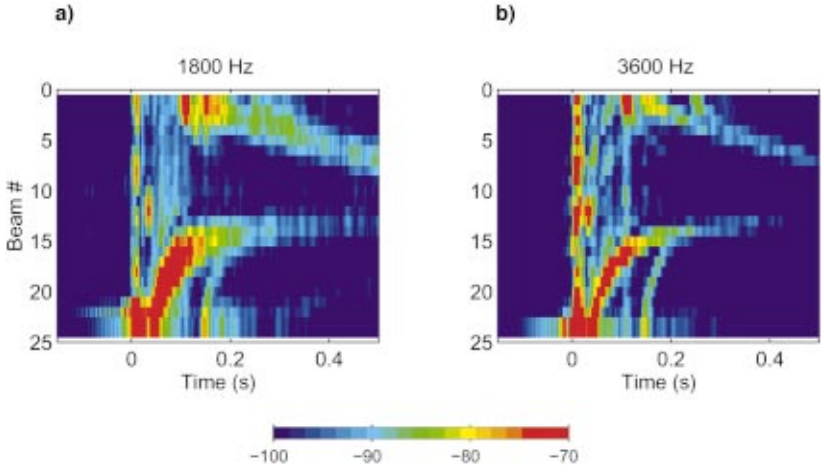


FIG. 23. Site 3 beam-time series (ten averages) using two sources at (a) 1800 Hz and (b) 3600 Hz. The quantity plotted is the ratio of received level to source level in dB.

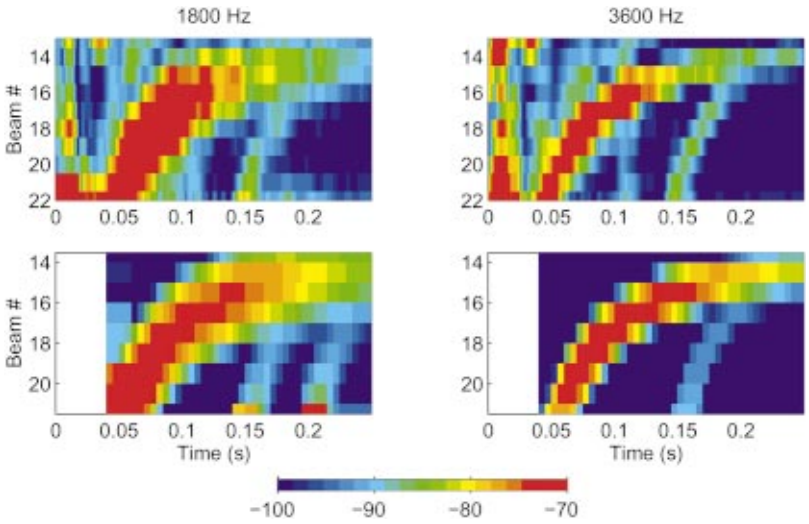


FIG. 24. Site 3 beam measured time series and model results for water–sediment interface scattering only (c) 1800 and (d) 3600 Hz. The lettering sequence follows a left-to-right, top-to-bottom order. At this site, no sub-bottom scattering contribution is apparent below 10 m.

frequency reverberation around site 2 (sub-bottom scattering) was dominated by Rayleigh statistics, while the statistics at site 3 (interface scattering) were strongly non-Rayleigh, thus appearing to confirm the hypothesis.

E. Site 2 and site 3 bottom scattering strength

Scattering strength results are shown in Fig. 25. Detailed quantitative analyses of these data will be undertaken at a later time. Rather, it is the intent here to present the data, and provide some observations on the frequency and angular dependencies in light of the knowledge already obtained from analysis of the beam-time series data.

Site 2 bottom scattering strength data (red curve) are strongly dependent upon frequency. The monotonically in creasing frequency dependence is perhaps due to the fact that the mean scatterer (sub-bottom shell and coral fragments) dimensions are smaller than a wavelength. The observed angular dependence is flatter than Lambert's law, particularly below 3600 Hz; this may also be due to sub-bottom scattering. Various authors [e.g., Novarini and Caruthers (1998), Holland and Neumann (1998)] have shown flatter than Lambert's law angular dependence in model and/or data results associated with sub-bottom scattering. However, the angular dependence of the observed scattering will depend heavily on the scattering kernel, the sediment sound speed profile between the water–sediment interface and the scatterers, and the source/receiver beam pattern.[7] Overall scattering strengths are less than or equal to the coefficient used by Mackenzie (1961) whose oft-cited frequency independent measurements were conducted in a deep-water silty-clay environment.

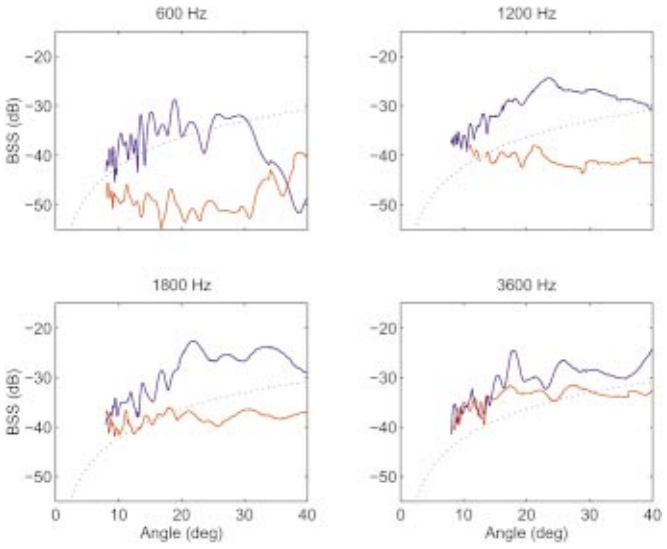


FIG. 25. Scattering strength results at sites 2 (red) and 3 (blue); also shown is Lambert's law with a coefficient of −27 dB (dotted line).

Scattering strength results at site 3 shown in Fig. 25 (blue curve) are roughly independent of frequency below 20° grazing angle and substantially larger than at site 2 for frequencies below 3600 Hz. Below 3600 Hz, the angular dependence is similar to Lambert's law at low angles but changes slope at intermediate angles. For example, at 1200 Hz, the slope of the scattering strength changes at 22 deg. The bottom loss data in Fig. 20 show that at 1200 Hz, there is a critical angle at 22 deg. A similar slope change and correlation with critical angles can be observed at 1800 and 600 Hz. The change of slope in the scattering strength is almost certainly due to a change of mechanism from interface scattering at low angles to sub-bottom volume scattering at intermediate angles.

IV. SUMMARY AND CONCLUSIONS

A direct-path shallow-water bottom scattering measurement technique has been presented. The principal advantage of the direct path measurement technique is that many of the assumptions that plague extraction of scattering strength from reverberation data are eliminated, so that the frequency and angular dependence of bottom scatter can be studied more precisely. Thus, the direct-path measurements can be used to examine the validity of the assumptions commonly applied to extract scattering strength from reverberation measurements.

Another important advantage of the direct-path technique is the ability to reveal the dominant scattering mechanism, i.e., to determine the relative contributions of water–sediment interface, sub-bottom volume or sub-bottom

horizon scattering. It is important to determine the scattering mechanism in order for physically based parameters to be extracted. The scattering mechanism may also be an invaluable indicator for predicting the general reverberation statistics.

The usual method for determining the scattering mechanism is based on inference from the geoacoustic parameters and/or the angular/frequency dependence of the data. We employed these approaches. However, we also demonstrated a significant new advance of measurement technique in which sub-bottom scattering can be directly observed. For example, at site 2, sub-bottom scattering from 23-m sub-bottom was apparent at 1800 Hz simply by inspection of the beam time series data. The scattering mechanism at this site was a series of thin, sandy shelly layers.

It was observed that even in shallow water, scattering can arise from relatively deep in the sediment (in this case 25 wavelengths). This has implications for both model and database development, inasmuch as in certain environments, the environmental description and modeling will have to consider commensurate sub-bottom depths.

ACKNOWLEDGMENTS

We gratefully acknowledge the North Atlantic Treaty Organization as sponsoring agency of this work. We also wish to acknowledge the captain, officers, crew, and the scientific crew aboard the NRV ALLIANCE (for the bottom scatter and environmental data) and RV MANNING (for the calibration). We express particular appreciation to Mr. Piero Boni, whose data acquisition skills and initiative were crucial to the success of the experiment.

APPENDIX A: SOURCE CALIBRATION

Calibration measurements were conducted on each of the source arrays using the geometry shown in Fig. A1 from a small workboat. The source array was placed at approximately 11 m below the surface and rotated in the horizontal plane by a stepper motor at 2.59 deg/s. Pulses of 15-ms duration were transmitted with a repetition rate of 0.863 Hz, yielding one sample every 3 deg. The receiver, a B&K 8101 hydrophone, was also set at a depth of approximately 11 m. The distance between the source and receiver was measured based on travel time for each deployment but was nominally 5 m.

The connectors used for the 2.5-m segments holding the source array were not completely rigid; thus, there was some play in the system, of the order of a few degrees. The effect of this play was minimized by rotating at a constant speed and checked by conducting the rotation in two directions. It was not possible to determine the orientation of the array relative to the receiver with any precision. Thus, the peaks of the theoretical beam pattern were used to fix the measurements in absolute angle.

The data were peak-level processed using a window from 10 to 14 ms which eliminated contamination from source turn-on/turn-off transients and surface path corruption at about 15 ms after the pulse initiation (see Fig. A2).

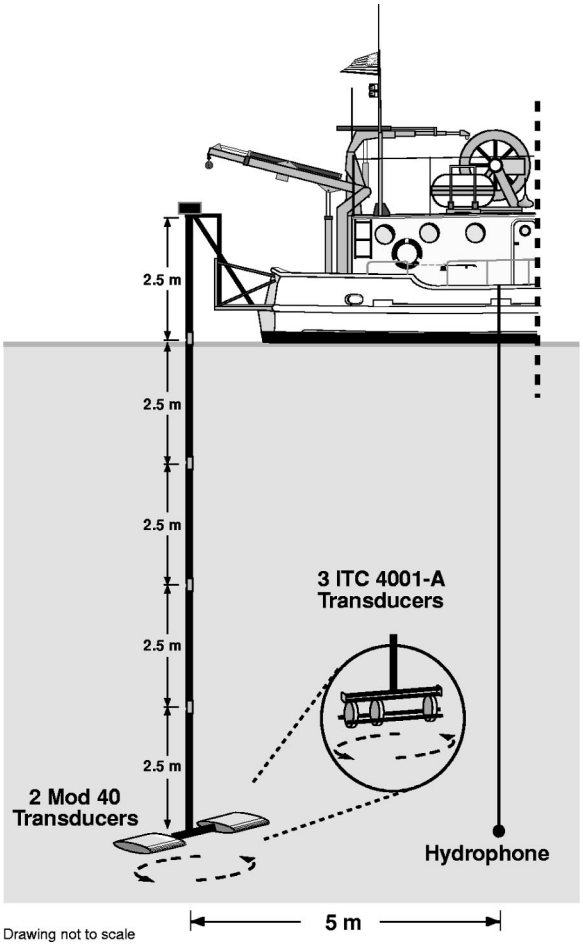


FIG. A1. Experimental setup for beam-pattern measurements. The source rotation (i.e., measurement) plane is in the horizontal.

In Fig. A3 the beam pattern measurements from the Mod40 array are provided at 600 and 800 Hz. The solid line is the theoretical result for two omnidirectional point sources at a distance of 5 m. The overall agreement is excellent except near the nulls. The slight asymmetry observed in the nulls at 600 Hz is due to the fact that the amplitudes of the two sources are slightly different.

Figure A4 provides the beam-pattern measurements and theory from the ITC array at 1200, 1800, and 3600 Hz. At these frequencies there are substantial differences between

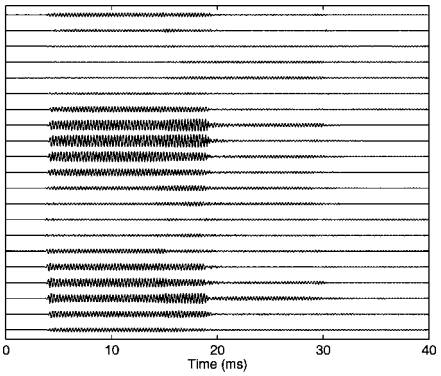


FIG. A2. Raw time series (18-deg increments) from beam-pattern measurements on ITC array at 3600 Hz. Note surface arrival apparent at about 15 ms.

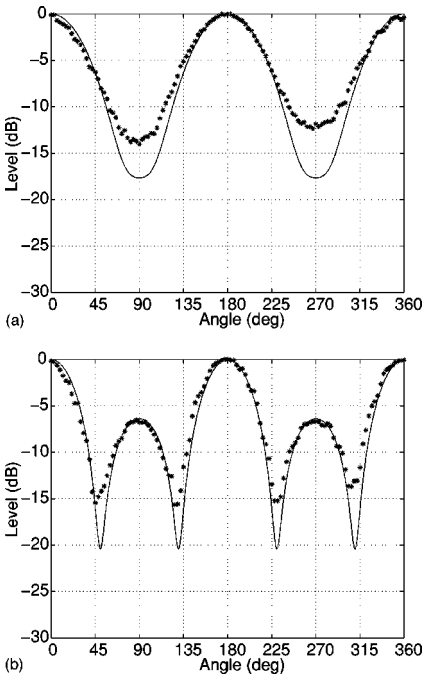


FIG. A3. Beam-pattern measurements (*) and theory (solid line) for Mod40 array at (a) 600 Hz, (b) 800 Hz.

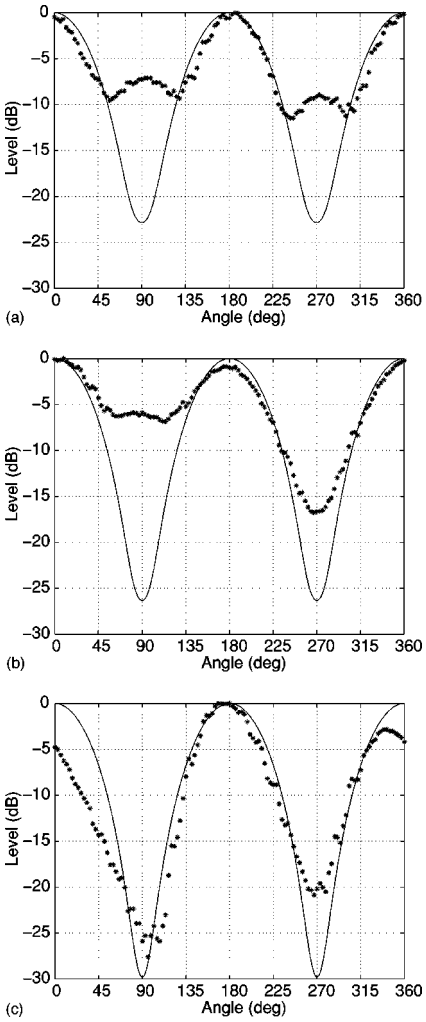


FIG. A4. Beam-pattern measurements (*) and theory (solid line) for ITC array at (a) 1200 Hz, (b) 1800 Hz, and (c) 3600 Hz.

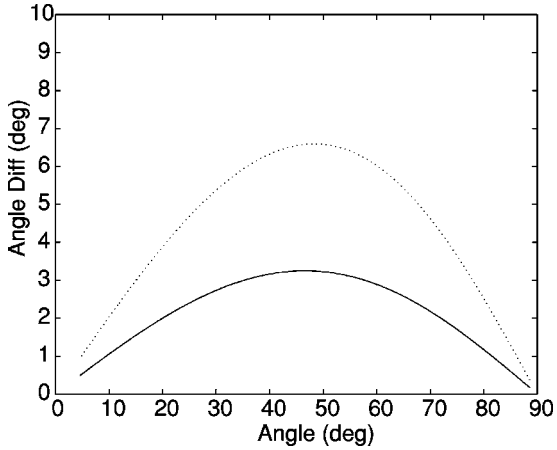


FIG. B1. Difference between incident and scattered angles for the Mod40 source array (dotted) and for the ITC array (solid). Source height is 20 m above the seafloor.

the theory and the measurements. Those differences are primarily due to the presence of a third transducer in the array (see sketch in Fig. A1). A secondary effect is that the individual transducers are not omnidirectional. The measured beam patterns in Figs. A3 and A4 were used in the data processing (to extract bottom scattering strength) and in the modeling.

APPENDIX B: MEASUREMENT BISTATIC ANGLES

Figure B1 shows the difference in incident and scattered angles due to the vertical offset between source and the two receive arrays. Note that the bistatic angle is maximum at 45 deg and tends toward zero at 0 and 90 degrees. For the results presented in this paper, the reported angle is the angle corresponding to the mean source–receiver height. Figure B2 shows the relationship between time and mean angle for the ITC array at sites 2 and 3.

The half-angle approximation (computing the monostatic scattering strength at the mean of the incident and scattered angles) is valid when the scattering function is linear between the two angles. A Lambert scattering function, for

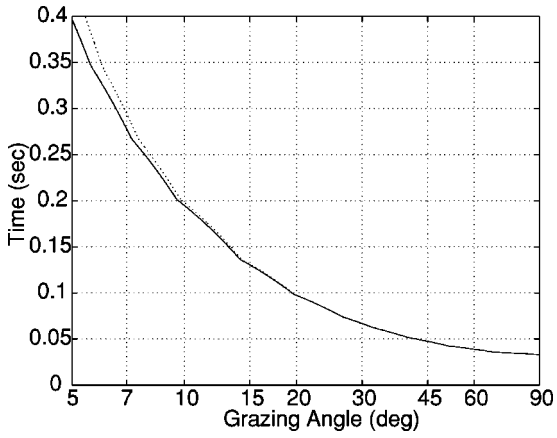


FIG. B2. Relationship between mean grazing angle at water–sediment interface angle and time for geometry at site 2 (solid line) and site 3 (dotted line).

example, is linear around 45 deg, satisfying the half-angle requirement.

APPENDIX C: MODEL FOR BOTTOM SCATTERING

The Holland and Neumann (1998) model departs from most other bottom-scatter models in that it computes received level in the time domain from the sediment/basement interfaces and a distribution of sub-bottom scatterers, then processes the data for scattering strength using the same assumptions employed in the data processing. The importance of the approach is that it provides a result that is directly comparable with measured data. Other sub-bottom scattering models (e.g., that compute the scattering strength in the frequency domain) do not account for the time-angle artifacts that exist in measured data when sub-bottom scattering plays a significant role. These artifacts exist because typical bottom-scatter data processing corrects for insonified area at the water–sediment interface and not for insonified volume or insonified area at a sub-bottom layer horizon.

Another important aspect of the model that separates is that it treats realistic sound speed profiles. Many other current models assume either isovelocity sediment sound speed profiles [e.g., Yamomoto (1996) or Novarini and Caruthers (1998)] or use an unrealistic sound-speed profile [e.g., Mourad and Jackson (1996)].

In this paper, the scattering is computed from the water–sediment interface and a sub-bottom layer interface. The received level for each beam is computed as a sum of delayed, scaled, replicas of the transmit source intensity $S(t)$

$$\mathrm{RL}(t;\theta_m) = 10\log\left[\sum_{k=1}^{2}\sum_{i=1}^{N} S(t-t_i)10^{\mathrm{TL}_{i,k}^{\mathrm{in}}/10}10^{\mathrm{TL}_{i,k}^{\mathrm{out}}/10} \times \sigma_k(\theta_{i,k}^{\mathrm{in}},\theta_{i,k}^{\mathrm{out}})A\right], \quad \text{(C1)}$$

where k is the index for the interface, $\mathrm{TL}_i^{\mathrm{in}}$ and $\mathrm{TL}_i^{\mathrm{out}}$ are the transmission losses (including the beam pattern) from the source to scattering cell i and from cell i back to the receiver respectively, σ is the scattering kernel (using the composite roughness theory of Jackson *et al.* (1986) for both interfaces), and A is the area of the cell. For the model used in this study, ray theory was used to compute the TL terms and the grid size was set to 4 m^2. Contributions from the direct blast and the normal incidence air–sea reflection were not included in the calculations. Source levels and beam patterns for the source and receiver were computed based on the calibration results (see Appendix A) and on an ideal Hanning shaded array, respectively.

[1]In addition to the disadvantages posed by the following assumptions, the method is not well suited for probing the angular dependence of bottom scattering.

[2]That is, i.e., the incident and scattered angles are different in the vertical plane.

[3]The sound-velocity profile effects both the TL and the mapping of time to angle. The difference between the TL from the measured profiles and that used in Eq. (1) (r^{-4}) is negligible. The mapping of time to angle is done via eigenray analysis for each profile.

[4]Based on the position of the sandy shelly layer observed in the less disturbed 1-m gravity core, it is presumed that the upper 0.2 m of the piston core have been lost and/or compacted.

[5]The layer is also observed in the seismic data of Fig. 14, and may be of similar constituency as the shelly sand layers observed at 0.5- and 5-m depth.

[6]As with the sediment interface roughness parameters, the layer roughness parameters are physically unrealistic because the scattering almost certainly arises from scattering from shells embedded in the layer rather than from layer roughness.

[7]Holland and Neumann (1998) showed that processing artifacts could artificially flatten the angular dependence when the scattering arises from the sub-bottom. This happens because while the processing generally corrects for the beam pattern at the water–sediment interface, the scattering actually arises from higher (i.e., less beam loss in this experiment) angles.

Abraham, D. A., and Holland, C. W. (**1998**). ''Statistical Analysis of Low-frequency Active Sonar Reverberation in Shallow Water,'' in *Proceedings of the Fourth European Conference on Underwater Acoustics* (CNR-IDAC, Rome), Ed Alippi and Cannelli, pp. 95–100.

Blanc, S., Novarini, J. C., and Nunez, A. L. (**1978**). ''Bottom reverberation-derive scattering strengths in shallow water off the Argentinian coast,'' J. Acoust. Soc. Am. **63**, 1342–1346.

Cable, P. G., Frech, K. D., O'Connor, J. C., and Steele, J. M. (**1997**). ''Reverberation-derived shallow-water bottom scattering strength,'' IEEE J. Ocean Eng. **22**, 534–540.

Chapman, R. P., and Harris, J. H. (**1962**). ''Surface backscattering strengths measured with explosive sources,'' J. Acoust. Soc. Am. **34**, 1592–1597.

Cole, B. F., and Podeswa, E. M. (**1974**). ''Shallow water reverberation under downward refraction conditions,'' J. Acoust. Soc. Am. **56**, 374–377.

Ellis, D. D. (**1995**). ''A shallow water normal mode reverberation model,'' J. Acoust. Soc. Am. **97**, 2804–2814.

Heard, G. J., McDonald, M., Chapman, N. R., and Jashke, L. (**1997**). ''Underwater lightbulb implosions: A useful acoustic source,'' Proc. IEEE Oceans '97 (unpublished).

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

Holland, C. W., and Neumann, P. (**1998**). ''Sub-bottom scattering: A modeling approach,'' J. Acoust. Soc. Am. **104**, 1363–1373.

Jackson, D. R., Winebrenner, D. P., and Ishimaru, A. (**1986**). ''Application of the composite roughness model to high frequency bottom backscattering,'' J. Acoust. Soc. Am. **79**, 1410–1422.

Mackenzie, K. V. (**1961**). ''Bottom reverberation for 530 and 1030 cps sound in deep water,'' J. Acoust. Soc. Am. **33**, 1498–1504.

McDonald, B. E., and Holland, C. W. (**2000**). ''Decay of large underwater bubble oscillations,'' J. Acoust. Soc. Am. **107**, 3084–3088.

Mourad, P. D., and Jackson, D. R. (**1996**). ''A model/data comparison for low frequency bottom backscatter,'' J. Acoust. Soc. Am. **94**, 344–358.

Novarini, J. C., and Caruthers, J. (**1998**). ''A simplified approach to backscattering from a rough seafloor with sediment inhomogeneities,'' IEEE J. Ocean Eng. **23**, 157–166.

Scanlon, G. A., Bourke, R. H., and Wilson, J. H. (**1996**). ''Estimation of bottom scattering strength from measured and modeled mid-frequency sonar reverberation levels,'' IEEE J. Ocean Eng. **21**, 440–450.

Sylva, P., Menard, P., and Roy, D. (**1986**). ''A reconfigurable real-time interpolation beamformer,'' IEEE J. Ocean Eng. **11**, 123–128.

Urick, R. J. (**1970**). ''Reverberation-derived scattering strength of the shallow sea bed,'' J. Acoust. Soc. Am. **48**, 392.

Yamamoto, T. (**1996**). ''Acoustic scattering in the ocean from velocity and density fluctuations in the sediment,'' J. Acoust. Soc. Am. **99**, 866–879.

Zhou, J., Dinghua, G., Erchang, S., and Ensheng, L. (**1982**). ''Long range reverberation and bottom scattering strength in shallow water,'' Chin. J. Ac. **1**, 54–63.

Low-grazing-angle monostatic acoustic reverberation from rough and heterogeneous seafloors

Robert J. Greaves[a)]
Massachusetts Institute of Technology/Woods Hole Oceanographic Institution, Joint Program, Woods Hole, Massachusetts 02543

Ralph A. Stephen
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

(Received 26 September 1998; revised 26 May 2000; accepted 13 June 2000)

Numerical wavefield modeling, based on time domain finite difference solutions to the full elastic wave equation, is used to quantitatively predict the sensitivity of monostatic backscatter to variations in geological properties of the seabed. This article addresses the hypothesis that observed backscatter signals from the seafloor are produced by a combination of seafloor (interface) and subseafloor (volume) scattering from structures having variations at scale lengths comparable to the wavelength of the insonifying acoustic field. Wavelength-scale seafloor roughness and subseafloor volume heterogeneity parameters are defined using stochastic spatial probability functions having Gaussian autocorrelations. The range of the variations in these parameters is constrained to realistic values based on estimates derived from seafloor bathymetry and other geologic data. Modeling results show that backscattering from rough, basaltic (hard) bottoms, in the absence of large-scale seafloor slope, is dominated by rough surface scattering with little contribution from volume scattering. Contrary to this, sediment (soft) bottoms with subseafloor volume heterogeneity, with or without seafloor roughness, produce significant backscattering signals compared to a homogeneous sediment bottom. © *2000 Acoustical Society of America.* [S0001-4966(00)03309-9]

PACS numbers: 43.30.Gv, 43.30.Hw, 43.30.Ft [DLB]

I. INTRODUCTION

Numerical wavefield modeling, based on the full elastic wave equation, is used to investigate and quantify the sensitivity of acoustic/elastic seafloor scattering to variations in wavelength-scale geological properties of the seabed. ''Seabed'' in this case refers to the seafloor and the subseafloor penetrated by energy from the incident wavefield. The modeling is motivated by low-grazing-angle (<20 degrees) monostatic acoustic reverberation experiments such as those performed in the ARSRP (Acoustic Reverberation Special Research Program) 1993 Acoustics Cruise in the Atlantic Natural Laboratory region of the Mid-Atlantic Ridge (ARSRP Initial Report, 1993). The quantitative interpretation of backscattered signals from actual ARSRP data in terms of seabed geological characteristics will be discussed in another paper. The goal of this study is to quantify the sensitivity of the backscattered signal to variations in correlation length and standard deviation of seafloor roughness and subseafloor volume heterogeneity that occur at wavelength scales. In order to maintain the focus on this goal we have not included regional seafloor dips, large-scale faults, nor subseafloor trends, such as vertical velocity gradients, in our numerical simulations.

We will show that geological variations in bathymetry and subbottom properties over distances comparable to the wavelength of the propagating sound cause scattering of sound from the seabed. For simplicity, ''wavelength scale'' is assumed to be from about $\frac{1}{10}\lambda$ to 10λ, where λ is the wavelength of sound in water. In this study the dominant wavelength in water of the incident and scattered field is assumed to be 6 m, corresponding to 250 Hz. The bandwidth of the linear frequency modulated acoustic source used in the ARSRP experiments was 200 to 255 Hz. Thus in describing variation in seafloor structure, ''wavelength scale'' would include all features from about 0.6 to 60 m. On the other hand, for subseafloor heterogeneity, where the velocities are higher, this wavelength scale definition will include variations over significantly greater distances. Also, by this definition, ''small scale'' then refers to all features or variations smaller than $\frac{1}{10}\lambda$ and ''large scale'' to those larger than 10λ. As described in Fig. 1 of Goff and Jordan (1988), seafloor bathymetry that varies over the range of this scale can be functionally modeled by stochastic spatial distributions. In our study, stochastic distributions are used to model both bathymetry and subseafloor heterogeneity that occur at wavelength scales.

Greaves and Stephen (1997) analyzed monostatic acoustic backscatter data from the Atlantic Natural Laboratory and concluded that ''observed variations in scattering strength are larger than can be explained by data error or seafloor dip,'' and suggested that ''other characteristics of steeply dipping areas, such as subsurface properties or smaller scale surface features, strongly affect the level of backscattered signals.'' One purpose of the study reported here is to determine if wavelength-scale roughness of the seafloor and heterogeneity of the subseafloor can produce variations in backscattered signal intensity that are of sufficient magnitude to explain the strong, fine-scale variations observed in monostatic low-grazing-angle reverberation data. This would be

[a)]Present address: MIT, Earth Resources Laboratory, 42 Carleton St., Cambridge, MA 02142.

consistent with the concept that the maximum energy level of the scattered wavefield, in nonspecular directions, occurs when the wave number ($k=2\pi/\lambda$) multiplied by the scale dimension of the interface or volume heterogeneity approaches unity. Frankel and Clayton (1986) and Dougherty and Stephen (1988) showed, using finite-difference modeling, that this peak in scattering does, in fact, occur in the vicinity of $ka=1$, where a is the correlation length of volume heterogeneity. Scattering from small-scale and large-scale heterogeneity is much less significant (Aki and Richards, 1980, pp. 748–750).

In this study, we use a time-domain finite-difference numerical wavefield simulation method to generate the scattered wavefield resulting from a variety of seabed models. Within the geophysics research community, this is considered an appropriate numerical modeling approach for wavelength-scale heterogeneity (Aki and Richards, 1980, p. 749). The approach has also been applied to geoacoustic and sea surface scattering problems (Stephen 1990, 1996; Hastings *et al.*, 1997, for example) and engineering acoustics (Schneider and Shlager, 1997, for example). Since we are simulating full elastic wave propagation, the numerical simulation includes all possible body and interface waves (in 2D), as well as multiple scattering interactions, for geological models with spatially variable elastic properties.

II. BACKGROUND

Scattering of low-grazing-angle sound waves from a very rough seafloor has traditionally been estimated by Lambert's rule which states that the intensity of the backscattered (monostatic) field from a unit seafloor area is proportional to the intensity of the incident wavefield and the square of the sine of the grazing angle (Jensen *et al.*, 1994). An example of this type of interpretation for ARSRP data is given by Caruthers *et al.* (1997). They primarily attribute variations in scattering to variations in local grazing angle which can, in turn, be defined as a function of local seafloor slope. Smaller scale structure and geological variability are qualitatively accounted for by variation of the Lambert coefficient. However, the distance over which the slope is estimated in such interpretations is usually much larger than the incident wavelength, and consequently the characteristics of the seafloor that would most strongly affect the scattered field are not quantitatively accounted for. The weakness of using Lambert's rule to predict scattering is that it is based on phenomenological evidence derived from scattering studies but is not based on any of the physical mechanisms that produce scattered signals.

In a review of long-range acoustic reverberation, Baggeroer and Dyer (1986) cite some examples that suggest backscattering is a more complex process than can be explained or predicted by just grazing angle and insonification intensity. In one example they show that a very deep ridge in the Arctic Ocean, assumed to be below the main channel of sound propagation, produces backscatter signal as strong as a much shallower ridge, even though the deeper ridge is insonified by a much weaker signal. They also point out that there are cases of high backscatter intensities that do not correlate with any high standing seafloor features. This suggests that strong scattering can be caused by seafloor characteristics that do not require the presence of large-scale structural features.

The variation of low-grazing-angle acoustic reverberation signals in terms of wavelength-scale seafloor characteristics is not well understood. Dyer *et al.* (1993) proposed that the backscatter is produced by scattering from seafloor facets that are within a few degrees of being normal to the incident wavefield. In this theory, facets that are at least a wavelength in size will be ''self-selected'' as the dominant source of backscattered signals. In essence, this is similar to the idea that scattering is dominated by wavelength-scale features, except for the added emphasis on facet orientation. Makris *et al.* (1995) suggest that the strongest reverberation signals are returned from large scale scarps on seafloor ridges that face the source direction but do not focus on the importance of wavelength scale features.

In an effort to provide a more physically satisfying explanation of the signal variations observed in acoustic reverberation data, wave equation modeling has been used to demonstrate the scattering of low-grazing-angle incident wavefields from the seafloor. Stephen and Dougherty (1993) compared scattering from rough seafloor defined with elastic versus acoustic properties. They also demonstrated the effects of sediment cover over a rough basalt basement. Their results showed that it is important to treat the seafloor as a fluid/elastic boundary since shear body waves and interface waves, and the rescattering of these waves back into the water column, are important scattering mechanisms that are not included in acoustic models (e.g., Kirchhoff modeling does not include shear waves since the Kirchhoff integral does not include particular motion tangent to the wave front (Berkhout, 1987, p. 233). In particular, secondary scattering of Stoneley waves is an important source of energy in the backscattered wavefield. Stephen and Dougherty also show that homogeneous sediment cover significantly decreases the amount of energy that is scattered into the water column, suggesting that geological properties of the subseafloor may be predicted from the characteristics of the scattered signal. Robertsson and Levander (1995) investigated the effects of sediment velocity, attenuation, and thickness on bottom scattering and came to the same conclusion, i.e., that sediment cover, even thin sediment cover on the order of a wavelength in thickness, strongly reduces the backscattered signal. They also reconfirm that Stoneley waves interacting with the rough seafloor are an important source of scattered energy. Stephen and Swift (1994, Fig. 6) showed that scattering from a wavelength-scale facet on the seafloor is dominated by diffraction scattering and that specular reflection and transmission from a wavelength-scale facet is not an important component of the scattered field.

In order to thoroughly investigate the mechanisms of scattering from the seafloor, it is also necessary to consider scattering from volume heterogeneity in the subseafloor. Dougherty and Stephen (1988) identify the mechanisms of volume scattering and show that this scattering varies with the spatial scale (quantified as the correlation length of the velocity and density distributions) of the volume heterogeneity. Swift and Stephen (1994) apply the same techniques to

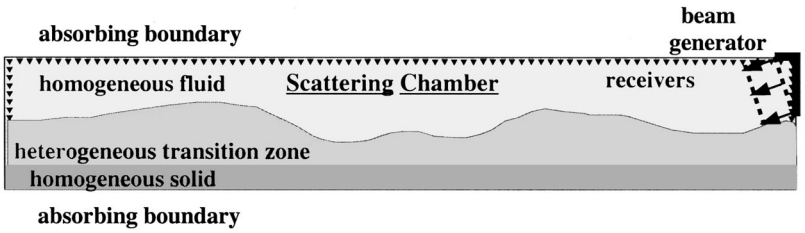


FIG. 1. The numerical scattering chamber (NSC) consists of an inner ''scattering chamber,'' in which a geological model is defined, surrounded by an absorbing boundary region. The scattering chamber includes a homogeneous fluid zone, a heterogeneous transition zone (seafloor model), and a homogeneous solid zone at its base. A Gaussian beam pulse is introduced into the scattering chamber in the upper right corner of the model space, propagating at a specified grazing angle. Receivers are located in the water column along the top and sides of the scattering chamber. Scatter direction in the NSC is measured as the counterclockwise grazing angle relative to the propagation direction.

investigate and quantify volume scattering, below flat seafloor models, when the incident wavefield is a Gaussian pulse-beam incident at a low grazing-angle. The Gaussian pulse-beam used as the source wavefield approximates the insonification in acoustic reverberation experiments. Their results suggest that volume scattering can be an important source of energy in the backscattered signal.

III. THE NUMERICAL SCATTERING CHAMBER

In this study, we use the numerical scattering chamber (NSC), described by Stephen and Swift (1994), to generate scattered wavefields from a variety of seabed models with well-defined seafloor roughness and subseafloor volume heterogeneity. The NSC is an elastic-wavefield-propagation forward-modeling program, designed specifically to address problems of seabed scattering at low grazing angles. The wavefield propagation algorithm is based on a second-order finite-difference formulation, with model parameters defined on a staggered grid, as described by Virieux (1986). The source wavefield in the NSC is a Gaussian pulse-beam that is incident at a particular low grazing angle (Fig. 1). In order to simulate a monostatic reverberation experiment, a beamforming process is applied to the scattered-field time-series data to calculate scattering coefficients as a function of scattering angle. The scattering coefficient in the monostatic backscatter direction is then compared as specific geologic parameters are varied.

The numerical wavefield can be ''recorded'' in two ways. One way is to store the particle displacements at all grid points at specified time steps. These ''snapshots'' of the time progression of the wavefield are particularly useful for gaining insight into the mechanisms of scattering, and they show the full variety of wave types that occur. Rather than displaying displacement vectors at each grid point, we display ''amplitude density,'' which is proportional to the square root of the energy density and maintains the sign of the amplitude wave field [see Dougherty and Stephen (1988)].

The other way that the propagating wavefield is recorded simulates ''receivers'' that would be used during actual data acquisition. In this case, the displacement or pressure at chosen grid points are recorded at every time step, producing a time series of the wavefield amplitude observed at each point in space. In our modeling, the pressure time series are recorded at grid locations along the sides and top of the scattering chamber, as indicated in Fig. 1. Only time series at grid locations in the water column (dependent on the shape of the model) are recorded in our modeling. Using the beamforming process defined by Stephen and Swift (1994) the time-series data are reduced to scattering coefficients as a function of scattering angle. Scattering angle is defined as the rotation angle relative to a horizontal vector pointing in the direction of the source. All scatter directions greater than 90 degrees correspond to forward scatter and less than 90 degrees to backward scatter. The scattering coefficient is the time average pressure field intensity, calculated as a function of propagation (scattering) angle, normalized by the intensity of the incident (reference) field. Before making this calculation, the calculated incident wavefield is removed from the pressure time series data at each receiver location.

The grid spacing is constrained by the need to minimize grid dispersion and yet maximize the speed of model computation. The grid spacing (Δx) was chosen as 0.4 m, which is $\frac{1}{15}$ of the source wavelength (6 m) at the peak frequency of pressure in water. Grid dispersion is always present to some extent in finite-difference modeling, but becomes problematic when the propagation path length is much greater than the wavelength. Kelly *et al.* (1976) suggest a ''rule of thumb'' for acceptable grid dispersion effects (for a second-order approximation to the wave equation) in homogeneous media is that the grid spacing should be at least ten grid points per wavelength, i.e., $\Delta x \leqslant \lambda_{\min}/10$. Since wavelength decreases with velocity, this rule of thumb suggests that the lowest velocity that should be included in a geological model is $v_{\min}=\lambda_{\min}\ f_p=10\Delta x f_p$, where f_p is the peak frequency of the source wavefield. In our modeling $f_p=250$ Hz, so that $v_{\min}=1$ km/s. Propagation over distances more than a wavelength in media with lower velocities will produce extended propagation wavelets due to the grid dispersion. However, in practice, when using heterogeneous models, we found that significantly lower velocity regions can be included without causing unacceptable grid dispersion effects. This is due to the fact that the waves propagate only very short distances in these low velocity regions. In our modeling we were able to achieve acceptable results as long as the smallest mean velocity (P or S wave velocity) was greater than about 0.5 km/s (i.e., five grid points per wavelength) and the velocity due to heterogeneity did not drop below 0.2 km/s anywhere within the models.

A finite difference calculation can also become numeri-

cally unstable if the time step chosen is too large. The stability criteria for propagation in homogeneous media using the second-order staggered grid scheme is (Virieux, 1986):

$$\Delta t \leqslant \frac{\Delta x}{\sqrt{2} v_{\max}}, \tag{3.1}$$

where $v_{\max}$ is the maximum compressional velocity in the model. For our modeling, Δt was chosen as 0.000 04 s (i.e., $\frac{1}{100}$ of a wave period at 250 Hz), which satisfies this stability criteria for $v_{\max} \leqslant 7$ km/s. However, sufficient stability conditions for heterogeneous media have not been defined (Stephen, 1990). Therefore, numerical stability in the modeling of heterogeneous media must be judged for each model computation. Within the range of heterogeneous models used in this study, the NSC proved to be numerically stable.

IV. MODELS WITH HOMOGENEOUS SUBSEAFLOOR

The scattered acoustic wavefield returned from the seabed contains products of both seafloor and subseafloor interaction with the incident and transmitted wavefield. In order to simplify the quantification of this scattered field in terms of seabed geological characteristics, we initially consider the problems of seafloor (interface roughness) scattering separately from subseafloor (volume heterogeneity) scattering.

The rough seafloor models are defined by stochastic height probability functions with the characteristic dimensionless parameters, ka and $k\sigma$, where $k = 2\pi/\lambda$ is the wave number of the incident field having wavelength λ, a is the (auto)correlation length, and σ is the standard deviation of the distribution. These spatial distributions are created by filtering normalized random distributions to stochastic probability distributions having Gaussian autocorrelation functions. The spatial distributions are then scaled such that their standard deviation in amplitude is the desired height, σ. This method of model generation is adapted from the work of Frankel and Clayton (1986) and Dougherty and Stephen (1988). Our method differs in that we found it more accurate to iterate the distribution filter parameters until the desired autocorrelation length was achieved.

Correlation length of a function having a Gaussian autocorrelation is defined as the lag distance at which the autocorrelation amplitude has decreased to $1/e$ of its peak value (Ogilvy, 1991, p. 14). This explicit definition of correlation length is a strong motivation for using Gaussian distributions in modeling where quantification of model characteristics is needed. Goff and Jordan (1988, Fig. 1) suggest that a broad range of scales of seafloor topography can be adequately represented by a stochastic spatial probability functions. They also state (p. 13 603) that seafloor asymmetry caused by such features as fault scarps and flat sediment ponds are not fully represented by Gaussian seafloor models. In our study such features are generally larger than what is considered ''wavelength scale'' and are, therefore, not included in the models. Our primary motivation for using Gaussian spatial distributions for the geologic models in this study is that such models can be easily generated with well-defined correlation lengths and standard deviation in amplitude. We recognize that Gaussian models may not be the optimal representation of natural phenomena, such as seabed spatial characteristics. In initial modeling (Greaves, 1998), scattering from Gaussian models was compared to scattering from Von Karman self-similar models [as defined by Frankel and Clayton (1986)] with equivalent spatial scales. The resulting scattering coefficient functions were almost identical, and certainly within the experimental error of the numerical results.

It is important to consider that each seafloor model is just a particular deterministic realization of all possible seafloor models having the same values for the parameters ka and $k\sigma$. In numerical modeling, different realizations are generated by initiating the model-building process with different random series. In our modeling we used three realizations of each model. It is not possible to distinguish or parametrize the ''randomness'' of each realization, so that the scattering from the three realizations is used to estimate a level of variation in scattered energy below which we cannot distinguish one realization from another. We refer to this as the ''realization error.'' Seabed property variations that generate signal variation at or below this level cannot be resolved in the scattering coefficient functions. In general, the scattering functions for the three realizations of each model remained within 1 or 2 dB.

In our geological model generation we constrained the spatial parameters to the ranges $0.5 \leqslant a \leqslant 6$ m for correlation length and $0.5 \leqslant \sigma \leqslant 6$ m for standard deviation of the rough seafloor bathymetry and $1 \leqslant a \leqslant 12$ m and $0 \leqslant \sigma \leqslant 50\%$ perturbation in velocity and density of volume heterogeneity. Note that for modeling with a 250-Hz sound signal, $k \approx 1$ (in water), so that $ka \approx a$ and $k\sigma \approx \sigma$ for rough seafloor. We consider these parameter constraints to be reasonable for seabed spatial property characteristics in regions of young oceanic crust.

For all of the homogeneous rough seafloor models, the upper half of the model space is defined as the water column, with compressional velocity $v_p^w = 1.5$ km/s, density $\rho^w = 1.0$ kg/m^3 and no shear velocity. The lower half of each model is a homogeneous solid basement with properties appropriate for basaltic (igneous) basement rock, $v_p^b = 4.0$ km/s, $v_s^b = 2.0$ km/s and $\rho^b = 2.5$ kg/m^3. Such seafloor is considered a ''hard bottom'' because $v_p^b \geqslant v_s^b \geqslant v_p^w$. In contrast, a ''soft bottom'' seafloor has sediment velocities (v_p^s, v_s^s) such that $v_p^s \geqslant v_p^w \geqslant v_s^s$, representing mud or ooze.

As shown in Fig. 1, the incident Gaussian pulse beam is introduced at a grazing angle of 15 degrees on the right side of the model space and propagates from right to left. The compressional and shear critical grazing angles for a model with these properties are, respectively,

$$\theta_c^p = \cos^{-1}\left(\frac{v_p^w}{v_p^b}\right) = 68 \text{ degrees} \tag{4.1}$$

and

$$\theta_c^s = \cos^{-1}\left(\frac{v_p^w}{v_s^b}\right) = 41 \text{ degrees}, \tag{4.2}$$

such that the 15 degree source wavefield is incident at a

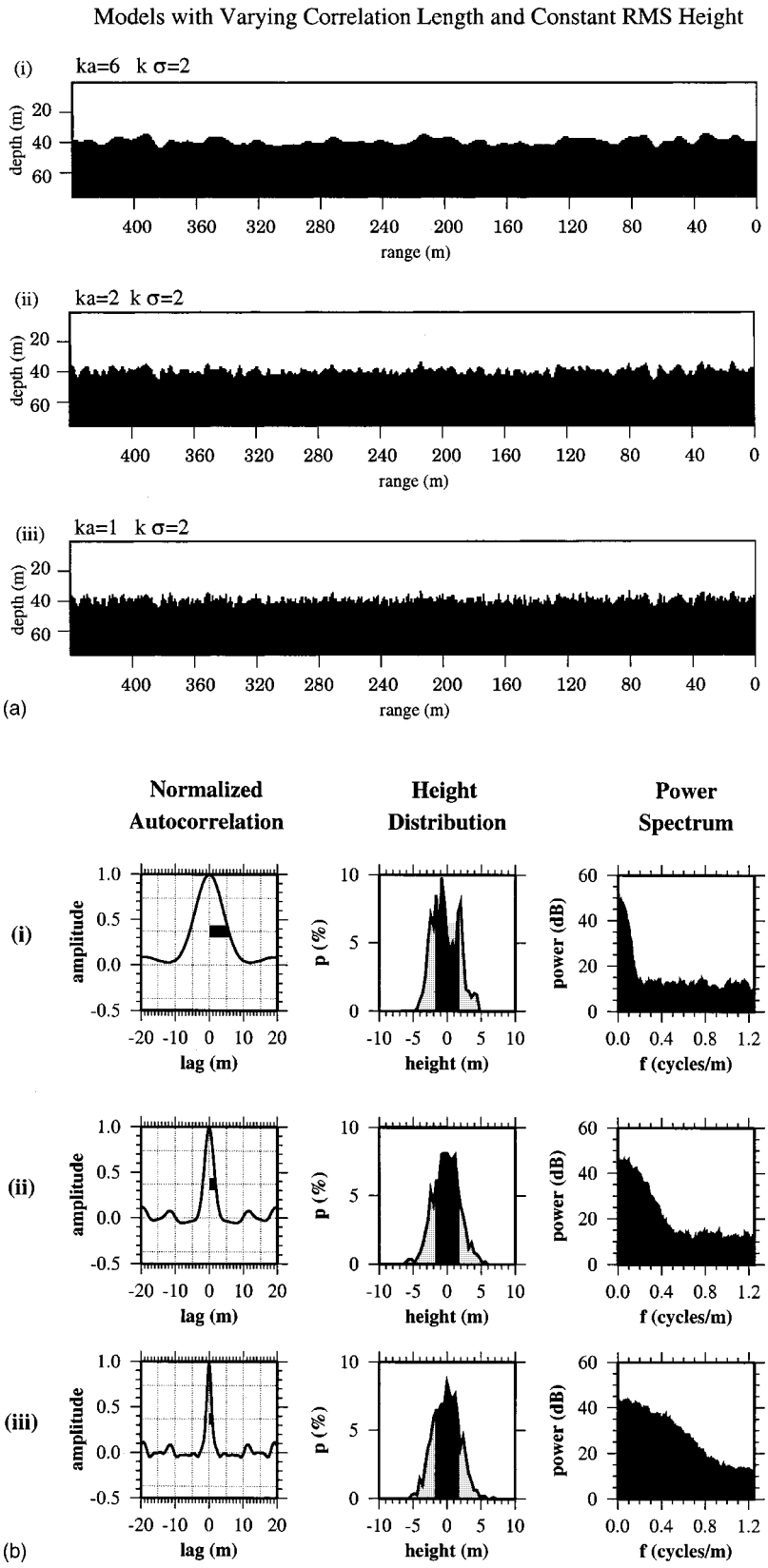


FIG. 2. (a) These models have horizontal rough seafloor with different correlation lengths. The seafloor models have Gaussian autocorrelations with correlation lengths: (*i*) $ka=6$, (*ii*) $ka=2$, and (*iii*) $ka=1$. The spatial distribution has been scaled such that the rms height of the distribution is $k\sigma=2$. As *ka* decreases, higher frequency variability is observed overriding the same low-frequency characteristics. Subseafloor velocity and density are $v_p^b=4.0$ km/s, $v_s^b=2.0$ km/s, and $\rho^b=2.5$ kg/m^3, simulating a basaltic seafloor. (b) The statistical properties of the seafloor models in (a) show that as the correlation length (black bar in autocorrelation plots) decreases, the corner frequency of the power spectrum increases. The models are defined by filtering a random series such that its autocorrelation function is Gaussian and then scaling the resulting random series to have the desired rms height. Although the height probability function is not exactly the same in each model, they do have exactly the same rms height and 68% of the height values fall within one standard deviation of the mean (black regions). (c) These energy density diagrams show the distribution of compressional and shear amplitude density at 160 ms (4000th time step). The incident beam wavefield (P_1) is reflected from the seafloor (P_1P_1) and has a direct wave root (DWR). The surface roughness in each model produces primary diffractions in the water column and diffracts energy into the subseafloor. Both diffracted compressional (P_1P_2) and diffracted shear (P_1S_2) waves are observed in the subseafloor. Surface (Stoneley) waves propagate in both the forward and backward directions and produce secondary diffractions that add to the total scattered wavefield in the water. Both compressional head waves, pHW, and shear head waves, sHW, are observed in the water column, as well as a shear head wave ($P_1P_2S_2$) in the solid. These wave types occur in each model but differ in detail.

subcritical grazing angle and is totally internally reflected on a smooth and horizontal seafloor. For such a seafloor, there is no transmission of either compressional or converted shear waves into the subseafloor. The only scattered field produced in the water column for such a model is the specular reflection. However, when the seafloor is rough, a wide variety of body and interface waves are generated in the seabed that produce scattered signals in the water column.

A. Scattering versus *ka*

Figure 2(a) shows a series of horizontal, rough-seafloor models with different correlation lengths but a constant stan-

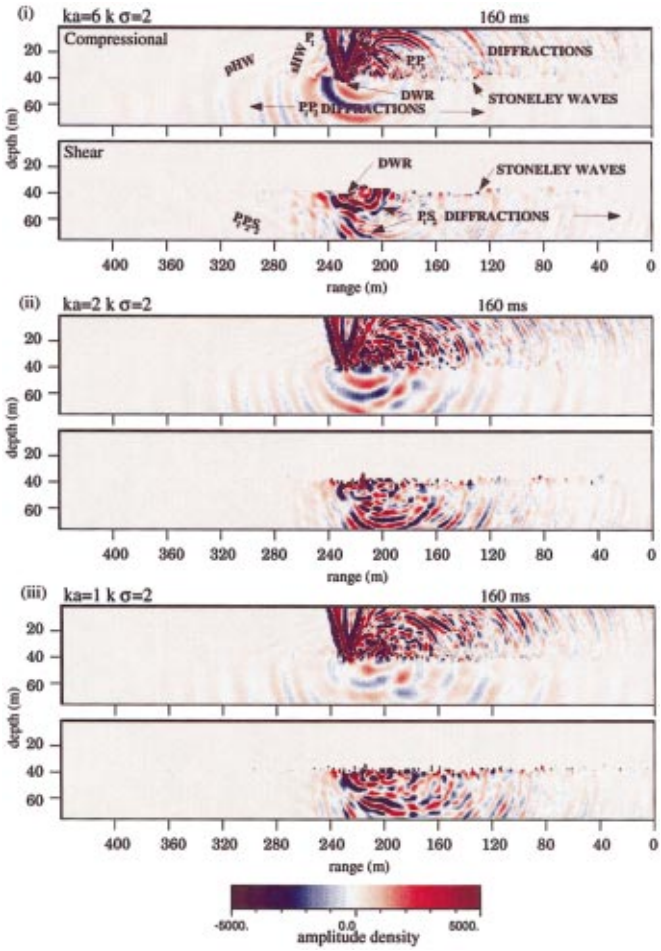


FIG. 2. (*Continued.*)

dard deviation (rms height for zero mean distributions) in height. The dimensionless parameters ka and $k\sigma$ are used to describe these surfaces. The homogeneous subseafloor is defined with the basaltic hard-bottom parameters. The seafloor models are produced by filtering an initial random series to produce models with specific correlation lengths and then scaling them to have a certain rms height. Note that as ka decreases, the low-frequency features are maintained while high-frequency variations are increased. The surface characteristics of these Gaussian models are illustrated by the plots of autocorrelation, height probability functions, and power spectra in Fig. 2(b). In each case, the rms height was scaled to $\sigma = 1.91$ m so that $k\sigma = 2$. Correlation lengths were varied from $ka = 0.5$ to $ka = 6$, but only distributions having $ka = 1$, 2, and 6 are shown in these plots. As correlation length decreases, power is distributed over a wider range of length scales, as indicated by the broadening of the spectra and lowering of its peak value. The details of the height probability function vary but always have 68% of height values falling within the standard deviation, as expected for Gaussian distributions.

The snapshots of wavefield amplitude density after 160 ms of propagation in the scattering chamber are shown in Fig. 2(c). As indicated in the figure, the energy represented in these diagrams corresponds to a variety of elastic and acoustic waves. The direct wave (P_1) is reflected (P_1P_1)

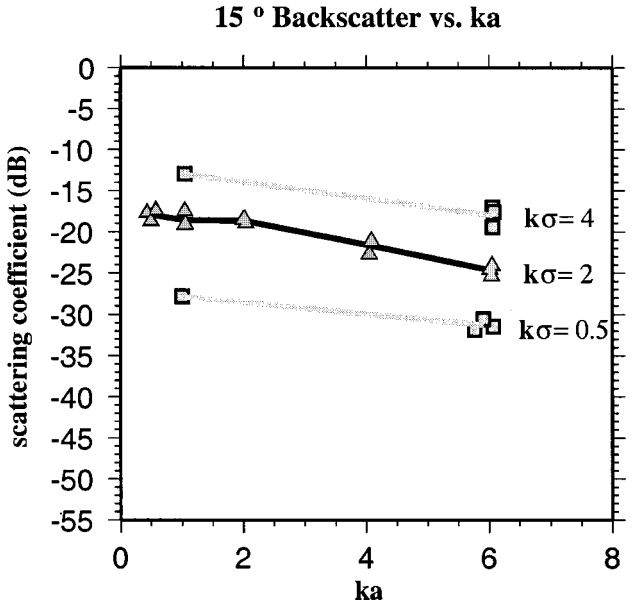


FIG. 3. Backscattering coefficients for horizontal, rough seafloor with different correlation lengths. In the monostatic backscatter direction (15 degrees) the variation in scattering coefficient as a function of ka shows an approximately linear decrease as ka increases. The black line corresponds to the results for $k\sigma = 2$ shown in (a). The gray lines correspond to results for models with the same ka but scaled such that $k\sigma = 4$ and $k\sigma = 0.5$. The symbols plot the values calculated for individual model realizations. The spread in the scattering coefficients for different realizations is about ± 1 dB.

from the seafloor as well as being diffracted from the surface roughness into both the water column and the seabed. Strong point diffractions occur where the half-width of a particular seafloor bathymetric feature is on the order of the incident wavelength (Dougherty and Stephen, 1987). Scattering at the seafloor also acts as the local source of interface waves that propagate in both the forward and backward directions. In this figure, Stoneley waves are clearly identifiable.

Although the grazing angle of the insonifying field at the seafloor is still subcritical, some of the incident energy is diffracted into the subseafloor. Wavefields propagate in the subseafloor as diffracted compressional body waves, diffracted shear body waves, Stoneley waves, and pseudo-Rayleigh waves. These are generated at wavelength scale roughness elements on the seafloor which act as secondary point sources. The diffracted compressional and shear waves, and the pseudo-Rayleigh waves can reradiate energy into the water column as head waves (pHW and sHW). All four wave types can radiate energy back into the water column by diffraction from other roughness elements. Another shear body wave ($P_1P_2S_2$) is generated by the interaction of the diffracted compressional wave with the interface. Pseudo-Rayleigh waves are possible for these models but are not positively identified in these figures because they are kinematically very close to the shear waves propagating along the interface. The total scattered wavefield in the water column is a combination of primary diffractions of the direct wave from seafloor features, internal reverberation within these features, and secondary diffractions of interface waves. A complete description of the types of acoustic and elastic propagation that are possible at a homogeneous liquid–solid boundary, insonified by a point source above the interface,

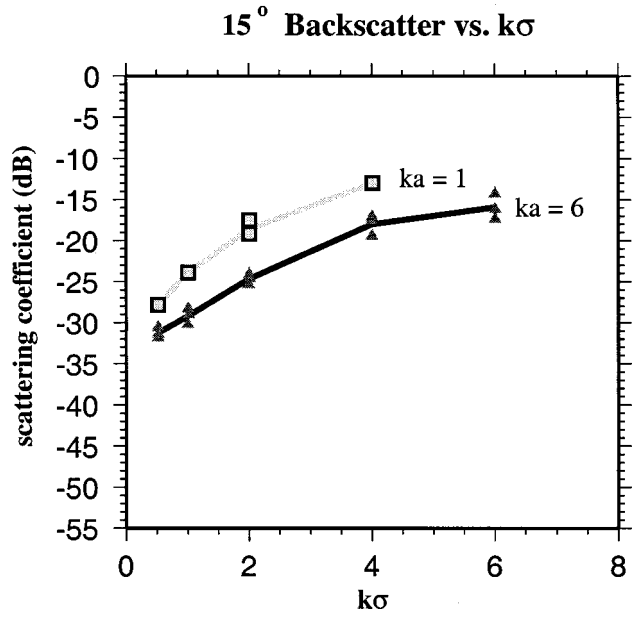


FIG. 4. Scattering coefficients for horizontal rough seafloor with different rms heights. The subseafloor is modeled with homogeneous basaltic properties. In the monostatic backscatter direction (15 degrees) the variation in scattering coefficient as a function of $k\sigma$ shows an increase in backscatter intensity as $k\sigma$ increases. The black line corresponds to the results for $ka=6$. The gray line and symbols correspond to results for models with $ka=1$ and show the same variation. The flattening of the trend suggests that there is some limit to the level of backscatter that is reached when $k\sigma$ becomes large.

is described in Brekhovskikh (1960). The excitation of diffracted energy from a single roughness element on a fluid–solid interface is summarized in Fig. 6 of Stephen and Swift (1994).

Visual comparison of the energy density diagrams suggests that as the value of *ka* decreases, the number of individual diffractions propagating in the water column increases. This suggests that the total scattered energy is increasing as the correlation length of the surface roughness decreases.

In Fig. 3 the scattering coefficients in the 15 degree (monostatic) backscatter direction are plotted for the surface roughness models. The realization error average is ± 1 dB. The results for $k\sigma=2$ show clearly that the scattering increases as *ka* decreases. Results from the models calculated with $k\sigma=4$ and $k\sigma=0.5$ show a similar trend. Between $ka=2$ and $ka=0.5$ the scattering coefficients are almost constant, suggesting that this region may be the peak of the backscattering, although we have no results for smaller values of *ka* to prove this. The grid spacing of 0.4 m does not allow definition of models with $ka<0.4$. Within the realization error, models within $0.5 \leqslant ka \leqslant 2$ cannot be clearly distinguished based on scattering coefficients. For $ka>2$, a linear approximation to the sensitivity suggests that the scattering coefficient decreases at a rate of about -1.5 dB per unit *ka*.

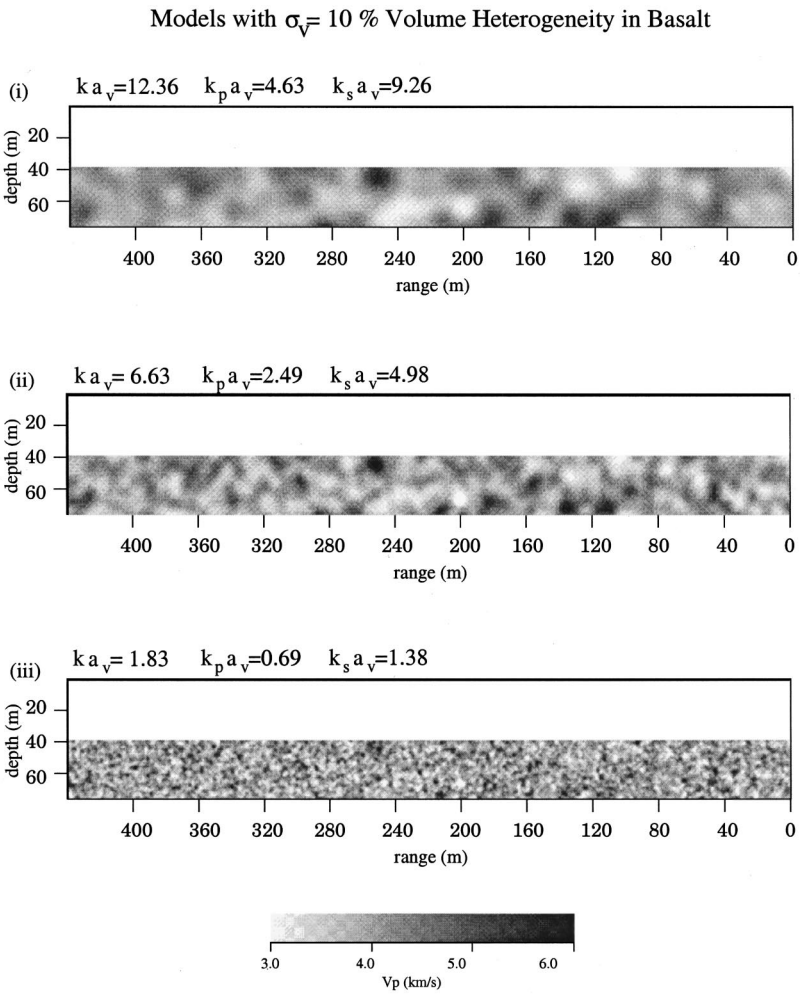


FIG. 5. Flat-seafloor models are shown with subseafloor volume heterogeneity. All of the models have volume heterogeneity with rms perturbation amplitude of 10% but have correlation lengths (i) $a_v=11.8$ m, (ii) $a_v=6.3$ m, and (iii) $a_v=1.7$ m. Average velocity and density of the subseafloor corresponds to the basalt (hard-bottom) model. When a source wavefield is incident on the seafloor at a subcritical grazing angle, the primary source of energy penetrating into the subseafloor will be the direct wave root.

B. Scattering versus $k\sigma$

In the next series of models, the correlation length is held constant while the seafloor rms height is varied. The 15-degree (monostatic) backscatter coefficients for all realizations of models with $ka=6$ and variable $k\sigma$ are plotted in Fig. 4. Again, the realizations remain grouped with differences in detail and a realization error of about ± 1 dB. The plot shows that scattering coefficients increase rapidly as $k\sigma$ increases for both models with $ka=1$ and $ka=6$. From $k\sigma=1.0$ to $k\sigma=6$, scattering increases by about 13 dB, which is 2.6 dB per unit $k\sigma$. This is more than double the variation observed over the same range in ka. This suggests that backscattering is much more sensitive to variation in surface height than to correlation length.

V. MODELS WITH HETEROGENEOUS SUBSEAFLOOR

The numerical calculation of scattering effects of rough seafloor with homogeneous subseafloor has shown that a significant portion of energy from low-grazing-angle incident wavefields penetrates into the subseafloor in rough seafloor models. Even in the case where total internal reflection occurs, energy is present in the shallow subseafloor via the direct wave root (Stephen and Bolmer, 1985; Brekhovskikh, 1960). All energy reaching the subseafloor is then available for elastic wavefield scattering if volume heterogeneity in velocity and density is encountered. When such scattering occurs, some of the scattered waves are directed upwards and return to the seafloor. Dougherty and Stephen (1988) showed that scattering from volume heterogeneity excites interface waves, which propagate along the seafloor and scatter back into the water column when further heterogeneity at the interface is encountered. Energy that is reradiated from the seafloor back into the water column from this interaction is observed as seabed volume scattering.

As with seafloor scattering, the spatial scales over which the volume heterogeneity occurs, compared to the wavelength in the subseafloor medium, are expected to have a significant effect on the characteristics of the volume scattering wavefield. In this study, volume heterogeneity at wavelength scales is defined as isotropic heterogeneity in velocity and density, with correlation lengths and rms perturbation amplitudes defined by stochastic Gaussian functions using the same procedure as for rough seafloor but applied in two dimensions.

There are essentially no measurements of fine-scale (in the sense of 6-m wavelength) lateral variation in subseafloor properties anywhere in the deep ocean. Goldberg and Sun (1997) analyzed well log data from ODP Hole 504B (on the flank of the East Pacific Rise) to estimate the scales of vertical heterogeneity in the shallow crust (<2.1 km). They found that the subseafloor was dominated by variations having correlation lengths of 1 and 10 m. Their interpretation of this is that the 1-m heterogeneity is characteristic of pillow basalts and the 10-m heterogeneity is characteristic of lava flows and large-scale faulting. We model exposed igneous basement areas as being isotropically heterogeneous with correlation lengths predominantly on the order of 10 m. We assume that lateral heterogeneity and vertical heterogeneity at our wavelength scale are similar.

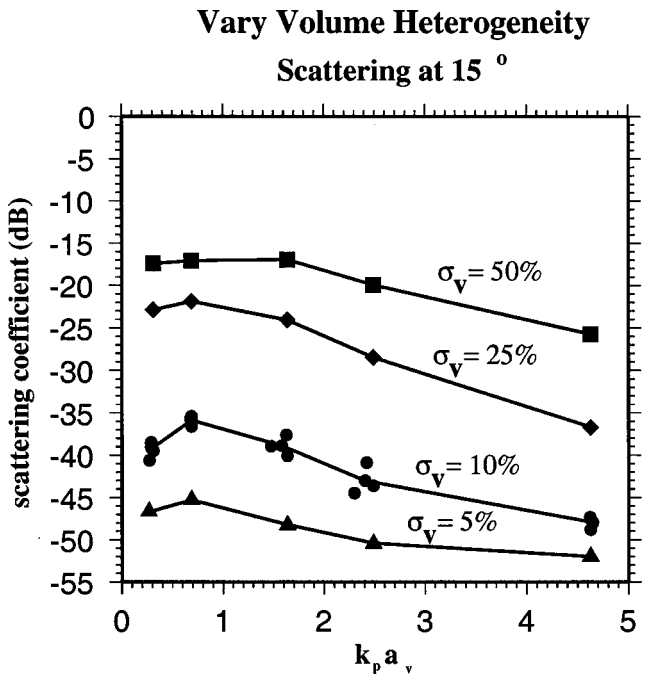


FIG. 6. Scattering coefficients for basaltic flat-seafloor models with different perturbation amplitudes in velocity and density. The amplitude of the scattering coefficients increases as the rms perturbation amplitude increases. This figure compares the scattering coefficients for the flat-seafloor models with volume heterogeneity having $\sigma_v=5\%$, 10%, 25%, and 50% for different volume correlation lengths. The sensitivity to correlation length of the volume heterogeneity remains about the same as the rms perturbation amplitude varies and is similar to that of rough seafloor scattering. The sensitivity to variation in the perturbation amplitude is quite strong, but only very high rms perturbation produces signal levels in the -10- to -20-dB range that is typical of scattering from a rough seafloor.

Seabed models with heterogeneous distributions of subseafloor velocity and density are constructed by defining 2-D distributions of perturbations in these properties, relative to a specified homogeneous background model. To create isotropic heterogeneity, 2-D stochastic distributions with Gaussian autocorrelations are generated having specified correlation length (a_v) and standard deviation of the perturbation amplitude (σ_v), equal in all directions within the subseafloor. The rms perturbation amplitude in these models is defined as a percentage of the mean (homogeneous) velocity and density. In all models, the range of perturbed velocity is constrained to $(v_p, v_s)<7.0$ km/s to maintain numerical stability. Shear wave velocity is calculated from compressional wave velocity assuming Poisson's ratios of 0.33 for hard bottoms and 0.40 for soft bottoms. Density is calculated assuming a Nafe–Drake- [derived from Fig. 11 in Ludwig *et al.* (1970)] type linear relationship to compressional velocity, i.e., $\rho=0.3788v_p+c$, where the constant, c, is determined by setting ρ and v_p to the desired mean values for the subseafloor.

A. Scattering versus a_v and σ_v

Figure 5 shows three examples of a flat seafloor with isotropic volume heterogeneity. In each case, the background homogeneous (mean) model is defined with the hard-bottom (basaltic) model properties. The perturbation amplitude of

$\sigma_v = 10\%$ applies to all three medium parameters, i.e., compressional and shear velocity and density. The model variable is the correlation length of the spatial distribution of the perturbation. For volume scattering, the dimensionless descriptor, ka_v, depends on the velocity of the propagation mode in the subseafloor, whereas in the seafloor roughness (interface) modeling *ka* was only dependent on the water velocity. For compressional waves, $k_p a_v = 2\pi a_v / \lambda_p$, where λ_p is the wavelength defined by the mean compressional velocity. For shear waves, $k_s a_v = 2\pi a_v / \lambda_s$, where λ_s is the wavelength defined by the mean shear velocity. For the following discussion we use the values of $k_p a_v$ when referring to volume heterogeneity and *ka* when referring to seafloor roughness.

An inherent problem in determining the scattering effects of volume heterogeneity is that in order to isolate the scattering due to volume heterogeneity from surface scattering, the seafloor is assumed to be flat. For source wavefields incident at subcritical grazing angles, this modeling, at first glance, appears to be a comparison of volume scattering with wavefield energy only reaching the subsurface via the direct wave root. The direct wave root is evanescent below the seafloor and, consequently, it makes direct contact with only the shallowest portion of the subseafloor model. The resulting scattered-field intensity varies as the correlation length varies, but the magnitude of the scattered field is very small and may not reflect the magnitude of volume scattering that could occur if more energy penetrated the seafloor, as would be the case with a rough seafloor or a low-velocity subseafloor, i.e., when the incident wavefield grazing angle is greater than the critical grazing angle.

However, further consideration suggests that to some extent the spatial perturbations in these models transform the flat seafloor into what is effectively a rough seafloor, i.e., variations in subseafloor velocity and density at the seafloor cause the seafloor to appear to the incident wavefield as an irregular boundary. That is, as the velocity varies laterally along the seafloor, the critical grazing angle varies, which can be thought of as equivalent to a homogeneous model with a rough surface. This allows more incident energy to scatter into the subseafloor as converted shear and compressional waves than is available from just the direct wave root.

Figure 6 shows the scattering coefficients calculated for hard, flat-seafloor models with isotropic heterogeneous subseafloor having $\sigma_v = 5\%$, 10%, 25%, and 50% for a range of correlation lengths (a_v). The plot shows that there is a peak in the scattering coefficient in the vicinity of $k_p a_v \approx 1$ and that volume scattering decreases above $k_p a_v \approx 1$ at a rate of about -3 dB per unit $k_p a_v$. Variations in scattering coefficients attributable to different realizations (for $\sigma_v = 10\%$) are about ± 1 dB, suggesting that changes in $k_p a_v$ greater than about one-third of a unit can be identified. The sensitivity to $k_p a_v$ is as strong as that observed for variation in rms height ($k\sigma$) of the seafloor roughness. The rapid increase in the scattering coefficients as the perturbation amplitude is increased indicates that volume scattering is very sensitive to this parameter, just as seafloor scattering is strongly sensitive to variation in seafloor height. When $\sigma_v = 50\%$, the magnitude of the volume scattering coefficients are comparable to rough seafloor scattering coefficients. However, all of the other volume heterogeneities produce coefficients that are 5–10 dB smaller than signals produced by rough seafloor scattering alone.

In order to evaluate whether this variation is attributable to variation in scattering from the volume heterogeneity or to variation in the amount of energy that penetrates the seafloor interface, seafloor roughness is added to the models with volume heterogeneity.

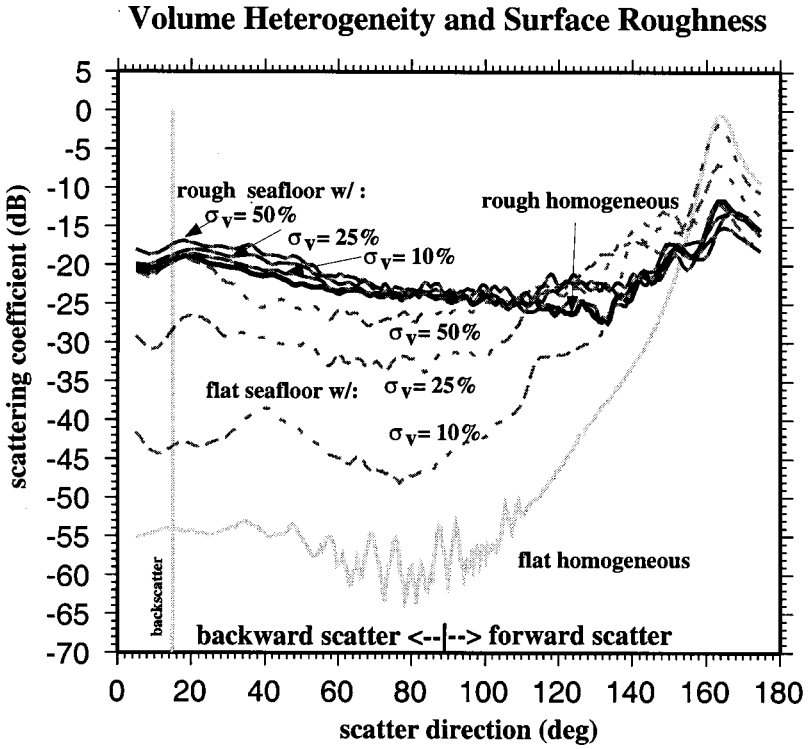


FIG. 7. The complete scattering functions calculated for rough basaltic seafloor with heterogeneous subseafloor differ very little from the same seafloor with homogeneous subseafloor. This is true even in the case of very high rms perturbation amplitude. The solid line is the scattering function of rough seafloor with a homogeneous subseafloor. The upper gray-dashed lines are for the same rough seafloor with different volume heterogeneities. The lower dashed lines are results for flat seafloor with the same volume heterogeneities. The baseline is the background noise level calculated as the scattering from a horizontal flat and homogeneous basaltic seafloor. For all these models the mean velocity and density are $v_p^b = 4.0$ km/s, $v_s^b = 2.0$ km/s, and $\rho^b = 2.5$ kg/m^3. Also, for each model, the seafloor roughness is defined with $ka = 6.1$ and $k\sigma = 4.0$ and the volume heterogeneity is defined with $ka_v = 6.6$, which in the subseafloor translates to $k_p a_v = 2.5$.

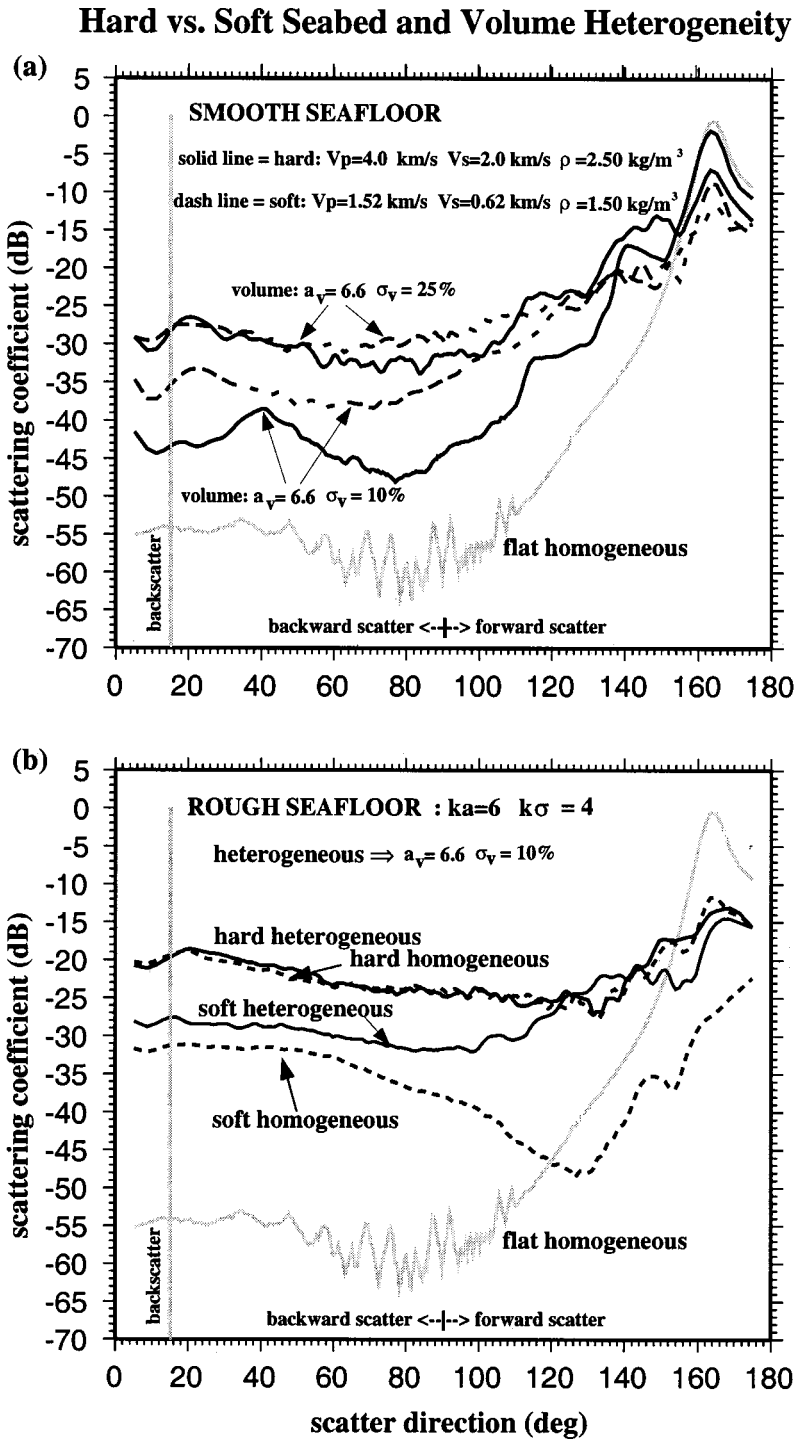


FIG. 8. Comparisons of scattering functions for hard (basalt) and soft (sediment) seafloor for different values of σ_v. (a) Flat heterogeneous models with either hard (solid lines) or soft (dashed lines) bottom subseafloor mean velocity and density generate comparable scattering functions when the volume heterogeneity is the $\sigma_v = 25\%$. For $\sigma_v = 10\%$ the scattering from the heterogeneous soft seafloor is significantly higher than from the hard bottom. (b) Adding subseafloor heterogeneity to a rough ($ka = 6$ and $k\sigma = 4$) hard bottom (solid line) does not substantially change the scattering function compared to the result from a homogeneous subseafloor (dotted line). For a soft bottom the difference between the scattering from a heterogeneous and a homogeneous subseafloor model is a stronger (about 5 dB in the backscatter direction). These scattering functions are for the case of $\sigma_v = 10\%$. (c) The same as (b), except that $\sigma_v = 25\%$. At this high level of perturbation amplitude in the heterogeneity, soft bottoms produce much higher scattering coefficients than the same rough seafloor with a homogeneous bottom, about 10 dB in the backscatter direction. Heterogeneity in hard bottoms produces almost no difference in scattering coefficients in any direction.

B. Rough seafloor and isotropic volume heterogeneity

In order to consider the combined effects of volume heterogeneity and seafloor roughness, scattering from a rough and heterogeneous hard bottom seabed simulation is compared to results for the same seafloor with homogeneous subseafloor, and to results for a flat but heterogeneous seabed. The rough seafloor in these models is defined with $ka = 6$ and $k\sigma = 4$, and the isotropic volume heterogeneity with $k_p a_v = 2.5$. Comparisons are made for three volume perturbation amplitudes, $\sigma_v = 10\%$, $\sigma_v = 25\%$, and $\sigma_v = 50\%$. Figure 7 compares the complete scattering coefficient functions for these models to those of flat-seafloor models with the same heterogeneity. When the rough seafloor is added to the heterogeneous volume models, the resulting scattering functions are almost identical to the case of the homogeneous rough surface. This implies that even with more energy penetrating the seafloor, the volume heterogeneity has little effect on the total scattered field. Even the rough-surface model with $\sigma_v = 50\%$ shows an increase of only about 2 dB over the homogeneous model. Because the rough homogeneous model and the flat heterogeneous model with $\sigma_v = 50\%$ have similar scattering levels, particularly in the backscatter directions, we might expect that the model with

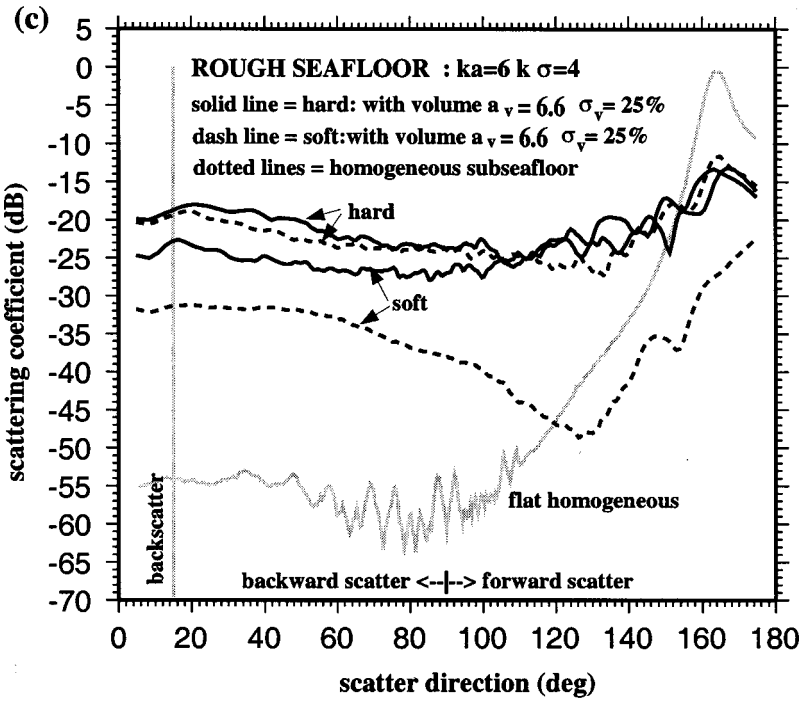


FIG. 8. (*Continued.*)

both the rough surface and volume heterogeneity would yield much higher scattering coefficients than either of the original models. However, the observed scattering coefficients are only increased slightly relative to the homogeneous model, suggesting that the ''volume scattering'' is predominantly scattering from the volume heterogeneity very close to the seafloor and, in effect, is ''acting'' like a rough seafloor. When volume heterogeneity is included in a rough-seafloor model, it essentially adds more roughness to the seafloor from the standpoint of interaction with the incident wavefield.

Although this is a very limited set of possible models, these observations strongly suggest that volume heterogeneity, at least in hard bottom models, cannot be distinguished from rough surface scattering. Nor does scattering from volume heterogeneity substantially increase the net scattering observed from a rough seafloor without subseafloor volume heterogeneity. It seems that most of the signal that is scattered into the subseafloor becomes trapped there, undergoing multiple scattering with little signal returning to the seafloor. It is true that the backscattered signal has sensitivity to volume heterogeneity correlation length and perturbation amplitude as strong as that of seafloor correlation length and rms height variation. However, the intensity level of the volume scattering is low compared to that produced by seafloor scattering, unless the perturbation amplitude is in the range of 50%. This result is contrary to our expectation and suggests that scattering from rough hard bottoms can be well represented by scattering from the rough surface alone.

C. Hard versus soft bottoms

If the subseafloor is a ''soft-bottom'' model, i.e., has both a compressional velocity only slightly greater than water velocity and a shear velocity less than water velocity, the seafloor becomes essentially transparent to the incident wavefield because the critical grazing angle is almost always exceeded. In this case the energy transmitted into the subseafloor, and available for volume scattering, is at its maximum. Figure 8(a) compares the scattering coefficients calculated for models computed with and without volume heterogeneity, for seabeds with hard- and soft-bottom parameters. For the soft bottom we used $v_p^s = 1.52$ km/s, $v_s^s = 0.62$ km/s and $\rho^s = 1.5$ kg/m^3, which is within the range of seafloor sediment properties and yet allowed for numerical simulation of wave propagation without strong grid dispersion for our grid spacing. For $\sigma_v = 25\%$, the difference between the scattering coefficients for flat seafloors shows very little sensitivity to the mean velocity of the models, particularly in the 15-degree backscattering direction. However, for $\sigma_v = 10\%$ the scattering is significantly stronger ($\sim$10 dB) from soft flat bottoms than from hard flat bottoms.

If we consider models that include seafloor roughness and subbottom heterogeneity, significantly different results are obtained. Figures 8(b) and (c) compare scattering functions of rough heterogeneous models to those of rough homogeneous models for both soft and hard bottoms with $\sigma_v = 10\%$ and $\sigma_v = 25\%$. The rough hard-bottom model results show, in Fig. 7, that volume heterogeneity has little effect on the net scattered field when the seafloor is rough. Contrary to this, the rough heterogeneous soft-bottom models produce significantly higher scattering coefficients than the homogeneous models. In the 15-degree backscatter direction, the difference is about 10 dB for the $\sigma_v = 25\%$ model and about 5 dB for the $\sigma_v = 10\%$ model. Although $\sigma_v = 25\%$ heterogeneity in a soft bottom may be a somewhat high variability for normal sediment, the result does suggest that heterogeneity between $\sigma_v = 10\%$ and $\sigma_v = 25\%$ can produce significant backscatter signal variations on the order of 5 dB.

VI. DISCUSSION

A motivation for this study has been the hypothesis that some of the variations in signal intensity observed in monostatic reverberation data can be explained by the scattering effects of wavelength-scale roughness and/or heterogeneity. It was also believed that both seafloor (interface) heterogeneity and subseafloor (volume) heterogeneity generate comparable scattered field intensities and variations in intensity.

For rough-seafloor models, we found that backscatter is most sensitive to variation in the rms height of the stochastic seafloor structure. The maximum range in intensity associated with height variation is about 15 dB, whereas the maximum range observed for variation in correlation length is only about 5 dB. For a horizontal seabed that is assumed to be uniform in all other characteristics, monostatic signal intensities that vary more than 5 dB are a strong indication that seafloor height is varying. Our results suggest that as the rms height of seafloor roughness increases, the backscatter signal approaches a limit, the magnitude of which depends on the correlation length.

The most striking result of all the modeling performed for this study is that rough basaltic (hard) seafloor, with or without subseafloor volume heterogeneity, produces essentially the same scattered fields. In particular, the backscatter intensity is always within a few decibels, even when the rms perturbation amplitude is as large as 50%. The scattering from flat-seafloor models with subseafloor heterogeneity shows strong sensitivity to variation in correlation length and rms perturbation amplitude of the volume heterogeneity but much smaller magnitude. This sensitivity is at the same level as that measured for seafloor rms height variation. This leads to the conclusion that the volume scattering in hard bottoms is not contributing substantial energy to the observed scattered field. It is suggested that the primary effect of volume heterogeneity is to increase the apparent roughness of the seafloor. It appears that energy scattered into the subseafloor is trapped in the multiple scattering environment such that very little energy is returned to the seafloor from below.

Contrary to this, the addition of volume heterogeneity to a rough seafloor model with soft bottom volume properties very significantly increases the scattering coefficient. This suggests that heterogeneity in soft bottom regions, such as sediment ponds, can produce backscattered signals up to 8 dB above the background noise level.

VII. CONCLUSION

The results of the numerical wavefield modeling of a variety of seabed models with well-defined geological characteristics have led to several important conclusions that impact the interpretation of monostatic reverberation data. However, it must be emphasized that these results do not account for the effects of seafloor dip or any specific large-scale seafloor features, such as fault scarps, that persist over more than a few wavelengths, nor do they include effects of large-scale subseafloor geological variations, such as velocity gradients or attenuation.

Variation in seafloor backscattering is most strongly affected by variation in the rms height of the wavelength-scale seafloor roughness and it is affected much less by variation in seafloor correlation length. If the subseafloor is approximated as a homogeneous medium, observed backscatter intensity can be interpreted in terms of seafloor height variations in wavelength scale roughness.

Another important conclusion is that the scattering generated by insonification of a hard-bottom seabed with low-grazing-angle wavefields is dominated by seafloor roughness scattering. Volume scattering, from a hard bottom subseafloor with volume heterogeneity alone, can produce strong backscatter signals when the perturbation amplitude is high ($\sim$50%). However, when combined with surface roughness, volume scattering does not appear to significantly alter the total backscattered signal. In contrast, when the volume heterogeneity occurs in low-velocity materials such as sediment, volume scattering can produce a significant increase in observed backscatter signals compared to homogeneous sediment bottoms. In some cases, these signals can be as strong (or stronger) as backscatter from smooth hard-bottom areas such as basaltic sheet flows. The real impact of volume heterogeneity is that the velocity variation at the seafloor adds to the apparent roughness of the seafloor, which increases the backscattered field.

ACKNOWLEDGMENTS

This work was funded by the Office of Naval Research under Grant Nos. N00014-93-1-1352, N00014-90-J-1493, and N00014-95-1-0506. This article is WHOI Contribution Number 10186.

Aki, K., and Richards, P. G. (**1980**). *Quantitative Seismology, Theory and Methods* (Freeman, New York), Vol. 2.

ARSRP (**1993**). *Acoustic Reverberation Special Research Program, Initial Report, Acoustics Experiment, R/V Cory Chouest, 5–26 July 1993*, Office of Naval Research.

Baggeroer, A. B., and Dyer, I. (**1986**). ''Long range, low frequency acoustic backscattering: a review,'' in *Ocean Seismo-Acoustics, Low-frequency Underwater Acoustics*, edited by T. Akal and J. M. Berkson (Plenum, New York), pp. 313–326.

Berkhout, A. J. (**1987**). *Applied Seismic Wave Theory* (Elsevier, New York).

Brekhovskikh, L. M. (**1960**). *Waves in Layered Media* (Academic, New York).

Caruthers, J. W., Yoerger, E. J., and Novarini, J. C. (**1997**). ''Modeling low-frequency reverberation near the Mid-Atlantic Ridge and comparison with ARSRP data,'' J. Acoust. Soc. Am. **101**, 2555–2565.

Dougherty, M. E., and Stephen, R. A. (**1987**). ''Geoacoustic scattering from seafloor features in the ROSE area,'' J. Acoust. Soc. Am. **82**, 238–256.

Dougherty, M. E., and Stephen, R. A. (**1988**). ''Seismic energy partitioning and scattering in laterally heterogeneous ocean crust,'' Pure Appl. Geophys. **128(1/2)**, 195–229.

Dyer, I., Baggeroer, A. B., Schmidt, H., and Fricke, J. R. (**1993**). ''Discrete backscatter can be dominant in rough bottom reverberation,'' in *Ocean Reverberation*, edited by D. D. Ellis, J. R. Preston, and H. G. Urban (Kluwer Academic, Dordrecht), pp. 51–57.

Frankel, A., and Clayton, R. W. (**1986**). ''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**(B6), 6465–6489.

Goff, J. A., and Jordan, T. H. (**1988**). ''Stochastic modeling of seafloor morphology: Inversion of Sea Beam data for second-order statistics,'' J. Geophys. Res. **93**(B11), 13589–13608.

Goldberg, D., and Sun, Y. F. (**1997**). ''Seismic structure of the upper oceanic crust revealed by *in situ* Q logs,'' Geophys. Res. Lett. **24**(3), 333–336.

Greaves, R. J. (**1998**). ''Seismic scattering of low-grazing-angle acoustic waves incident on the seafloor,'' Ph.D. thesis, Massachusetts Institute of Technology/Woods Hole Oceanographic Institution.

Greaves, R. J., and Stephen, R. A. (**1997**). ''Seafloor acoustic backscattering from different geological provinces in the Atlantic Natural Laboratory,'' J. Acoust. Soc. Am. **101**, 193–208.

Hastings, F. D., Schneider, J. B., and Broschat, S. L. (**1997**). ''A finite-difference time-domain solution to scattering from a rough surface pressure-release surface,'' J. Acoust. Soc. Am. **102**, 3394–3400.

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

Kelly, K. R., Ward, R. W., Treitel, S., and Alford, R. M. (**1976**). ''Synthetic seismograms: A finite-difference approach,'' Geophysics **41**(1), 2–27.

Ludwig, W. J., Nafe, J. E., and Drake, C. L. (**1970**). ''Seismic refraction,'' in *The Sea*, edited by A. E. Maxwell (Wiley, New York), Vol. 4, pp. 53–84.

Makris, N. C., Avelino, L. Z., and Menis, R. (**1995**). ''Deterministic reverberation from ocean ridges,'' J. Acoust. Soc. Am. **97**, 3547–3574.

Ogilvy, J. A. (**1991**). *Theory of Wave Scattering from Random Rough Surfaces* (Institute of Physics Publishing, Bristol, England).

Robertsson, J. O. A., and Levander, A. (**1995**). ''A numerical study of seafloor scattering,'' J. Acoust. Soc. Am. **97**, 3532–3546.

Schneider, J., and Shlager, K. L. (**1997**). ''FDTD simulations of TEM horns and the implications for staircased representations,'' IEEE Trans. Antennas Propag. **45**, 1830–1838.

Stephen, R. A. (**1990**). ''Solutions to range-dependent benchmark problems by the finite-difference method,'' J. Acoust. Soc. Am. **87**, 1527–1534.

Stephen, R. A. (**1996**). ''Modeling sea surface scattering by the time-domain finite-difference method,'' J. Acoust. Soc. Am. **100**, 2070–2078.

Stephen, R. A., and Bolmer, S. T. (**1985**). ''The direct wave root in marine seismology,'' Bull. Seismol. Soc. Am. **75**(1), 57–67.

Stephen, R. A., and Dougherty, M. E. (**1993**). ''Canonical seafloor models and the finite difference method for low-angle acoustic backscatter,'' in *Computational Acoustics—Volume 1*, edited by R. L. Lau and A. R. Robinson (Elsevier Science, New York), pp. 227–246.

Stephen, R. A., and Swift, S. A. (**1994**). ''Modeling seafloor geoacoustic interaction with a numerical scattering chamber,'' J. Acoust. Soc. Am. **96**, 973–990.

Swift, S. A., and Stephen, R. A. (**1994**). ''The scattering of a low-angle pulse beam from seafloor volume heterogeneities,'' J. Acoust. Soc. Am. **96**, 991–1001.

Virieux, J. (**1986**). ''P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method,'' Geophysics **51**(4), 889–901.

Resonant response of complex shell structures

Douglas M. Photiadis,[a] Brian H. Houston, E. G. Williams, and J. A. Bucaro
Naval Research Laboratory, Washington, DC 20375-5000

(Received 25 May 1999; revised 1 April 2000; accepted 5 April 2000)

Resonances of shell structures play a prominent role in determining their acoustic characteristics. Among the most important of such characteristics are radiated noise, scattering, and self-noise levels. The nature of the structural resonances, for example the number of modes within a particular frequency band or the overall spatial dependence of the response, determines the importance of these phenomena in various contexts and what steps may be taken to modify their acoustic impact. In the past, knowledge of the resonant response of highly idealized systems has been used to guide the development of quiet platforms. The results of a scale model investigation of the effects of increased complexity on the nature of the resonances of submerged shell structures are reported. Effects presented here are flexural Bloch mode resonances, localized resonances resulting from structural irregularity, and the effects of non-axisymmetry on these phenomena. The implications of these results on acoustic design and the development of systems is discussed.
 [S0001-4966(00)02008-7]

PACS numbers: 43.40.Ey [CBB]

INTRODUCTION

Developing an understanding of the resonant response of shell structures is a crucial aspect of developing quiet structures and high-performance acoustic systems. Most of our knowledge concerning this issue derives from the study of highly idealized systems, systems as simplified as infinite flat plates or infinite cylindrical shells. While some characteristics of these systems persist in real marine structures, other characteristics, the structural resonances and their coupling to the external fluid being two of the more important, can be greatly modified by structural complexity. We have found from a scale model investigation that the changes resulting from the system complexity can be dramatic, impacting (1) the basic frequency bands in which the structure has the largest acoustic signature, (2) the basic mode types giving rise to acoustic signatures, and (3) the general spatial extent of the system resonances. All of these aspects of marine systems are important in the design of quiet platforms and high-performance acoustic systems.

We report here on the results of a research effort at Naval Research Laboratory (NRL) examining the influence of complexity on the acoustics of submerged shell structures in the mid-frequency range, i.e., the range $2<ka<20$ with k the acoustic wave number and a the radius of the structure. This effort consisted of predicting, measuring, and analyzing the resonant response and acoustic behavior of a sequence of models, proceeding from a highly idealized system (a simple cylindrical shell) to a highly complex model (an irregularly ribbed shell with a large ''random'' array of oscillators attached internally to the structure). Some of the results discussed here have been previously reported; we have included these results in order to present an overall picture of the effects of structural complexity on the acoustic response of systems. Aspects we are presenting for the first time are discussion and derivation of the expressions for the density of states of the various models, the real space response and radiation from a point excited framed shell with many internal oscillators, and structural acoustic design implications associated with our findings.

Many other researchers have attacked the problem of the effects of complexity in structural acoustics. The most recent flurry of results concerned ''fuzzy'' structure models,[1–3] but certainly there has also been a great deal of other research. We have restricted our discussion here only to work which directly relates to the experiments and associated interpretation we present below. Otherwise, this paper would become a general review article, as opposed to an attempt to present a unified view of our recent findings. The following general physical picture of the salient effects of increasing complexity has been developed.

In the uniform shell with no internal structure, it is well known that the resonances of the system are simple helical standing waves extending throughout the structure.[4] Those resonances whose wavelengths are greater than the acoustic wavelength, i.e., the modes associated with shear or compressional vibrational energy, are well coupled to the fluid and control the far-field acoustic signature arising from vibrations of the shell. The great majority of the resonances, i.e., the modes associated with flexural vibration, have short wavelengths, and, while relevant to the near-field acoustic properties of the system, they do not radiate to the far field. The development of acoustic systems for a structure well described by this idealization requires a detailed understanding of the properties of these modes both in the ideal case and under perturbations. Evidently, if one is most concerned with the far-field acoustic signatures of this simple system, the great majority of the resonances may be ignored.

As the complexity of the system is increased, the number of resonances well coupled to the acoustic fluid increases significantly, and the properties of these resonances change markedly. Generally, as the system becomes less uniform, the resonances of the system become more and more local,

[a]Electronic mail: doug@genah.nrl.navy.mil

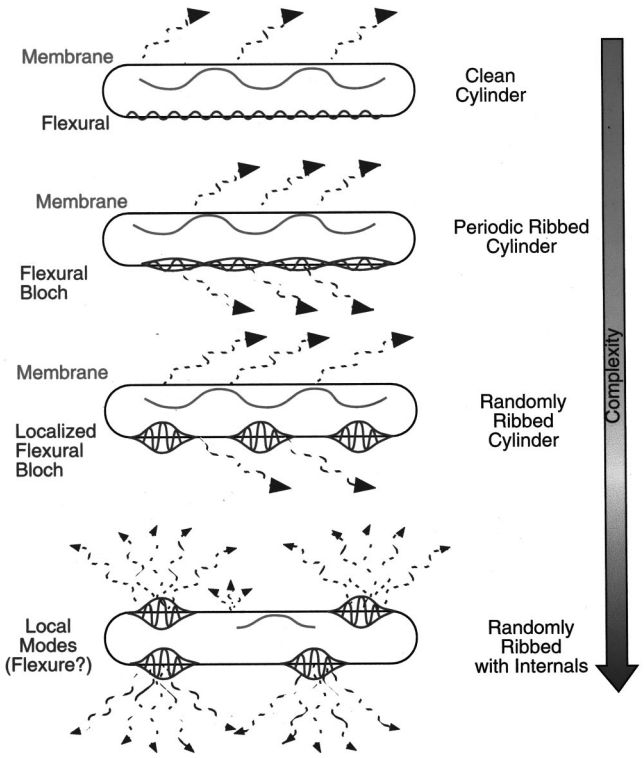


FIG. 1. Structural acoustics of ribbed shells as a function of complexity.

associated with particular locations along the structure, rather than with particular wave numbers. Further, the resonance frequencies associated with a particular motion of a given type, for example, a coupled frame-shell bay resonance, now vary in a somewhat random fashion as we traverse the shell. In some ways, the complex system is much more difficult to describe than an idealized structure; there are certainly many more parameters to evaluate. In other ways, the system in fact becomes simpler, with large-scale coherent vibration being suppressed and self-averaging processes eliminating the need for many details. These changes in the acoustic properties of the system as the complexity is increased are illustrated in Fig. 1, which shows an artist's depiction of the various vibration modes and their coupling to the external acoustic field.

In describing the effects of complexity, we focus on two aspects of the resonant structure of the systems: the spectral density or density of states (the number of resonances per unit frequency) and the overall spatial character of the resonances. These quantities are at the heart of many computations of vibro-acoustic quantities, particularly if one is interested in frequency-band-averaged quantities, as is likely to be the case for complex systems. In general, the modes dominating the density of states, the flexural modes, also dominate the dynamic motion of the structure in that any vibrational energy imparted to the system in a given frequency band is much more likely to excite bending than any other motion. The mode shapes are also clearly an important quantity, determining the nature of the acoustic coupling from one point to another along the structure and also determining the coupling of surface vibration to bulk acoustic waves in the surrounding fluid. In simple systems, for example, the nearly pure subsonic nature of the flexural modes enables the much more sparsely distributed shear and longitudinal waves to control the radiation and scattering characteristics of the system. In contrast, the corresponding flexural modes of complex structures often acquire a supersonic component, and the overall radiation and scattering properties of such systems, while influenced by the mode shapes, therefore more closely reflect the spectral density of the system.

The outline of the article is as follows. In Secs. I–IV, each step in the sequence of models shown in the figure above is discussed. We begin in Sec. I with a brief review of the ideal cylindrical shell and then examine the periodically framed case, the irregularly framed case, and finally a case including irregular framing and a pseudo-random array of attached internal oscillators. In Sec. V, a brief discussion is given of the impact of these results on the design of stealthy acoustic structures.

In the following sections, to simplify many of the theoretical discussions, we focus on the role of the cylindrical portion of the shell as opposed to the end caps. We assume the cylinder to be baffled, extended to ∞ by a rigid cylinder, and we further assume periodic boundary conditions on the elastic waves, $\psi(x)=\psi(x+L)$ with L being the length of the structure. These assumptions are not particularly realistic, but in this context, in which we hope to present a broad picture of the role of complexity in the structural acoustics of shell structures, the simplified modeling is sufficient and indeed beneficial.

I. UNIFORM CYLINDRICAL SHELL

The dynamics of the uniform cylindrical are well understood.[4,5] The resonances of the shell are traveling waves of the form

$$\psi_{nm}(\theta,z)=A^{-1/2}\exp(in\theta)\exp(2\pi mz/L)\epsilon_{nm}, \qquad (1)$$

where A is the area of the cylindrical shell and ϵ_{nm} is a polarization vector. The mode shapes are simple sinusoids with $k_\theta=n/R$ and $k_z=2\pi m/L$. Every resonance of the system extends over the entire structure with an average mean square amplitude which is constant and given by A^{-1}; i.e., the normal modes of the system are extended.

The resonance frequencies of the system are determined by the constraint that the helical wave number of a particular resonance must coincide with the helical wave number of one of the natural propagating waves of the cylinder, quasi-flexural, quasi-shear or quasi-longitudinal, and the resonances are thus classified as such. The individual resonances must therefore have a frequency and wave number which satisfy the dispersion relations of the shell, and the pattern of resonances will be a series of discrete frequencies that follow the dispersion curves.[6]

In Fig. 2, the measured surface velocity[7] of a simple cylindrical shell under point force excitation is shown in wave-number-frequency space for the $n=1$ azimuthal mode, the other azimuthal modes of the shell being fairly similar. One can see the clear decomposition between the large-wave-number, slow flexural waves and the small-wave-

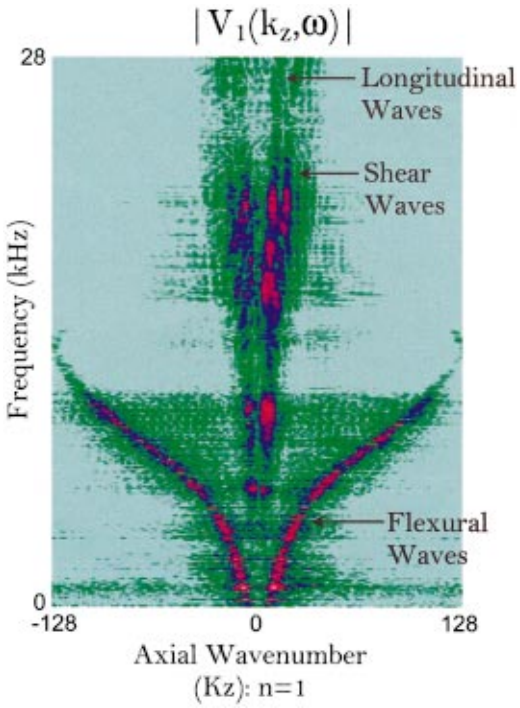


FIG. 2. Experimental dispersion curves for the $n=1$ mode of a cylindrical shell (from Ref. 7, p. 121).

number, fast membrane waves. Further, one can directly observe the resonances of the system as large enhancements at discrete frequency-wave-number values, with just a few membrane resonances in this frequency range and a large number of flexural modes.

Theoretically, the number of resonances in a particular frequency band may in general be obtained quantitatively as[8]

$$n(\omega)\Delta\omega=\sum_{nm}\int_{\omega}^{\omega+\Delta\omega}d\omega\,\delta(\omega-\omega_{nm}), \tag{2}$$

where the sum over nm is a sum over the resonant states of the system, the δ-function restriction giving the correct count of the resonances. Assuming that the wavelengths of the modes are short compared to the length and radius of the shell, L, $R\gg\lambda$, one may replace the sum over (n,m) by an integral in k-space, whence the δ-function restricts the integral to fall on the (k_θ,k_z) dispersion curves $C_i(k)$ for the ith wave type:

$$n(\omega)\Delta\omega\approx\frac{A}{(2\pi)^2}\sum_i\int_{C_i(\omega)}dC\,\frac{1}{|\nabla\omega_i(k)|}. \tag{3}$$

A simple expression for the density of states can be obtained from the above integral provided the dispersion relations can be approximated analytically. Neglecting the anisotropy in the shell response, i.e., taking the dispersion curves to be given by $\omega_i=c_i|k|$, a reasonable approximation for frequencies greater than the ring frequency, one may show that

$$n_i(\omega)\approx\frac{\omega A}{(2\pi)}\frac{1}{c_{pi}c_{gi}}, \tag{4}$$

where c_{pi} and c_{gi} are the isotropic phase and group speeds of the ith wave type, respectively, and A is the surface area of the shell. The expression given for the density of resonant modes by Eq. (4) is just the formula for the density of mode type i in a flat rectangular area A, a standard quantity in mathematical physics.[9] In a structural acoustics context, Maidanik[10] in 1962 derived an identical expression specialized for flexural plate modes. One may observe that in the mid-frequency range the density of states for a cylindrical shell is a relatively smooth function dominated by flexural waves because of their small wavelength and associated slow wave speed.

Of course, for the pseudo-longitudinal and pseudo-shear waves, the frequency will often not be large enough for the approximations above to be valid, and the sum over n in Eq. (2) will need to be retained in order to see the ''discrete'' enhancements at the cutoff frequencies for the in-plane modes. In the overall density of states this effect is small since there are so many more flexural modes than fast modes.

The vibrational and near-field acoustic behavior of the system is therefore generally governed by the excitation of the closely spaced flexural resonances, but in the far field, the acoustic signature is controlled by the small-wave-number, rather sparsely spaced resonances associated with the fast modes. The far-field acoustic signature of the structure has a fairly strong frequency dependence but a relatively simple angular response. (Recall that a structure-borne wave with wave number k produces an acoustic highlight at the angle $k_0\cos\theta=k$, where $k_0=\omega/c$ is the acoustic wave number in the fluid.) The backscattering signature of a uniform unframed shell is shown in Fig. 3(a). The ''chalice'' structure formed about the beam is due to scattering from the membrane modes,[11] while the flexural modes give no contribution to the far field.

II. PERIODICALLY FRAMED SHELL

The addition of ribs produces some significant changes in the resonances of the shell. The normal modes now occur in frequency bands, and there are discrete gaps in the spectrum as a result of Bragg scattering from the array of frames.[12] The mode shapes take on the Bloch or Floquet form appropriate to the periodicity a,

$$\psi_{n\alpha k_B}(\theta,z)=A^{-1/2}\exp(in\theta)\exp(ik_Bz)U_{n\alpha k_B}(z) \tag{5}$$

$$=A^{-1/2}\exp(in\theta)\exp(ik_Bz)\times\sum_j c^j_{n\alpha k_B}\exp(2\pi j/a). \tag{6}$$

Here, α is the band index, k_B is the Bloch or Floquet wave number, and $U_{n\alpha k_B}(z)$ is the single bay wave function. The Bloch wave number k_B is restricted by the boundary conditions to the values $k_m=2\pi m/L$, and the mode shapes are periodic, but they are no longer simple sinusoids. Instead, the single bay wave function $U_{n\alpha k_B}(z)$ which repeats along the length of the shell with a phase shift may be quite complicated. Nevertheless, averaged over a single bay, the mean square magnitude of each of the normal modes is constant over the entire system and given again by A^{-1}. Hence, all of

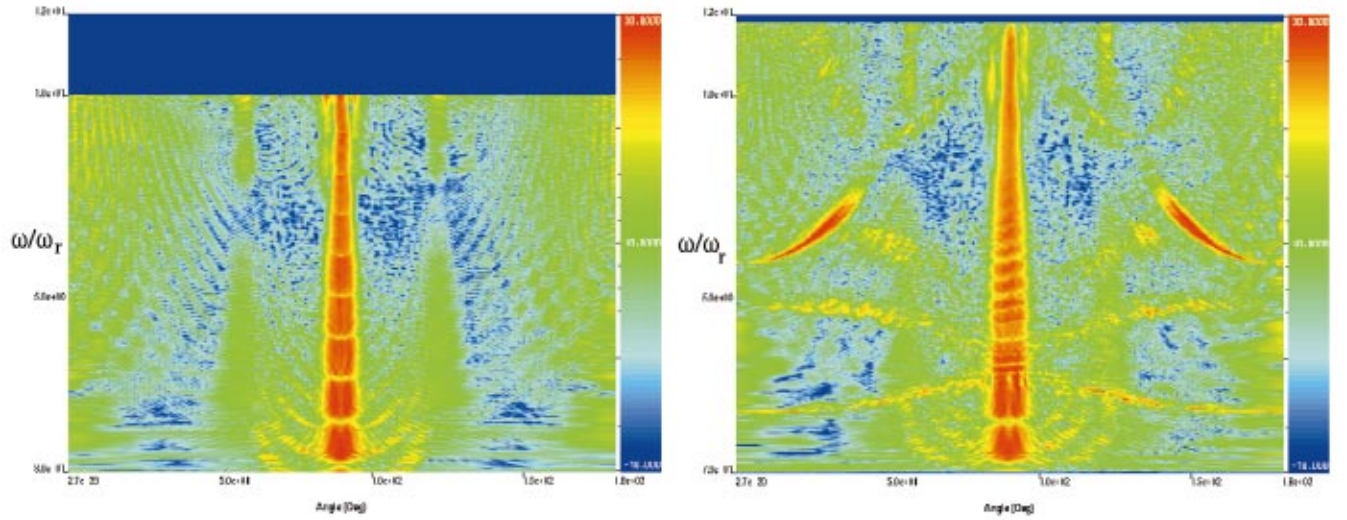


FIG. 3. Comparison of the backscattering cross section of an unribbed and ribbed shell (from Ref. 14, p. 2851).

the resonances of the system are extended, just as for the uniform shell.

To understand the distribution of the resonant modes of the ribbed shell, consider a particular azimuthal mode n. The Bloch wave number k_B of this mode is approximately related to the natural wave number $k(\omega)$ of the same mode of the uniform shell via[13]

$$\cos k_B a = \frac{\cos(k(\omega)a+\delta)}{|t|}, \tag{7}$$

where the single frame transmission coefficient for the elastic wave propagating on an unribbed shell is $t=|t|\exp(i\delta)$. The resonance frequencies of the system are determined by the constraint that the Bloch wave number k_B satisfy the dispersion equation, Eq. (7), and at the same time the boundary conditions. The frequency gaps appear when $\cos(k(\omega)a+\delta)$ is near its maximum and there is no real solution for k_B (because k_B must be real in order to satisfy the boundary conditions); i.e., the gaps are centered near the wave numbers $(k(\omega)a+\delta)=n\pi$ at which coherent backscattering from the array of frames dominates, thus preventing wave propagation.

Note that while the resonances can still be identified by wave type, the clean decomposition into ''fast'' modes and ''slow'' modes now refers only to the group speeds and not to the phase speeds, because the Bloch modes may contain fast Fourier components resulting from scattering by the periodic array of frames. Mathematically, this is apparent in the expression for the Bloch modes, Eq. (6), which contains multiple Fourier components. This aspect is quite important in the context of radiation and scattering as shown in Fig. 3, which compares the backscattering cross section of framed and unframed shells.[14] The slow group speed features in the cross section are due to coupling of the acoustic field to the supersonic components of the flexural Bloch modes. The angular dependence is predicted once again via simple phase matching, $k_0\cos\theta=k$.

The density of states of the system is significantly altered from that of the uniform shell. Proceeding in a similar manner as in the previous section, we have in analogy to Eq. (3)

$$n(\omega)\approx\frac{A}{(2\pi)^2}\sum_{i,\alpha}\int_{C_{i\alpha}(\omega)} dl\,\frac{1}{|\nabla\omega_{i\alpha}(k)|}. \tag{8}$$

In this case, the sum is now over the various bands α in addition to the wave types and the integral is restricted to the first Brillioun zone in (q_z-k_θ) space, where q_z is the Bloch wave number in the z direction. The resulting density of states, $n(\omega)$, is still dominated by the flexural modes in the mid-frequency range, but may differ significantly from that of the uniform shell because of the clustering of states into bands.

Geometrically, the resonances of the framed shell fall on the disconnected surface determined by the dispersion equation $D(q_z,k_\theta,\omega)=0$. The branches of the surface are smooth, bounded from below in frequency space, and hence exhibit at least one extremum at which the gradient $|\nabla\omega_{i\alpha}(k)|$ vanishes. The detailed analysis of this surface is interesting but beyond the scope of this article; we merely note here that it is straightforward to show, based on Eq. (7), that a minimum in a branch of the dispersion surface can occur at $k_\theta=n/R=0$, $q_z=0$, π/a if the single rib transmission coefficient is slowly varying in a particular frequency range. At frequencies below this cutoff, none of the resonances associated with this ''sheet'' of the dispersion surface will contribute to the density of states, while above the cutoff the number of resonances contributed by the branch in question will vary smoothly. At the cutoff frequency the density of states exhibits a singularity owing to the vanishing of the

group velocity $|\nabla \omega_{i\alpha}(k)|$. This behavior is in significant contrast to the smooth, featureless behavior of the density of states of the smooth shell.

III. IRREGULARLY FRAMED SHELL

We examine here the effect of irregularity in the frames which does not destroy the axisymmetry in the system, the first level of real complexity that one might add. The resonances of the system therefore still consist of individual Fourier components,

$$\psi_{n\beta}(\theta,z)=\exp(in\theta)\phi_{n\beta}(z), \tag{9}$$

but are no longer simple elastic Bloch or Floquet modes. Here, β is merely a label for the various resonances which occur for a given azimuthal order n.

The resonance spectrum of the system depends in a natural way on the amount of irregularity in the system: If the irregularity is small, the resonance frequencies are weakly shifted from those of the associated periodic system, while if the irregularity is large, the resonances are spread in a more or less random way. In the weakly irregular case, the total density of resonant modes of the system is therefore changed from that of the ideal periodic system only in the vicinity of the frequency band edges. In this frequency region, the sharp edge in the distribution of the resonances (a singularity in the density of states) is smoothed out by the irregularity. This may in some circumstances be useful, because the high concentration of resonances near a band edge may produce unwanted acoustic behavior. Nevertheless, the qualitative impact of irregularity on the total density of states is relatively small.[15] We describe the density of states here as the *total* density of states to emphasize that because the system is not translation invariant, the density of states may change as the position is varied; and, further, the local density of states, the number of resonances per unit frequency observed at a particular position, may differ from the total density of states. Generally, the local density of states will be reduced from its value in the periodic or uniform system in which the local density of states is the same as the total density of states.

The mode shapes $\phi_{n\beta}(z)$ are in many cases significantly modified from the Bloch or Floquet modes of the associated periodic system, even by small amounts of irregularity.[16,17] The form of $\phi_{n\beta}(z)$ has been shown under very general circumstances to be a standing mode with an exponential envelope,[18–20]

$$|\phi_{n\beta}(z)|\rightarrow\exp(-|z|/l_{n\beta})=\exp(-\gamma_{n\beta}|z|) \quad \text{as} \quad z\rightarrow\infty. \tag{10}$$

This phenomena is termed Anderson localization,[21] and the quantities $l_{n\beta}$ and $\gamma_{n\beta}$ are called the localization length and localization factor, respectively. The spatial nature of the modes of system has thus been fundamentally changed by the irregularity. A new length scale determined by complex multiple scattering phenomena, the localization length, has been generated.

The relevance of irregularity that does not break the axisymmetry of the system thus depends critically on the localization length: If this length scale is significantly smaller than both the size of the system and the damping length of the particular mode, then this change in the resonant mode shapes significantly impacts the acoustic behavior of the system. The ''damping length'' of a particular mode is the distance over which a packet of vibrational energy may propagate before being significantly attenuated by internal damping mechanisms; i.e., $l_d \approx c_g \tau$ with c_g the group speed and τ a typical damping time.[22] In Fig. 4 are shown experimental measurements of the spatial response of an almost periodically ribbed shell subjected to a point force excitation.[22] The deviation of the ribs from their nominal ideal periodic locations was measured to be $\pm 2\%$ of the average spacing. The surface response was obtained from near-field acoustic holography (NAH) measurements. For a given value of n, one is observing at most only a few resonances of the system, and in many cases only a single mode. In any case, provided one is further than a single localization length from the source, the observed spatial decay rate is the modal localization factor.[22] For small values of n, the magnitude of the modal responses are nearly uniform across the structure (when they are not radiation damped), but as n increases the modes become increasingly localized. Already at $n=11$ this phenomena is fairly significant, but for the larger values of n, it is completely dominant. The character of the system response has changed entirely from standing waves extended over the entire structure to standing waves localized in the vicinity of ''defects.''

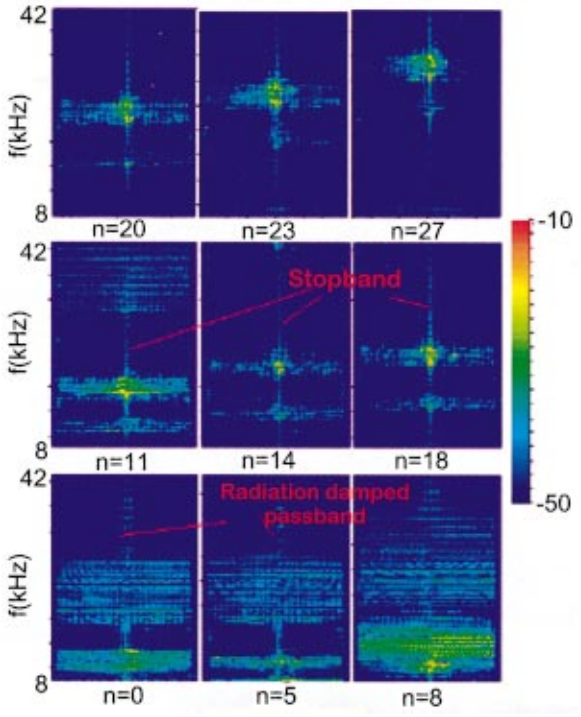


FIG. 4. Response of the various azimuthal modes of an irregularly ribbed shell to a point excitation. The horizontal coordinate is axial position along the shell while the vertical coordinate is frequency. The point drive is applied in the center of the shell and the localization of vibration about the source for the higher modes is evident (from Ref. 22, p. 1377).

The structural acoustics of the system have thus been significantly impacted by even this small amount of irregularity. The principal effect of adding irregularity to the sys-

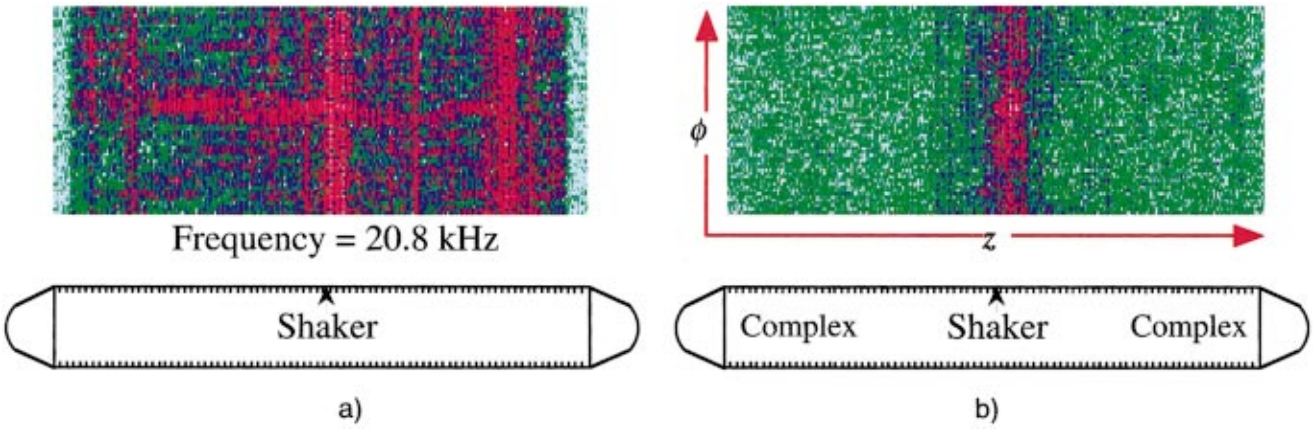


FIG. 5. Spatial response of an irregularly ribbed shell compared to the response of an identical shell with a large array of internal oscillators.

tem has been to introduce a new dynamic length scale, a localization length which specifies the structural ''decoupling'' length. The modes of the system are no longer extended with a mean square value of A^{-1}, and the typical magnitude of the modal response is significantly increased, now being of order $(2\pi R*l)^{-1}$. Hence the flexural resonances, when localized, are even more dominant in the irregular shell than in the periodically framed case. The real space response of the shell with irregular rib spacings is shown in Fig. 5 compared to the response of the same shell with internal oscillators. The strong enhancements of the response of the irregularly ribbed shell at certain axial locations results from matching to local resonance frequencies, a phenomena which cannot occur on a periodic shell.

Despite the dramatic change in the structural acoustic near field, the far-field acoustic signature of the system, being governed by the supersonic small n structural modes, is only weakly affected by the irregularity in the system. This is apparent in Fig. 4 in which the $n=0,5,8$ modes are seen to be extended, just as one would expect for a Bloch mode of a periodic system. As the irregularity is increased, theoretical studies[23] predict that the small n modes will be influenced by irregularity, and accordingly the far-field acoustic signature will be modified; but because a great majority of the structural resonances is still nonradiating, the acoustic near field rather than the far field will be most influenced by complexity which preserves the axisymmetry of the system.

IV. NON-AXISYMMETRIC INTERNAL STRUCTURE

In this section, the complexity of the system is increased in two ways: First, the axisymmetry in the structure is destroyed, and, second, additional internal degrees of freedom are added to the system. Further, the added structure is not so dominant as to render the framed shell to be isotropic; rather, the effects of the frames are still of significance. These properties, the anisotropy together with the non-axisymmetry, significantly reduce our ability to make theoretical predic-

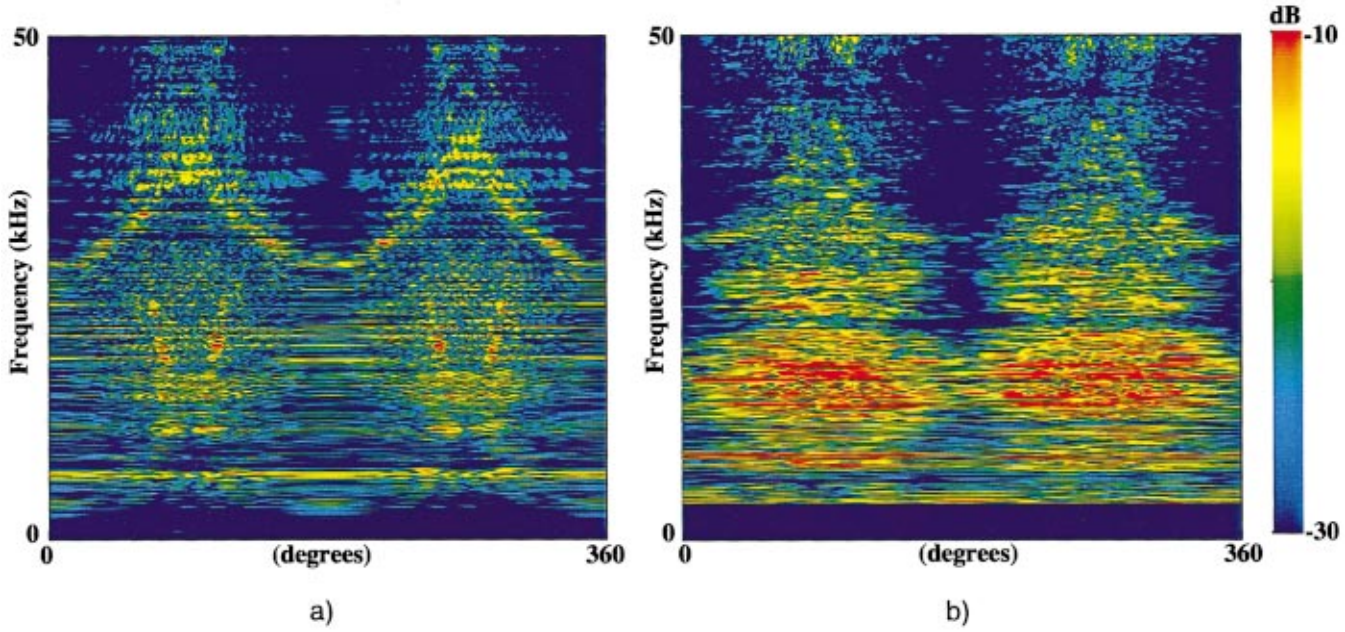


FIG. 6. Radiation resulting from a point drive applied to an irregularly ribbed shell with an array of internal oscillators; (a) irregularly framed shell with no internals and (b) irregularly framed shell with oscillators.

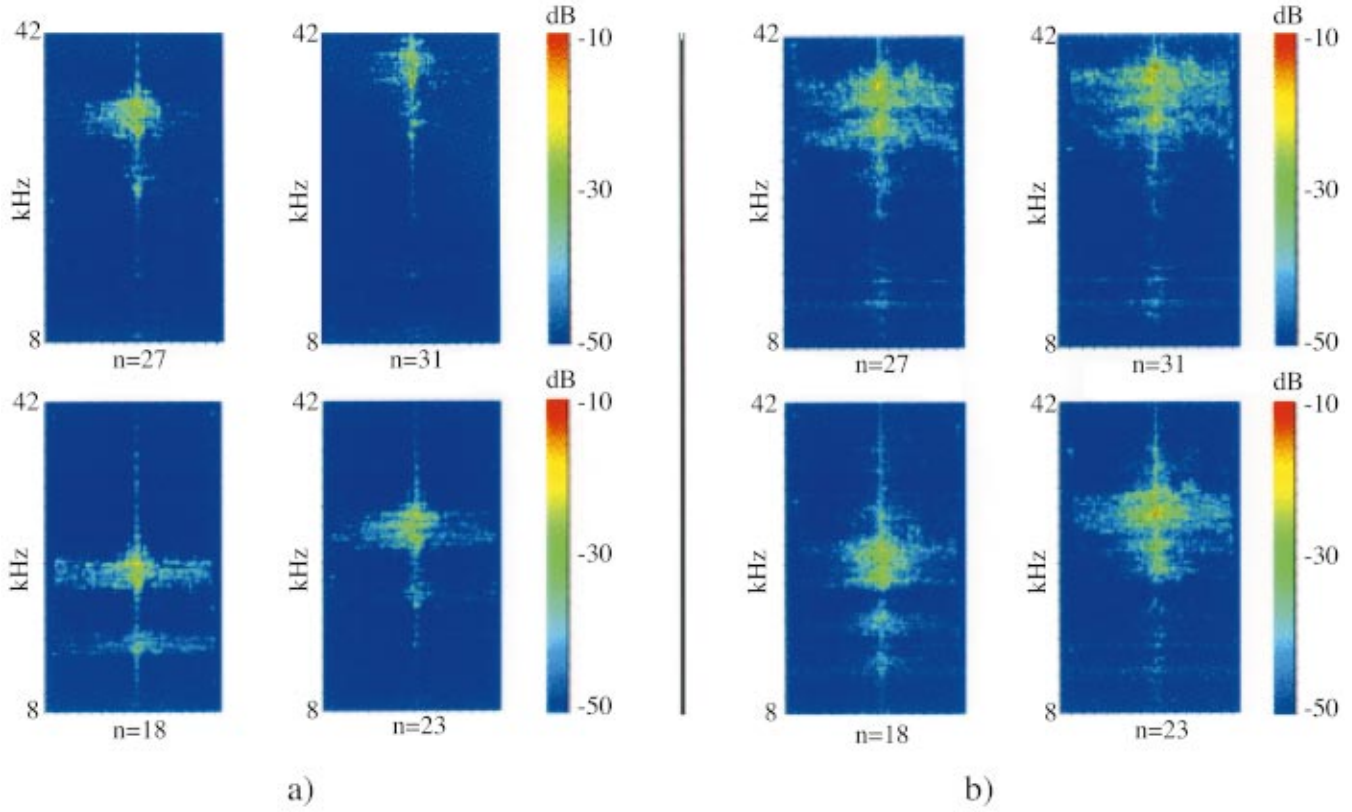


FIG. 7. Comparison of the large n modal response of an irregularly ribbed shell to the response of an identical shell with a large array of internal oscillators.

tions of the nature of the resonances of the system. Typically, an isotropic 2-D system with irregularity exhibits less mode localization than a similar 1-D system,[24] the effect of the irregularity being to destroy the coherence in the system rather than to actually produce localized standing modes. It is difficult, however, to make general conclusions in a case in which there is significant anisotropy. Let us therefore immediately examine the experimental results.

The model on which the experiments were performed was the irregularly framed model examined in the previous section modified by the addition of 880 pseudo-randomly arranged oscillators. Each oscillator adds a single internal degree of freedom within the frequency band in question, $10\,\text{kHz} < f < 40\,\text{kHz}$, and the spectrum of oscillator frequencies were chosen to cover this band. Details regarding the model are contained elsewhere.[25,26] Unlike in the previous examples, a modal decomposition azimuthally does not reveal the normal modes of the system. In Fig. 5 is shown a typical real space response to a point force of the oscillator shell along with that of the ''simple'' irregularly framed shell. The very significant spatial localization of the response is apparent. Quite clearly, the long-range coherent modes which persisted on the irregularly framed shell have been heavily suppressed by the internal oscillators.

Some further aspects of the resonant behavior can be uncovered by examining a few additional pieces of data. Figure 6(a) shows the radiation from a point-driven irregularly framed shell with no internals while Fig. 6(b) shows the corresponding radiated noise signature for the shell with internal oscillators. In both cases, the shaker is oriented normal to the rotation plane with a corresponding symmetry in the radiation signature. The presence of the oscillators has had a profound impact on the character of the signature. The scattering data[25,26] also show this very large impact. Notice a couple of features of the data. Nowhere to be found are the clear, coherent far-field features indicative of extended wave propagation. Instead, the far field has a noisy character indicative of the superposition of many incoherent sources. In fact, the data has been successfully accounted for by a model invoking only scattering from a stochastic distribution of local impedance discontinuities.[27] Also note that there are frequency bands in which the far-field levels are enhanced. A possible explanation for this phenomena, particularly the very strong band in the 10–20-kHz range, is radiation from large-n flexural Bloch modes which are allowed to radiate because of the presence of the non-axisymmetry. That is, the previously nonradiating large-n modes excited to fairly high levels in all the previous shells have now developed a radiating component. Note that the shell with no internals also shows some signature in this region, perhaps indicating that the very weak deviations from axisymmetry in the scale model may be coupling these same modes into radiation. Considering the far-field result together with the near-field measurement in Fig. 5 one may fairly conclude that the excitation of coherent extended modes, if it exists at all, is weak.

We have not observed a substantial increase in the internal damping of the structure due to the attached oscillators. Note that the radiated noise levels for the shell with oscillators [Fig. 6(b)] are increased relative to the levels for

the simple framed shell [Fig. 6(a)]. This result indicates that the radiation damping of the resonant modes has increased and that any increase in the internal damping of the system is a much smaller effect. The scattering data leads one to a similar conclusion.[25] The number of internal oscillators in the experiment is evidently too small to give rise to the apparent damping effects discussed by Soize and others.[1–3]

An additional insight can be gleaned from an examination of the response of some large-n azimuthal modes. Plots of the response of the $n=18,23,27,31$ modes are shown in Fig. 7 compared to the response of the corresponding modes on the ''simple'' irregularly framed shell. One must exercise care in drawing too many conclusions from this figure, since on the ''simple'' shell one is observing individual resonant modes while on the oscillator shell the response at a given frequency may well contain many resonant modes. Nevertheless, the similarity in the responses indicates that the large-n localized modes on the unframed shell persist in the shell with internal oscillators. It is hence quite likely that these modes are relevant players on any framed shell—an important piece of information.

The structural acoustic effects of adding the internal oscillators to the irregularly framed shell are quite significant. In terms of the acoustic far field, two important phenomena are observed: First, the bulk of the modes, including many of the previously nonradiating flexural resonances, now radiate; second, the clear signature resulting from the extended wave propagation of the small-n modes has been destroyed. In the near field, the localization has been increased by the additional complexity, a phenomena probably resulting from the destruction of the propagating small-n modes.

V. APPLICATION TO STRUCTURAL ACOUSTIC DESIGN

Internal complexity has been seen to give rise to several phenomena, some which are positive from the perspective of noise control and others which are negative. A dominant effect apparent in the experimental results is that internals which break the symmetry of the system generally increase the coupling of the structure to acoustic waves and are thus detrimental. On the other hand, internal complexity at the same time disturbs large-scale acoustic wave phenomena, including both elastic wave propagation and specular reflection, and hence reduces the far-field signatures associated with these processes. Further, we have found that even a small amount of internal complexity produces localization of a significant number of resonances of the structure, another positive aspect from an acoustic point of view. This understanding has implications for a number of applications in structural acoustics including quiet acoustic designs, acoustic treatments, and hull mounted sensor systems.

Suppose one is able to start from scratch and asks, how should a quiet underwater structure be designed? How much complexity should the structure have? Defining ''quiet'' as minimizing the acoustic coupling to the structure, the answer to this question is clear for a system with a small loss factor: The quiet, lightly damped, shell structure ideally should have no internals which break the symmetry of the system. For example, in the model experiments discussed in the preceding sections, all of which had small internal damping, the ''best'' model is the uniform shell. The addition of each level of internal complexity increased the overall far-field acoustic radiation and scattering levels. By preserving the symmetry of system as much as possible, axially and azimuthally, the minimum number of resonant modes are coupled to the fluid, and therefore the overall acoustic scattering and radiation levels will be minimized. We note that a similar conclusion could have been reached based on the concept that discontinuities generally increase the radiation resistance of the system.[10]

This conclusion raises an interesting point regarding the contribution to scattering and radiation from the well-known membrane and flexural Bloch waves. The acoustic highlights in angle and frequency associated with these structure-borne waves, so clear in the scattering from simple framed shells, are no longer apparent in acoustic far field of the shell with many internal oscillators. The internals are evidently disturbing the large-scale, coherent wave processes which give rise to the far-field acoustic signatures. Nevertheless, the competing tendency of the internals to increase the coupling of elastic surface vibration to radiating acoustic waves outweighs this benefit in lightly damped shell structures, and the overall scattering and radiation levels are therefore increased by internal complexity.

The presence of significant internal damping mechanisms in the structure can lead us to a different conclusion. In general, only those modes whose radiation damping is of the order of or exceeds the internal damping significantly couple the system to acoustic waves in the surrounding fluid. As we increase the internal complexity of a nominally uniform system, the radiation efficiency of the majority of the modes increases as we have seen, but provided the internal damping exceeds the radiation damping, these modes will remain essentially uncoupled from bulk acoustic waves in the surrounding fluid. At the same time, coherent processes, including not only wave propagation but also specular reflections, will be inhibited by the increased structural complexity and the acoustic signatures resulting from these processes will be reduced. The degree to which this may be accomplished of course depends on the amount of the internal damping, but in this scenario, internal complexity will improve the acoustic performance of the system, despite breaking the symmetry of the shell structure. Thus, the opposite conclusion regarding the effects of complexity has been reached, provided the internal damping of the structure is large enough.

These concepts present an opportunity for improved acoustic design of underwater structures. Irregularity in shell structures produces localization of a significant number of resonances of the system. The corresponding reduction of the structural acoustic coupling along the structure is desirable because it results in a reduction of the cross couplings in distributed sensing or control systems. The resulting changes in the far-field acoustic signature will also be beneficial provided the increased radiation damping of the majority of the modes remains smaller than the internal damping. And even in cases in which this criteria is violated, the resulting increases in the acoustic signature will be associated with local

vibrational behavior and hence easier to reduce or control than phenomena associated with extended wave propagation.

VI. CONCLUSION

We have presented some of the key results from our research effort exploring the impact of complexity on the structural acoustics of marine structures. The results presented here have allowed us to develop an overall picture of the changes in the vibro-acoustic response which occur in shell structures as the complexity of the system is increased.

Generally, the effects of the increasing complexity are to increase the spatial localization of vibrational energy and to increase the number of resonant modes which are well coupled to acoustic waves in the surrounding fluid. Some of the most important conclusions we have drawn are (1) nonaxisymmetric complexity significantly decreases the impact of large-scale resonances, while small-scale resonances persist; (2) a very small amount of irregularity can produce significant modifications (localization) of the acoustic near field without necessarily altering the acoustic far field; and (3) resonances previously uncoupled to the fluid, predominantly highly resonant large-*n* modes, are now, as a result of the increased complexity, coupled to the acoustic radiation field. The effects of complexity of the sort which occurs on actual marine structures are still unknown, but we expect that at least some of the results we have found will persist.

ACKNOWLEDGMENT

This research was supported by the Office of Naval Research.

[1] C. Soize, ''Probabilistic structural modeling in linear dynamic analysis of complex mechanical systems I. Theoretical elements,'' J. Recherche Aerospat. ONERA France **5**, 23–48 (1986).
[2] M. Strasburg and D. Feit, ''Vibration damping of large structures induced by attached small resonant structures,'' J. Acoust. Soc. Am. **99**, 335–344 (1996).
[3] A. D. Pierce, V. W. Sparrow, and D. A. Russell, ''Fundamental structural acoustic idealizations for structures with fuzzy internals,'' Trans. ASME, J. Vib. Acoust. **117**, 339–348 (1995).
[4] M. C. Junger and D. Feit, *Sound, Structures, and Their Interaction*, 2nd ed. (MIT, Cambridge, MA, 1972).
[5] R. N. Arnold and G. B. Warburton, ''Flexural vibrations of the walls of thin cylindrical shells having freely supported ends,'' Proc. R. Soc. London, Ser. A **197**, 238–256 (1949).
[6] J. F. M. Scott, ''The free modes of vibration of an infinite fluid-loaded cylindrical shell,'' J. Sound. Vib. **125**, 241 (1988).
[7] E. G. Williams, ''Supersonic acoustic intensity,'' J. Acoust. Soc. Am. **97**, 121–127 (1994).
[8] N. Ashcroft and D. Mermin, *Solid State Physics* (Saunders College, Philadelphia, PA, 1976), pp. 175–213, employs a similar approach in a condensed matter physics context.
[9] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), p. 761.
[10] G. Maidanik, ''Response of ribbed panels to reverberant acoustic fields,'' J. Acoust. Soc. Am. **34**, 809 (1962).
[11] M. L. Rumerman, ''Contribution of membrane wave reradiation to scattering from finite cylindrical steel shells in water,'' J. Acoust. Soc. Am. **93**, 55–65 (1993).
[12] L. Brillioun, *Wave Propagation in Periodic Structures* (Dover, New York, 1946).
[13] D. M. Photiadis, J. A. Bucaro, and B. H. Houston, ''Scattering from flexural waves on a ribbed cylindrical shell,'' J. Acoust. Soc. Am. **96**, Pt. 1, 2785–2790 (1994).
[14] B. H. Houston, J. A. Bucaro, and D. M. Photiadis, ''Broadband acoustic scattering from a ribbed shell,'' J. Acoust. Soc. Am. **98**, Pt. 1, 2851–2853 (1995).
[15] I. M. Lifshits, S. A. Gredeskul, and L. A. Pastur, *Introduction to the Theory of Disordered Systems* (Wiley, New York, 1988).
[16] C. H. Hodges and J. Woodhouse, ''Vibration isolation from irregularity in a nearly periodic structure: Theory and measurements,'' J. Acoust. Soc. Am. **74**, 894–905 (1983).
[17] D. M. Photiadis, ''Localization of helical flexural waves by irregularity,'' J. Acoust. Soc. Am. **96**, 2291–2301 (1994).
[18] D. C. Herbert and R. Jones, ''Localized states in disordered systems,'' J. Phys. C **4**, 1145–1161 (1971).
[19] H. L. Frisch and S. P. Lloyd, ''Electron levels in a one-dimensional random lattice,'' Phys. Rev. **120**, 1175–1189 (1960).
[20] R. E. Borland, ''The nature of electronic states in disordered one-dimensional systems,'' Proc. R. Soc. London, Ser. A **274**, 529 (1963).
[21] P. W. Anderson, ''Absence of diffusion in certain random lattices,'' Phys. Rev. **109**, 1492–1505 (1958).
[22] D. M. Photiadis and B. H. Houston, ''Anderson localization of vibration on a framed fluid-loaded cylinder,'' J. Acoust. Soc. Am. **106**, Pt. 1, 1377 (1999).
[23] D. M. Photiadis, ''The effect of irregularity on the scattering of acoustic waves from a ribbed plate,'' J. Acoust. Soc. Am. **91**, 1897–1903 (1992).
[24] P. Sheng, *Introduction to Wave Scattering, Localization, and Mesoscopic Phenomena* (Academic, San Diego, CA, 1995).
[25] D. M. Photiadis, J. A. Bucaro, and B. H. Houston, ''The effect of internal oscillators on the acoustic response of a submerged shell,'' J. Acoust. Soc. Am. **101**, 895 (1997).
[26] J. A. Bucaro, D. M. Photiadis, and B. H. Houston, ''Acoustic scattering from a submerged shell with many internal oscillators,'' Proceedings of the ASME, Noise Control and Acoustics Division, NCA **22**, 87–92 (1996).
[27] J. A. Bucaro, A. J. Romano, A. Sarkissian, D. M. Photiadis, and B. H. Houston, ''Local admittance model for acoustic scattering from a cylindrical shell with many internal oscillators,'' J. Acoust. Soc. Am. **103**, 1867 (1998).

Sensor scheme design for active structural acoustic control

Arthur P. Berkhoff[a)]
TNO Institute of Applied Physics, P.O. Box 155, 2600 AD Delft, The Netherlands

(Received 16 August 1999; accepted for publication 24 April 2000)

Efficient sensing schemes for the active reduction of sound radiation from plates are presented based on error signals derived from spatially weighted plate velocity or near-field pressure. The schemes result in near-optimal reductions as compared to weighting procedures derived from eigenvector or singular vector analysis of the radiation operator. Efficient control configurations are suggested using a, possibly analog, front-end implementing a bank of spatial weighting functions and a digital controller with a minimized number of input and output channels. The performance of different weighting functions is compared, as well as the performance of different frequency-dependent filtering functions. Design rules are given for the sensor spacing, the number of weighting functions, the number of actuators, and the corresponding controller dimensionality. © *2000 Acoustical Society of America.* [S0001-4966(00)02708-9]

PACS numbers: 43.40.Vn [PJR]

INTRODUCTION

Near-field sensing to predict acoustic radiation into free or enclosed spaces is often desirable in active structural acoustic control, in which structural actuators are used to reduce sound transmission. Placing sensors at a relatively large distance from the radiating structure for measurement of the total radiated sound power is not always possible and can deteriorate performance because of the delay between the actuator and sensor. One approach to reducing acoustic radiation is to use accelerometers for sensing the vibration levels. However, the reduction of vibration levels may lead to disappointing results[1] in terms of radiated sound power. The discrete wave number transform has been suggested as a solution to this problem and good results have been obtained in feedforward configurations.[2–5] In this method, signals from an array of structural sensors are delayed and subsequently combined to arrive at a decomposition of plane wave constituents, each of which has an associated propagation direction.

Sensing strategies using velocity information can also be based on so-called radiation modes,[6] which are the vibration patterns of a structure that radiate sound independently in free space[7–13] or in enclosed spaces.[14,15] The resulting control strategy is optimal with respect to the number of error signals for the controller. An advantage is that the schemes use real-valued spatial weighting functions without a delay. A disadvantage of the use of radiation modes is that the vibration patterns corresponding to the radiation modes depend on frequency, leading to a complicated control strategy involving spatio-temporal filtering. It has been demonstrated that the pertinent vibration patterns can be chosen to be independent of frequency.[16]

The use of frequency-independent spatial filters has also been suggested by Baumann *et al.*[17] Their approach, however, was based on implementing frequency-dependent radiation efficiencies of *structural* modes and the interaction between these modes, where the frequency-dependent part was implemented as temporal filters operating on the output of the spatial weighting functions, i.e., the structural modes. In many cases, particularly with thin plates, this is a rather inefficient procedure, and it may be more efficient to use spatial filters defined by *acoustic* modes. Other methods based on radiation modes were presented by Sommerfeldt and Lin,[18] who proposed an implementation using radiation mode sensing based on at least two sensors for each radiation mode.

The method for velocity sensing as used in this article was implicit in the of Baumann[17] and Borgiotti.[16] In the present article a radiation mode weighting scheme is also derived for near-field pressure sensing,[31] taking into account the smoothness of the weighting functions. The fixed spatial weighting functions are based on the eigenfunctions of the acoustic radiation operator while frequency-dependent recombination functions are used to reconstruct the exact radiation modes up to a predetermined maximum frequency. The spatial filters can be implemented in analog hardware and provide a reduction of the number of input channels to a digital controller. The frequency-dependent filtering can be implemented on the digital controller. A possible configuration for such a system has already been presented.[19] The present article gives more details on the methods and on the design parameters that influence the performance, both for velocity sensing and for near-field pressure sensing. The choice for the fixed spatial functions is investigated as well as the influence of the frequency-dependent recombination functions.

I. SENSING METHODS

A. Radiated power

The radiated power Π is given by

$$\Pi = \frac{1}{2}\mathrm{Re}\left\{ \int_{\mathbf{x}\in S} v^*(\mathbf{x}) p(\mathbf{x})\, d\mathbf{x} \right\}, \tag{1}$$

with $\mathbf{x}$ a spatial coordinate on the radiator, p the pressure, and v the velocity, where the asterisk is the complex conju-

a)Electronic mail: berkhoff@tpd.tno.nl

gate operator. As is discussed in Ref. 6, the sound radiation can be assumed to be due to a number of elemental radiators, which in this case are equivalent to small piston sources on a surface. Let $\mathbf{p}$ and $\mathbf{v}$ denote the vectors of the pressure p and the normal velocity v, respectively, and let $\mathbf{Z}$ be an impedance matrix. Then a relationship between pressure and velocity on the radiator surface can be formulated as[6]

$$\mathbf{p}=\mathbf{Z}\mathbf{v}, \tag{2}$$

and the radiated power of Eq. (1) as

$$\Pi=\frac{S}{2}\mathrm{Re}\{\mathbf{v}^H\mathbf{p}\}, \tag{3}$$

in which S is the area of each of the elemental radiators corresponding to the elements of $\mathbf{v}$ and $\mathbf{p}$, and H denotes the Hermitian transpose operator.

B. Velocity sensing

The radiated acoustic power at angular frequency ω can be formulated as[6]

$$\Pi(\omega)=\mathbf{v}^H(\omega)\mathbf{R}(\omega)\mathbf{v}(\omega), \tag{4}$$

with $\mathbf{v}(\omega)$ the vector of velocities of the radiating structure and $\mathbf{R}(\omega)$ the radiation matrix, which is real, symmetric, and positive definite. For radiators in a plane baffle this radiation matrix can be calculated analytically:

$$\mathbf{R}=\frac{\omega^2\rho S^2}{4\pi c}\begin{bmatrix} 1 & \frac{\sin(kr_{12})}{kr_{12}} & \cdots & \frac{\sin(kr_{1N})}{kr_{1N}} \\ \frac{\sin(kr_{21})}{kr_{21}} & 1 & & \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\sin(kr_{N1})}{kr_{N1}} & & & 1 \end{bmatrix}, \tag{5}$$

in which $r_{mn}=r_{nm}$ is the distance from element m to element n, $1\leqslant\{m,\ n\}\leqslant N$, and $k=\omega/c$ is the wave number with c the speed of sound. Further, ρ is the density of the acoustic medium.

Upon an eigen-decomposition of the radiation matrix

$$\mathbf{R}(\omega)=\mathbf{\Sigma}_\mathbf{R}^T(\omega)\mathbf{\Lambda}_\mathbf{R}(\omega)\mathbf{\Sigma}_\mathbf{R}(\omega), \tag{6}$$

where $\mathbf{\Sigma}_\mathbf{R}(\omega)$ is a real matrix containing a radiation mode shape in each row, and where $\mathbf{\Lambda}_\mathbf{R}(\omega)$ is a real diagonal matrix containing the radiation mode efficiencies, the expression for the radiated power becomes

$$\Pi(\omega)=\boldsymbol{\nu}_\mathbf{v}^\mathrm{H}(\omega)\mathbf{\Lambda}_\mathbf{R}(\omega)\boldsymbol{\nu}_\mathbf{v}(\omega), \tag{7}$$

where

$$\boldsymbol{\nu}_\mathbf{v}(\omega)=\mathbf{\Sigma}_\mathbf{R}(\omega)\mathbf{v}(\omega) \tag{8}$$

is the vector containing the amplitudes of the different radiation modes, i.e., the vector $\boldsymbol{\nu}_\mathbf{v}(\omega)$ contains the inner products of the radiation mode shape vectors with the velocity vector.

A related approach is based on fixed *frequency-independent* spatial weighting functions $\mathbf{E}$, in this article termed *elementary radiation shapes*. These functions are required to be complete with respect to radiated acoustic power[16] for all radiation mode shapes below a certain frequency $f_{\max}$. Then, at each frequency smaller than $f_{\max}$, the dominant radiation modes can be synthesized by a linear combination of the elementary radiation shapes. Since, in general, radiation mode shapes depend on frequency, the weighting factors involved in the linear combination are also functions of frequency. Let the weighting factors be the coefficients of the matrix $\mathbf{F}(\omega)$. Then

$$\mathbf{\Sigma}_\mathbf{R}(\omega)=\mathbf{F}(\omega)\mathbf{E}. \tag{9}$$

Using Eqs. (7)–(9), the approximation $\tilde{\Pi}(\omega)$ of the radiated power $\Pi(\omega)$ for frequencies smaller than $f_{\max}$ becomes

$$\tilde{\Pi}(\omega)=\mathbf{v}^H(\omega)\mathbf{E}^T\mathbf{F}^H(\omega)\mathbf{\Lambda}_\mathbf{R}(\omega)\mathbf{F}(\omega)\mathbf{E}\mathbf{v}(\omega). \tag{10}$$

Introducing the modified radiation mode strengths

$$\tilde{\boldsymbol{\nu}}_\mathbf{v}(\omega)=\mathbf{E}\mathbf{v}(\omega) \tag{11}$$

and the modified efficiency matrix

$$\tilde{\mathbf{\Lambda}}_\mathbf{R}(\omega)=\mathbf{F}^H(\omega)\mathbf{\Lambda}_\mathbf{R}(\omega)\mathbf{F}(\omega), \tag{12}$$

we arrive at

$$\tilde{\Pi}(\omega)=\tilde{\boldsymbol{\nu}}_\mathbf{v}^H(\omega)\tilde{\mathbf{\Lambda}}_\mathbf{R}(\omega)\tilde{\boldsymbol{\nu}}_\mathbf{v}(\omega). \tag{13}$$

With this equation, we have obtained the desired spatial weighting for the velocity signals $\mathbf{v}(\omega)$ through the frequency-independent $\mathbf{E}$ to obtain $\tilde{\boldsymbol{\nu}}_\mathbf{v}(\omega)$, followed by a frequency-dependent reconstruction of the radiation efficiency matrix by the matrix $\mathbf{F}(\omega)$ to obtain $\tilde{\mathbf{\Lambda}}(\omega)$. If the functions in $\mathbf{E}$ are orthonormal, then a right-multiplication of Eq. (9) with $\mathbf{E}^T$ gives an explicit expression for $\mathbf{F}(\omega)$:

$$\mathbf{F}(\omega)=\mathbf{\Sigma}_\mathbf{R}(\omega)\mathbf{E}^T. \tag{14}$$

A left multiplication of Eq. (11) with $\mathbf{E}^T$ and use of Eqs. (6), (10), and (14) shows that

$$\tilde{\mathbf{\Lambda}}_\mathbf{R}(\omega)=\mathbf{E}\mathbf{R}(\omega)\mathbf{E}^T. \tag{15}$$

It is straightforward to show that $\tilde{\mathbf{\Lambda}}_\mathbf{R}$ is positive definite, i.e., $\mathbf{x}^H\tilde{\mathbf{\Lambda}}_\mathbf{R}\mathbf{x}>0$, $\forall\mathbf{x}\neq\mathbf{0}$. In view of Eq. (13), it immediately follows that $\tilde{\Pi}=0\ \forall f<f_{\max}$ only if $\tilde{\boldsymbol{\nu}}_\mathbf{v}=\mathbf{0}$, $\forall f<f_{\max}$. As a consequence $\Pi=0\ \forall f<f_{\max}$, provided the elementary radiation shapes in $\mathbf{E}$ are complete for the propagating vibration patterns at $f<f_{\max}$. In other words the radiated power can be substantially reduced for all frequency components if a linear combination of the shapes in $\mathbf{E}$ can give a good approximation of the strongest radiation modes of $\mathbf{\Sigma}_\mathbf{R}$.

C. Pressure sensing

In this subsection it is shown that a radiation mode formulation also can be obtained for near-field pressure sensing. The explicit notation of the dependence on ω is omitted in the following. The starting point is again Eq. (3). In order to arrive at an expression for $\mathbf{v}$, Eq. (2) is formulated as the inverse operation

$$\mathbf{v}=\mathbf{Y}\mathbf{p}, \tag{16}$$

with $\mathbf{Y}$ the inverse of $\mathbf{Z}$. For radiators in a baffle the impedance matrix $\mathbf{Z}$ can be written as

$$\mathbf{Z}=\rho c\begin{bmatrix} 1-e^{ik\sqrt{S/\pi}} & -\frac{ikS}{2\pi}\frac{e^{ikr_{12}}}{r_{12}} & \cdots & -\frac{ikS}{2\pi}\frac{e^{ikr_{1N}}}{r_{1N}} \\ -\frac{ikS}{2\pi}\frac{e^{ikr_{21}}}{r_{21}} & 1-e^{ik\sqrt{S/\pi}} & & \\ \cdots & & & \\ -\frac{ikS}{2\pi}\frac{e^{ikr_{N1}}}{r_{N1}} & & & 1-e^{ik\sqrt{S/\pi}} \end{bmatrix}. \tag{17}$$

The techniques which are used[20] for deriving this equation are similar to the boundary element method[21] using analytically evaluated singularities. The expression for the radiated power becomes

$$\Pi=\frac{S}{4}\mathbf{p}^H(\mathbf{Y}^H+\mathbf{Y})\mathbf{p}=\mathbf{p}^H\mathbf{G}\mathbf{p}, \tag{18}$$

with $\mathbf{G}=(S/2)\mathrm{Re}\,\mathbf{Y}$. The inversion of the expression of $\mathbf{Z}$ in Eq. (17) to arrive at $\mathbf{Y}$ poses no numerical difficulties because the matrix $\mathbf{Z}$ is well behaved. However, due to the truncation of the analysis domain, the eigenvectors of $\mathbf{G}$ contain spatial high-frequency components at the edges of the computational domain, especially at low frequencies. The edge contributions of the radiation modes at temporal high frequencies are smaller. In many cases large edge effects are undesirable. In order to reduce these edge effects and increase smoothness, a cost function ϕ is introduced which contains two terms. The first term involving $\mathbf{v}-\mathbf{Y}\mathbf{p}$ corresponds to the standard solution of the inversion. The second term penalizes large values of $\mathbf{p}$. The cost function ϕ then can be expressed as

$$\phi=(\mathbf{v}-\mathbf{Y}\mathbf{p})^H(\mathbf{v}-\mathbf{Y}\mathbf{p})+\gamma\mathbf{p}^H\mathbf{p}, \tag{19}$$

with γ a suitably chosen parameter to balance the relative importance of the two cost terms. The latter equation can be written as

$$\phi=\mathbf{p}^H\mathbf{A}\mathbf{p}+\mathbf{p}^H\mathbf{b}+\mathbf{b}^H\mathbf{p}+c \tag{20}$$

with

$$\mathbf{A}=\mathbf{Y}^H\mathbf{Y}+\gamma\mathbf{I}, \tag{21}$$

$$\mathbf{b}=-\mathbf{Y}^H\mathbf{v}, \tag{22}$$

$$c=\mathbf{v}^H\mathbf{v}. \tag{23}$$

The minimum of Eq. (20) is obtained for $\mathbf{A}\mathbf{p}=-\mathbf{b}$, which leads to

$$(\mathbf{Y}^H\mathbf{Y}+\gamma\mathbf{I})\mathbf{p}=\mathbf{Y}^H\mathbf{v}. \tag{24}$$

Inversion of the latter equation yields

$$\mathbf{v}=\mathbf{Y}_\gamma\mathbf{p} \tag{25}$$

with

$$\mathbf{Y}_\gamma=[(\mathbf{Y}^H\mathbf{Y}+\gamma\mathbf{I})^{-1}\mathbf{Y}^H]^{-1}. \tag{26}$$

The corresponding expression for the radiated power can be written as

$$\Pi=\mathbf{p}^H\mathbf{G}_\gamma\mathbf{p}, \tag{27}$$

with $\mathbf{G}_\gamma=(S/2)\mathrm{Re}\,\mathbf{Y}_\gamma$. By using eigenvector analysis, the matrix $\mathbf{G}_\gamma$ can be diagonalized as in the preceding section:

$$\mathbf{G}_\gamma=\mathbf{\Sigma}_\mathbf{G}^T\mathbf{\Lambda}_\mathbf{G}\mathbf{\Sigma}_\mathbf{G}. \tag{28}$$

As with velocity sensing, the expression for the radiated power becomes

$$\Pi=\boldsymbol{\nu}_\mathbf{p}^H\mathbf{\Lambda}_\mathbf{G}\boldsymbol{\nu}_\mathbf{p}, \tag{29}$$

where

$$\boldsymbol{\nu}_\mathbf{p}=\mathbf{\Sigma}_\mathbf{G}\mathbf{p} \tag{30}$$

is the vector containing the strengths of the different radiation mode error signals. The method is related to Tikhonov regularization, such as discussed by Nelson.[22]

Small values of γ already lead to a considerable smoothing of the shapes of the radiation modes in $\mathbf{\Sigma}_\mathbf{G}$, as will be seen in the results. As with velocity sensing, pressure sensing can be combined with frequency-independent spatial weighting functions to arrive at simple sensing schemes with a relatively small number of error signals.

II. SIMULATION RESULTS

In this section, results are given of a feedforward control configuration using a simply supported steel plate of 60 cm×75 cm×2 mm. The configuration is used to reduce the sound transmitted through the plate with piezoelectric patch actuators and velocity or pressure sensors. The sensors and the piezoelectric actuators are assumed to be positioned on a regular grid with equal separation distances in each of the two in-plane directions.

A. Cost functions

In the control algorithm, which can be an iterative gradient descent method such as a (least-mean-square[1]) LMS-type algorithm, a small value of effort weighting was used to ensure convergence to a global minimum for underdetermined systems having less input channels than output channels. The cost functions are

$$J=\boldsymbol{\nu}^H\mathbf{\Lambda}\boldsymbol{\nu}+\beta\mathbf{u}^H\mathbf{u} \tag{31}$$

for the exact frequency-dependent radiation mode shapes,

$$J=\tilde{\boldsymbol{\nu}}^H\tilde{\mathbf{\Lambda}}\tilde{\boldsymbol{\nu}}+\beta\mathbf{u}^H\mathbf{u}, \tag{32}$$

for the frequency-independent spatial weighting functions, and

$$J=\tilde{\boldsymbol{\nu}}^H\tilde{\boldsymbol{\nu}}+\beta\mathbf{u}^H\mathbf{u} \tag{33}$$

for the frequency-independent spatial weighting functions with no additional frequency-dependent weighting. The value of β was set to 10^{-6} times the diagonal-mean of the Hessian matrix.

B. Influence of the numbers of actuators, sensors, and radiation modes on performance

This subsection determines the individual influence of the number of sensors, the number of radiation modes, and the number of actuators on the reduction of radiated sound

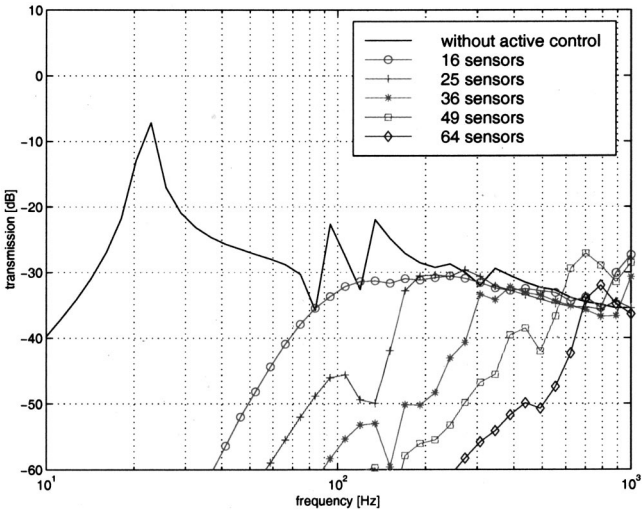


FIG. 1. Transmission of sound power through a steel plate of 2-mm thickness using 36 actuators and the radiated power as error criterion obtained from a varying number of accelerometers.

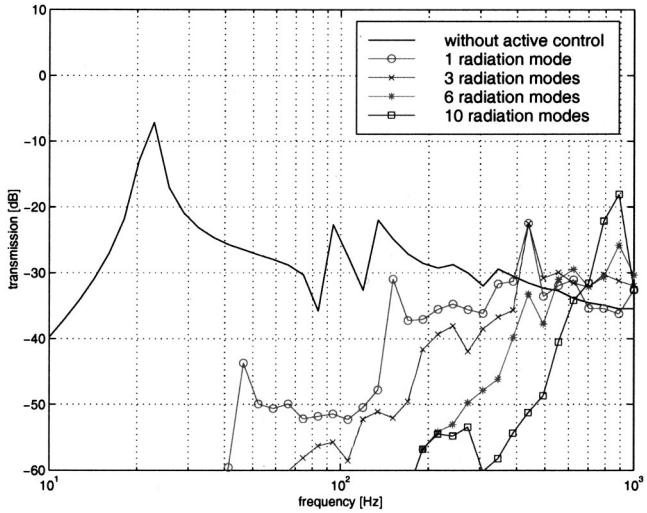


FIG. 2. As in Fig. 1, using 36 actuators, 64 sensors, and a varying number of radiation modes.

power. The cost function of Eq. (31) was used, of which the first part corresponds to the radiated sound power if all radiation modes are taken into account.

1. Sensors

The estimation of the structural vibration is governed by the sampling theorem: at least two sensors per structural bending wavelength $\lambda_B = 2\pi/k_B$ are required, with k_B the structural wave number. Assuming at least two sensors per structural wave length, the requirement for the sampling distance Δx becomes $\Delta x \lesssim \pi/k_B$. For a plate, the wave number for bending waves equals

$$k_B = \sqrt{\frac{\omega}{h}}\left(\frac{12\rho(1-\nu^2)}{E}\right)^{1/4} \tag{34}$$

with h the plate thickness, ρ the density of the plate, E Young's modulus, and ν the Poisson ratio. The parameters of the steel plate as used in the simulations are $\rho = 7.8 \times 10^3\,\mathrm{kg\,m^{-3}}$, $E = 2\times 10^{11}\,\mathrm{N\,m^{-2}}$, $\nu = 0.3$, and $h = 2$ mm. Figure 1 shows the performance for different numbers of sensors which are used to estimate radiated power; in this case a relatively large number of actuators is used (6×6). A description of the actuator placement is given in a subsequent section. From a comparison of the spacing between the sensors for a particular configuration and the structural wavelength at the maximum frequency where reductions can be obtained for that configuration, it can be concluded that slightly more than two acceleration sensors per structural wavelength are required. Because of the diagonal dominance of $\mathbf{Z}$ in Eq. (17) the design rules for velocity sensing also apply to near-field pressure sensing.

2. Radiation modes

The relationship between the number of radiation modes and the estimation of acoustic radiation for general geometries is similar to that for the separable geometry in the Appendix. For nonseparable geometries, analytical results may be impossible, but the relationship between the number of radiation modes and the ratio of wavelength to physical dimensions is similar. The magnitude of the spatial frequency content of the radiation modes has some resemblance with the ''beams'' of wave vector filtering.[8] Therefore, an approximation for the number of required radiation modes can be based on the number of propagating beams. For the rectangular geometry as used in the examples, we introduce the integer mode indexes $n_x \geqslant 0$ and $n_y \geqslant 0$ in the x and y directions, respectively. The radiation modes that are required to achieve sound power reductions for frequencies $f < f_{\max}$ are at least the combinations (n_x, n_y) that satisfy

$$\left(\frac{\pi n_x}{l_x}\right)^2 + \left(\frac{\pi n_y}{l_y}\right)^2 \lesssim k_{\max}^2, \tag{35}$$

with l_x and l_y the dimensions of the radiator and $k_{\max} = 2\pi f_{\max}/c$. Based on the expression (35), an explicit estimation of the number of radiation modes N can be obtained from a technique similar to that of Maa:[23]

$$N \approx \frac{1}{4} + \frac{k_{\max}}{2\pi}(l_x + l_y) + \frac{l_x l_y}{4\pi}k_{\max}^2. \tag{36}$$

Figure 2 shows the performance for a different number of radiation modes when a relatively large number of error sensors (10×10) and a relatively large number of actuators (6×6) are used. The results are in agreement with Eqs. (35) and (36).

The number of required radiation modes as given above is consistent with sampling requirements for continuous wave number transforms of acoustic radiation operators. For 2D geometries it has been shown[20] that the number of modes required equals $N = kL/\pi$, with L the length of the geometry. The number N equals the number of propagating modes corresponding to the spatial sampling and the wave number sampling[20] as defined by the discrete Fourier transform.[24] It is also interesting to compare the number of radiation modes with the number of modes for the discrete wave number transform,[2] as given by Wang.[25] Using the notation as above, Wang requires $N = kL/\pi + 1$ for 2D geometries. The difference of one is due to rounding and the additional require-

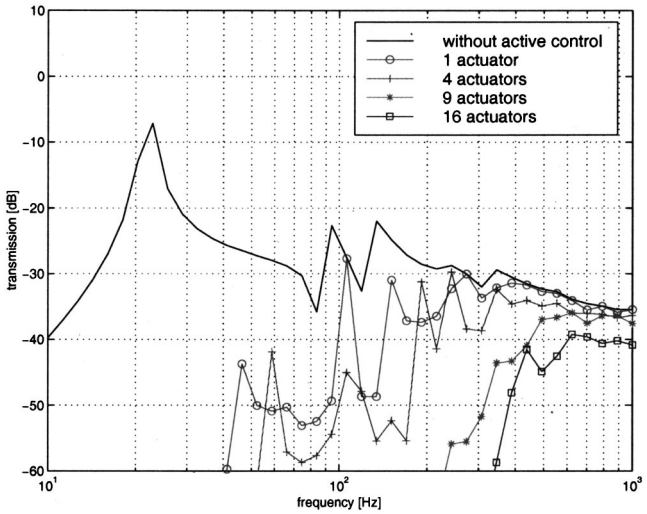


FIG. 3. As in Fig. 1, using a radiated power criterion computed from the plate velocity at 10×10 positions.

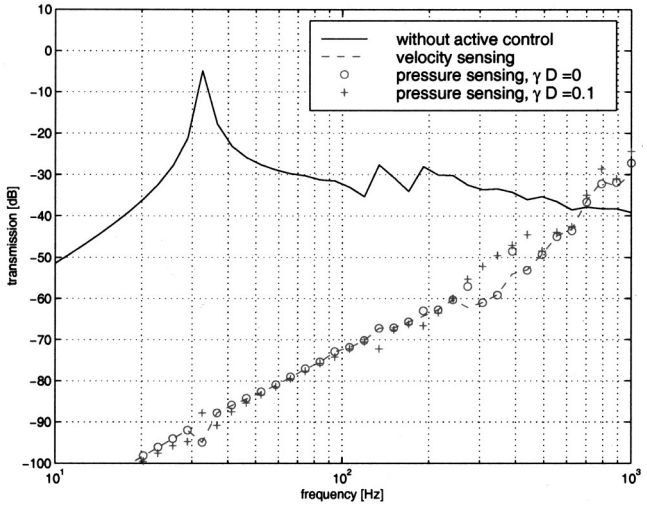


FIG. 4. Sound power transmitted through a plate without control and with active control for velocity sensing and pressure sensing.

ment in Wang's calculation to have at least one mode at low frequencies. Hence, for planar rectangular geometries, the required number of signals to be minimized by the controller is equal for radiation mode sensing, discrete Fourier transform techniques, and discrete wave number transforms.

3. Actuators

In an efficient system the degrees of freedom to be controlled by the actuators are approximately equal to the degrees of freedom measured by the error sensor. By virtue of their independence, radiation modes constitute the degrees of freedom of a sensor measuring radiated power. Hence, the number of actuators should be approximately equal to the number of radiation modes, provided certain reachability requirements[26] of the actuator configuration are fulfilled, i.e., the actuator configuration should be able to control those radiation modes. The actuator placement was based on the following rule: the actuators were taken to be as large as possible under the conditions that the actuators were not placed directly adjacent to each other and not directly at the boundaries. In Sec. III an example is given for such an actuator configuration. Figure 3 shows the performance for different numbers of actuators in the case of a relatively large number of error sensors (10×10) and a radiated power criterion. Indeed, it can be seen that the performance for a certain number of actuators in Fig. 3 approximately corresponds to the performance for the same number of radiation modes in Fig. 2.

4. Velocity versus pressure sensing

A comparison of the performance for velocity sensing using $\mathbf{\Sigma_R}$ and $\mathbf{\Lambda_R}$ and pressure sensing using $\mathbf{\Sigma_G}$ and $\mathbf{\Lambda_G}$ is given in Fig. 4. The radiation modes corresponding to pressure sensing were computed without regularization (indicated by $\gamma D=0$), and with a regularization value γ equalling 0.1 times the diagonal mean of $\mathbf{Y}^H\mathbf{Y}$ (indicated by $\gamma D=0.1$).

It can be seen that pressure sensing leads to (very slightly) smaller reductions in the frequency range of 250 to 500 Hz as compared to velocity sensing. It can also be seen that the regularization has a small effect on the reduction. However, it does have a large smoothing effect on the shapes of the radiation modes, as can be seen in Fig. 5 for the strongest radiation mode at low frequencies. The differences between radiation-mode-based pressure sensing and radiation-mode-based velocity sensing were also small for other configurations. The performance for a given number of radiation modes was approximately equal for both quantities. In the following, therefore, only results of velocity sensors are shown.

C. Frequency-independent spatial weighting functions

The use of frequency-independent mode shapes to approximate frequency-dependent radiation modes has been suggested by Borgiotti and Jones.[16] The suggestion was to use the radiation mode shapes at the highest controlled frequency. The question is whether the radiation mode shapes at the highest frequency are the only shapes that can be used.

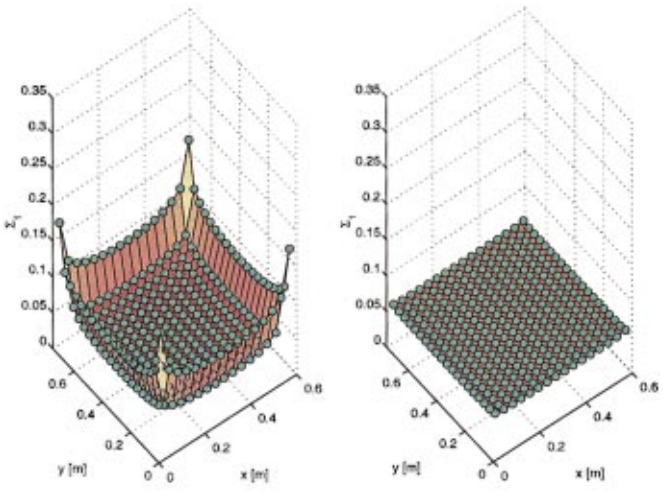


FIG. 5. Strongest pressure radiation mode at low frequencies without regularization (left) and with regularization (right).

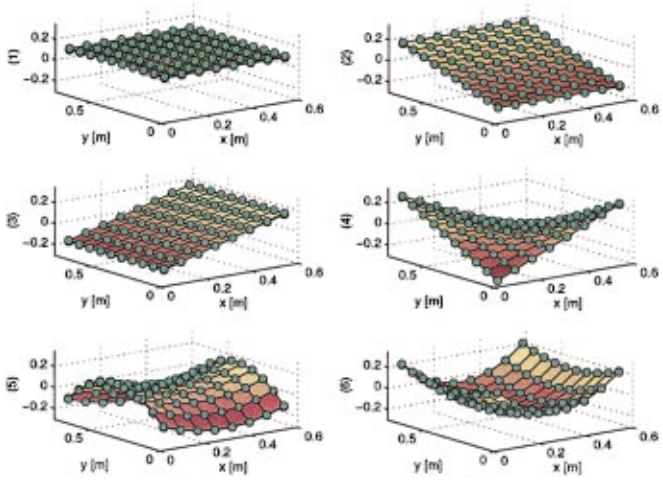


FIG. 6. Elementary radiation shapes **E** obtained as the radiation mode shapes $\mathbf{\Sigma}$ at $ka=0.1$.

The motivation in Ref. 16 to use the N propagating mode shapes at $f_{\max}$ is due to the nesting property in separable geometries. With these mode shapes, the acoustic far field can be fully reconstructed at any frequency $f<f_{\max}$. This same idea is then applied to nonseparable geometries, based on numerical calculation of the radiation modes. In view of the results of the Appendix, the numerically calculated radiation modes at low frequencies can also be considered. Although most of the latter modes are evanescent at low frequencies, they can still be useful at higher frequencies. Indeed, in the following it is shown that this is a viable alternative.

Two choices for the elementary radiation shapes were used: the radiation modes at low frequencies ($ka=0.1$, Fig. 6) and the radiation modes at high frequencies ($ka=3$, Fig. 7). Here, $a=(l_x l_y/\pi)^{1/2}$ is the effective radius and $ka=1$ corresponds to a frequency of 144 Hz. In addition, some choices for the frequency-dependent filtering are compared. A weighting according to Eq. (15), which accounts for the radiation efficiency of the elementary radiation shapes in **E**, would be an obvious choice. In Fig. 8, different choices for **E** and different choices for the frequency-dependent weighting $\tilde{\mathbf{\Lambda}}(\omega)$ are compared, using six elementary radiation shapes (i.e., $N=6$). Three choices of frequency-dependent weighting were used:

(1) Eq. (32) using the full matrix $\tilde{\mathbf{\Lambda}}(\omega)$ of Eq. (15);
(2) Eq. (32) using the diagonal of Eq. (15);[32] and
(3) Eq. (32) using the identity matrix for $\tilde{\mathbf{\Lambda}}(\omega)$, which corresponds to the use of Eq. (33).

The conclusions are as follows.

(i) The choice for **E** is not critical. At low frequencies, better results are obtained with the radiation mode shapes at low frequencies than with the radiation mode shapes at high frequencies. At high frequencies the differences are small.
(ii) The influence of frequency-dependent weighting by incorporating the radiation efficiencies in $\tilde{\mathbf{\Lambda}}(\omega)$ is negligible. This confirms the prediction in Sec. I that only $\tilde{\boldsymbol{\nu}}$ has to be driven to zero without frequency-dependent weighting by $\tilde{\mathbf{\Lambda}}(\omega)$.

Due to the latter conclusion, a sensor without any delay or inherent frequency-dependent processing can be arrived at by the use of the identity matrix for $\tilde{\mathbf{\Lambda}}$. This is an advantage if there is no time-advanced (feedforward) information of the disturbance signals available. In Fig. 8, results of only two choices for **E** were shown. Several other choices are possible. These choices are not critical provided all propagating wave number components below the maximum frequency can be represented by the particular choice of the elementary radiation shapes. A property of the radiation modes at high frequencies is that at low frequencies there may be more than one dominant volumetric mode. This is due to the fact that, in general, for arbitrary geometries, radiation modes are only independent at a single frequency.

If the elementary radiation shapes **E** equal the radiation modes at low frequencies, then the present approach can be seen as a generalization of the concept of reducing the volume velocity[27] of a radiating structure. The volume velocity corresponds to the strongest radiation mode at low frequencies. The present generalization extends the volume velocity approach to allow reductions at higher frequencies.

The number of structural sensors is larger than the number of radiation modes if the structural wavelength is smaller than the acoustic wavelength. This is the case for all frequencies below the critical frequency of the plate.[28] This means that for all frequencies below the critical frequency a reduction of the controller dimensionality is possible. Figure 9 shows a block diagram of such a system, with the digital controller depicted in Fig. 10. It is noted that the functions in **E** are real valued.[7] Therefore, simple weighting factors without memory can be used in the distribution network between the sensors and the controller.[16]

There are several possibilities for designing a radiation sensing scheme. The approach using frequency-independent spatial weighting functions has the advantage that for all radiator geometries, orthogonal radiation filter shapes can be determined; these shapes are not very critical as long as all

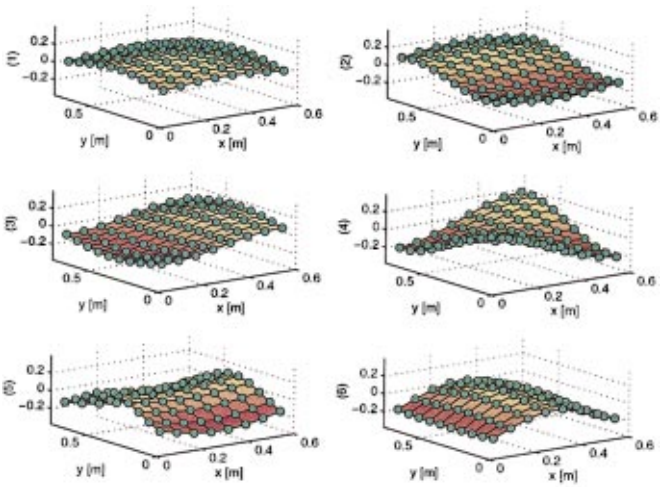


FIG. 7. As in Fig. 6, for $ka=3$.

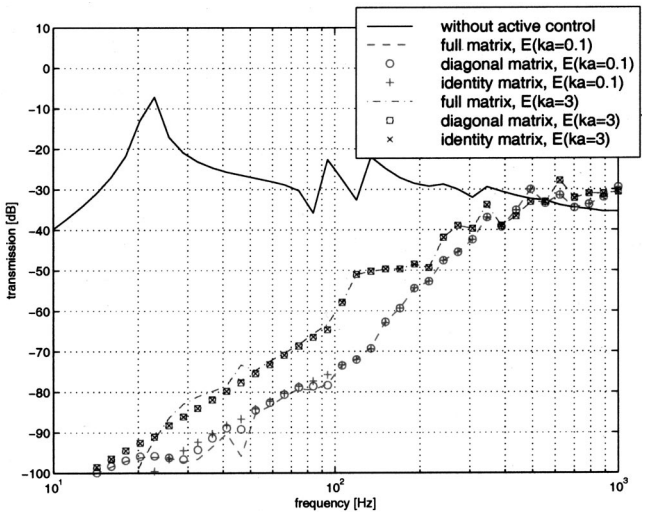


FIG. 8. Sound power transmitted through a plate without control and with active control for two different **E** and three different frequency-dependent weighting procedures.

propagating components are represented. In addition, frequency-dependent temporal weighting is not critical, at least not for free-field radiation. If radiation into enclosed spaces is considered, then frequency-dependent weighting can become important.[15] For free-field radiation, the corresponding radiation filters can be implemented with a simple frequency-independent weighting factor for each sensor to obtain the radiation filter outputs without any delay, as opposed to the method using the discrete wave number transform, which requires a delay for nearly all sensors. Simple analog front-end networks can be used to obtain the radiation filter outputs, leading to a reduction in computational burden for the digital controller, as compared to full spatio-temporal radiation mode filtering.

It would be interesting to reduce the number of sensors, such as with a sparse sensor configuration. However, without introducing additional knowledge of the system, the use of a sparsing configuration was found to be unsatisfactory.

III. EXAMPLE

An example is given for a steel plate of 60 cm×75 cm×3 mm. As compared to previous calculations, a thicker plate was used to illustrate the general applicability of the design rules. The maximum operating frequency $f_{\max}$ is 700

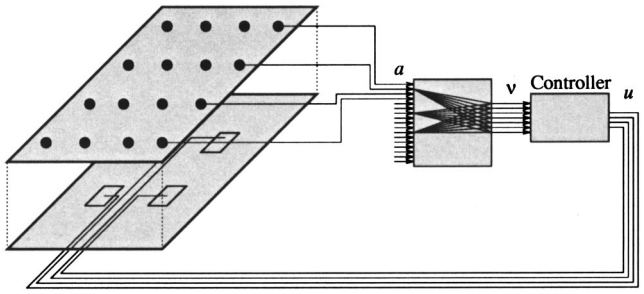


FIG. 9. Schematic diagram of control configuration implementing simplified radiation mode control strategy using piezoelectric actuators, accelerometers, adaptive controller, and a distribution network between sensors and controller. The sensors and actuators are mounted on the plate.

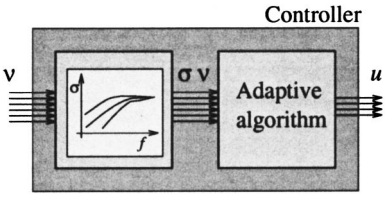


FIG. 10. Digital controller for simplified radiation mode control strategy consisting of filters and adaptive stochastic gradient algorithm.

Hz. The structural wave number at 700 Hz equals $k_B = 30.9\,\mathrm{m}^{-1}$, which leads to a maximum sensor spacing of $\Delta x = 0.10\,\mathrm{m}$. As a consequence the number of sensors was taken to be 7×8, uniformly distributed over the plate, which leads to spacings of 0.086 and 0.094 m, respectively. Using Eq. (35) or (36) with $k_{\max} = 12.8\,\mathrm{m}^{-1}$, we arrive at $N = 9$ error signals. The number of actuators should be approximately equal to N. With nine error signals and nine evenly distributed actuators the result was not exactly according to the specifications due to the cumulative effects on performance degradation. With a slightly more complicated system using 10 error signals and 12 evenly distributed actuators the result was according to the specifications. The actuator geometry is depicted in Fig. 11. The number of actuators can be reduced by one or two if the positions are not on a regular grid but are optimized.

Simulation results for exact frequency-dependent radiation modes, radiation filters designed for $ka = 0.1$, and radiation filters designed for $ka = 4$ are given in Fig. 12. For the latter two cases, there was no frequency-dependent weighting, i.e., $\tilde{\mathbf{\Lambda}} = \mathbf{I}$. It can be seen that there is not much difference between the curves. For all three methods, the requirement of reductions up to 700 Hz are met. At low and mid frequencies, exact radiation mode control and the radiation filters designed for $ka = 4$ are slightly better. At high frequencies above 700 Hz, however, the radiation filters designed for $ka = 0.1$ are better.

IV. CONCLUSIONS

Efficient sensing schemes have been presented based on the measurement of the velocity or pressure distribution of

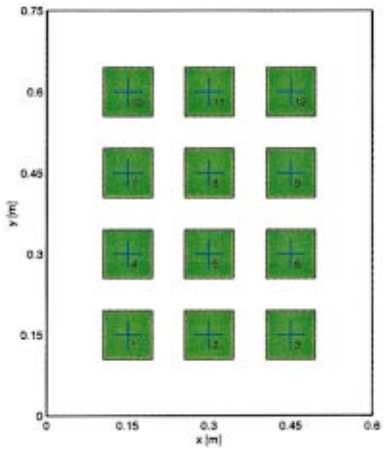


FIG. 11. Actuator configuration; the lines in the center of the actuator indicate relative piezoelectric activity.

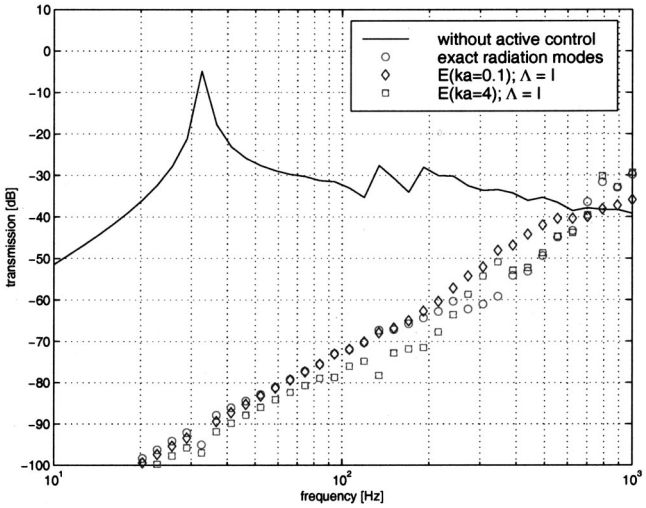


FIG. 12. Transmission of sound power through a steel plate of 60 cm ×75 cm ×3 mm using 7×8 acceleration sensors, ten radiation modes, and the actuator configuration of Fig. 11.

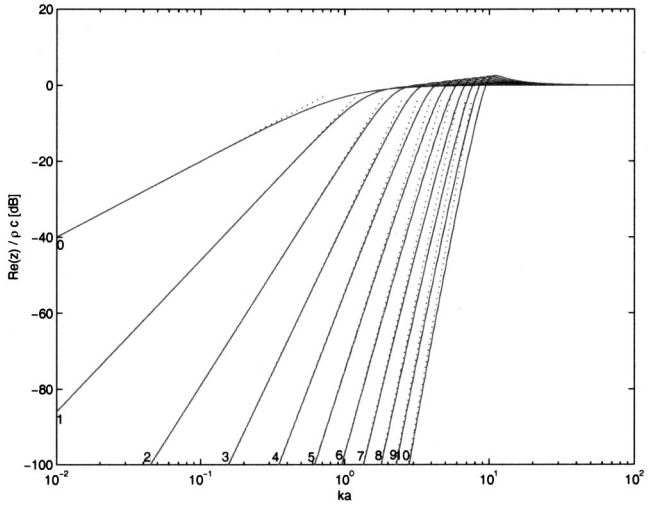


FIG. A1. Radiation resistance of the acoustic eigenfunctions for a sphere of order $l=0,...,10$ (solid line), and low-frequency approximation (dotted line).

radiating structures. The sensor signals can be coupled to simple distribution networks intended for the reduction of the number of channels input to a digital controller. As compared to full frequency-dependent radiation mode control, the schemes result in near-optimal performance at low frequencies. Simple guidelines for the design of such a scheme have been given, leading to requirements for the number of sensors, the number of actuators, and the dimensionality of the controller. It has been shown that the strongest radiation modes at the highest controlled frequency are not the only spatial weighting functions that can be used. Other, possibly simpler spatial weighting functions are also possible. In addition it has been shown that frequency-dependent weighting was not critical for the spatial weighting functions considered. It has been shown that, for a given number of error signals to be minimized by the controller, the frequency-independent weighting approaches lead to controllers with reduced complexity as compared to plane wave decomposition methods which are based on delaying the sensor signals.

ACKNOWLEDGMENTS

The author would like to thank M. W. R. M. xan Overbeek and J. W. Verheij for their suggestions and remarks.

APPENDIX: RADIATION MODE ASYMPTOTICS

The analysis in this Appendix shows that, for a particular separable geometry, the N nonevanescent modes at high frequencies correspond to the N modes with the highest radiation resistance at low frequencies. The results also show the asymptotic behavior at low frequencies of each of the modes as well as the transition point between evanescent behavior and nonevanescent behavior. The derivation essentially follows that of Sarkissian[9] and gives additional results for the radiation resistance at low frequencies. The analysis starts by expressing the pressure $p(r,\theta,\phi)$ at distance r and angle (θ,ϕ) in a fluid outside a sphere in terms of outgoing spherical waves,

$$p(r,\theta,\phi)=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}A_l^m h_l(kr)Y_l^m(\theta,\phi), \tag{A1}$$

with $k=\omega/c$ the wave number, c the speed of sound, ω the angular frequency, h_l the spherical Hankel functions of the first kind, and $Y_l^m(\theta,\phi)$ (Ref. 29) the spherical harmonics. The latter form an orthonormal set of functions on the surface of the sphere:

$$\int Y_l^m(\theta,\phi)Y_{l'}^{m'}(\theta,\phi)\,d\Omega=\delta_{ll'}\delta_{mm'}, \tag{A2}$$

with Ω the solid angle. Let ρ be the density and $\hat{n}$ the unit outward pointing normal. Then the normal velocity $v_n=\hat{n}\cdot\nabla p/(i\rho\omega)$ can be evaluated as

$$v_n=\frac{1}{i\rho c}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}A_l^m h_l'(kr)Y_l^m(\theta,\phi). \tag{A3}$$

Analogous to $z=p/v_n$, the impedance $z_{l,m}=r_{l,m}+ix_{l,m}$ of each surface pressure mode $p_l^m=A_l^m h_l(ka)Y_l^m(\theta,\phi)$ is found to be

$$z_{l,m}=i\rho c\frac{h_l(ka)}{h_l'(ka)}. \tag{A4}$$

Near the surface, the pressure of each mode is proportional to

$$p_l^m\propto\exp\left(\frac{i\rho\omega r}{z_{l,m}}\right). \tag{A5}$$

The nonpropagating evanescent modes are due to the imaginary component $x_{l,m}$ of $z_{l,m}$, and the propagating modes are due to the real component $r_{l,m}$ of $z_{l,m}$. Of interest for the present purpose is the real part $r_{l,m}$ of the impedance $z_{l,m}$. From a series analysis it is found (see also Ref. 30) for the evanescent modes,

$$r_{l,m} \approx \rho c \left[\frac{(ka)^{l+1}}{1 \cdot 3 \cdot 5 \cdots (2l-1)(l+1)} \right]^2, \quad l \gtrsim ka,$$
$$x_{l,m} \approx -\rho c \frac{ka}{l}, \tag{A6}$$

and, for the propagating modes, using the high-frequency approximation of the Hankel function $h_l(\zeta) \rightarrow (-i)^{l+1} \exp(i\zeta)/\zeta$, a real-valued impedance $z_{l,m} = r_{l,m}$:

$$z_{l,m} \approx \rho c, \quad l \lesssim ka. \tag{A7}$$

A plot of $r_{l,m}$ normalized to ρc is given in Fig. A1. The total radiated power Π is

$$\Pi = \frac{1}{2} \mathrm{Re} \left\{ \int_{\partial S} \sum_{l=0}^{\infty} \sum_{m=-l}^{l} (z_{l,m} p_l^m)^* p_l^m \, dS \right\}. \tag{A8}$$

With Eq. (A2), the radiated power can be simplified to

$$\Pi = \frac{a^2}{2} \sum_{l=0}^{\infty} \sum_{m=-l}^{l} |A_l^m|^2 |h_l(ka)|^2 r_{l,m}, \tag{A9}$$

which shows that the modes diagonalize the radiation operator, and, as a result, radiate independently. From Eq. (A6) it can be seen that, for $l \gtrsim ka$, $r_{l,m} > r_{l+1,m}$. As a consequence, the N propagating modes of Eq. (A7), which are sufficient to describe the radiation at the highest frequency, correspond to the N modes with the highest values of $r_{l,m}$ at low frequencies [Eq. (A6)].

[1] C. R. Fuller, S. J. Elliott, and P. A. Nelson, *Active Control of Vibration* (Academic, London, 1996).
[2] J. P. Maillard and C. R. Fuller, ''Advanced time domain wave-number sensing for structural acoustic systems. I. Theory and design,'' J. Acoust. Soc. Am. **95**, 3252–3261 (1994).
[3] J. P. Maillard and C. R. Fuller, ''Advanced time domain wave-number sensing for structural acoustic systems. II. Active radiation control of a simply supported beam.'' J. Acoust. Soc. Am. **95**, 3262–3272 (1994).
[4] J. P. Maillard and C. R. Fuller, ''Advanced time domain wave-number sensing for structural acoustic systems. Part III. Experiments on active broadband radiation control of a simply supported plate,'' J. Acoust. Soc. Am. **98**, 2613–2621 (1995).
[5] J. P. Maillard and C. R. Fuller, ''Comparison of two structural sensing approaches for active structural acoustic control,'' J. Acoust. Soc. Am. **103**, 396–400 (1998).
[6] S. J. Elliott and M. E. Johnson, ''Radiation modes and the active control of sound power.'' J. Acoust. Soc. Am. **94**, 2194–2204 (1993).
[7] G. V. Borgiotti, ''The power radiated by a vibrating body in an acoustic fluid and its determination from boundary measurements,'' J. Acoust. Soc. Am. **88**, 1884–1893 (1990).
[8] D. M. Photiadis, ''The relationship of singular value decomposition to wavevector filtering in sound radiation problems,'' J. Acoust. Soc. Am. **88**, 1152–1159 (1990).
[9] A. Sarkissian, ''Acoustic radiation from finite structures,'' J. Acoust. Soc. Am. **90**, 574–578 (1991).
[10] G. V. Borgiotti and K. E. Jones, ''The determination of the acoustic far field of a radiating body in an acoustic fluid from boundary measurements,'' J. Acoust. Soc. Am. **93**, 2788–2797 (1993).
[11] K. Naghshineh and G. H. Koopmann, ''Active control of sound power using acoustic basis functions as surface velocity filters,'' J. Acoust. Soc. Am. **93**, 2740–2752 (1993).
[12] K. A. Cunefare and M. N. Currey, ''On the exterior acoustic radiation modes of structures,'' J. Acoust. Soc. Am. **96**, 2302–2312 (1994).
[13] M. N. Currey and K. A. Cunefare, ''The radiation modes of baffled finite plates,'' J. Acoust. Soc. Am. **98**, 1570–1580 (1995).
[14] G. Verbeek, A. H. W. M. Kuijpers, and J. W. Verheij, ''Radiation modes and axisymmetric duct acoustics,'' in *Proceedings of Inter-Noise 98*, edited by V. C. Goodwin and D. C. Stevenson (New Zealand Acoustical Society Inc., 1998), pp. 1–4.
[15] B. S. Cazzolato, ''Sensing systems for active control of sound transmission into cavities,'' Ph.D. thesis, University of Adelaide, 1999.
[16] G. V. Borgiotti and K. E. Jones, ''Frequency independence property of radiation spatial filters,'' J. Acoust. Soc. Am. **96**, 3516–3524 (1994).
[17] W. T. Baumann, W. R. Saunders, and H. H. Robertshaw, ''Active suppression of acoustic radiation from impulsively excited structures,'' J. Acoust. Soc. Am. **90**, 3202–3208 (1991).
[18] S. D. Sommerfeldt and D. Lin, ''Estimating radiation modes for broadband measurement,'' J. Acoust. Soc. Am. **105**, 1088–1088 (1999).
[19] A. P. Berkhoff and E. J. J. Doppenberg, ''Active vibration control of light-weight structures: physical aspects related to reduction of sound radiation,'' in *Proc. Adaptronic 99*, edited by R. Wimmel (Adaptronic, Potsdam, 1999), pp. 101–114.
[20] A. P. Berkhoff, P. M. van den Berg, and J. M. Thijssen, ''Ultrasound wave propagation through rough interfaces: iterative methods,'' J. Acoust. Soc. Am. **99**, 1306–1314 (1996).
[21] R. D. Ciskowski and C. A. Brebbia, *Boundary Element Methods in Acoustics* (Elsevier, London, 1991).
[22] P. A. Nelson, ''Some inverse problems in acoustics,'' in *Proc. 6th International Congress on Sound and Vibration* (Technical University of Denmark, 1999), pp. 7–32.
[23] D.-Y. Maa, ''Distribution of eigentones in a rectangular chamber at low frequency range,'' J. Acoust. Soc. Am. **10**, 235–238 (1939).
[24] A. V. Oppenheim and R. W. Schafer, *Discrete-time Signal Processing* (Prentice–Hall, Englewood Cliffs, NJ, 1989).
[25] B.-T. Wang, ''The pvdf-based wave number domain sensing techniques for active sound radiation control from a simply supported beam,'' J. Acoust. Soc. Am. **103**, 1904–1915 (1998).
[26] K. J. Astrom and B. Wittenmark, *Computer Controlled Systems* (Prentice–Hall, London, 1984).
[27] M. E. Johnson and S. J. Elliott, ''Active control of sound radiation using volume velocity cancellation,'' J. Acoust. Soc. Am. **98**, 2174–2186 (1995).
[28] L. Cremer, M. Heckl, and E. E. Ungar, *Structure-borne Sound*, 2nd ed. (Springer-Verlag, Berlin, 1988).
[29] P. M. Morse and K. U. Ingard, *Theoretical Acoustics*, 2nd ed. (Princeton U. P., Princeton, 1986).
[30] E. Skudrzyk, *The Foundations of Acoustics* (Springer Verlag, Wien, 1971).
[31] B.-T. Wang and C. R. Fuller, ''Near-field pressure, intensity, and wavenumber distributions for active structural acoustic control of plate radiation: Theoretical analysis,'' J. Acoust. Soc. Am. **92**, 1489–1498 (1992).
[32] G. P. Gibbs, R. L. Clark, D. E. Cox, and J. S. Vipperman, ''Radiation modal expansion: Application to active structural acoustic control,'' J. Acoust. Soc. Am. **107**, 332–339 (2000).

The vibration behavior of railway track at high frequencies under multiple preloads and wheel interactions

T. X. Wu[a)]
State Key Laboratory of Vibration, Shock and Noise, Shanghai Jiao Tong University, Shanghai 200030, China

D. J. Thompson[b)]
Institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ, United Kingdom

(Received 16 August 1999; accepted for publication 2 June 2000)

The track foundation is preloaded by multiple wheel loads due to the train weight and, as the pad and ballast are nonlinear, their stiffness depends upon the preload in them. Due to the influence of these resilient components of the track, the track vibration is affected by the wheel loads. It is also affected by the wheel/rail interactions. In this article the preloads in the pad and ballast are calculated by considering the nonlinear properties of the track foundation, and thus the preloaded pad and ballast stiffnesses are determined. The vibration properties are explored for the track under multiple wheel loads and multiple wheel/rail interactions by comparing the results from different track models with and without these effects. It is found that the point receptance of the track is reduced and the vibration decay rate is enhanced at low frequencies due to the wheel loads. The effects of the wheel/rail interactions are most significant for frequencies 400–2000 Hz. Because of the wheel/rail interactions, the point receptance fluctuates and the vibration decay is enhanced in the regions around the wheels. © *2000 Acoustical Society of America.* [S0001-4966(00)02409-7]

PACS numbers: 43.40.At [PJR]

I. INTRODUCTION

The major source of railway rolling noise is the structural vibration of the wheels and rails.[1] One aspect of modeling required to predict this noise is therefore the track vibration. Different models have been developed to investigate the dynamic behavior of the railway track over the years.[2]

The effects on the rail vibration of the interaction with multiple wheels have recently been explored by Wu and Thompson.[3] Prior to this the wheel/rail interaction problem at high frequencies has usually been solved for one wheel interacting with the rail (Refs. 4–6). Igeland[7] considered a two-wheel bogie on a single rail and found some evidence of standing wave effects. The main conclusions of Ref. 3 are the following: The extent of the influence of multiple wheels can be seen in the reaction force generated at one wheel/rail contact due to roughness excitation at another. The magnitude of these forces is related to the wave propagation decay rate in the rail and the sum of the point receptances of the wheel (including contact spring) and rail. If the wave propagation decay rate or the point receptance sum is high, the interaction forces will be small. For the track with one or more additional wheels on it, cancellation of the first pinned-pinned resonance occurs in the point receptance when the excitation acts at mid-span.

The effects of the local preload in the track foundation, which is caused by the wheel load due to the train weight, have been studied in Ref. 8. The preload increases the pad and ballast stiffness in the loaded region because of the pronounced nonlinear behavior of the pad and ballast, hence modifying the rail vibration. It is found that the point receptance of a railway track at low frequencies is governed by the local supports near the excitation point and is reduced by the stiffened pad and ballast due to the wheel load. Moreover, the vibration decay at lower frequencies is higher in a preloaded area than in an unloaded area because the preloaded pad and ballast stiffness is higher than the unloaded stiffness.

However, only a single wheel load on the rail was considered in Ref. 8, which is clearly not the case in practice. If multiple wheel loads are taken into account, the static problem of the load distribution becomes very difficult to solve because the track foundation is nonlinear and the superposition principle does not hold. To carry out a comprehensive investigation of the vibration behavior for the track with multiple wheels on the rail, it is therefore desirable to take account of the effects of both the preload in the track foundation and the multiple wheel/rail interactions. The main barrier to doing this is how to calculate the reaction force in the track foundation caused by the multiple wheel loads.

In this article a new methodology is developed to calculate the reaction force in the nonlinear track foundation due to the multiple wheel loads. Although the pad and ballast stiffness is nonlinear, the dynamic behavior of the track may be represented by a linear model with stiffnesses that are chosen according to the quasi-static load, since the dynamic displacements are small. First, the nonlinear property of the track foundation is introduced. Based on this, the reaction force in the foundation caused by the multiple wheel loads is calculated. Then the preloaded pad and ballast stiffness at each support is obtained according to their load-stiffness re-

[a)]Current address: ISVR, University of Southampton, UK.
[b)]Author to whom correspondence should be addressed; electronic mail: djt@isvr.soton.ac.uk

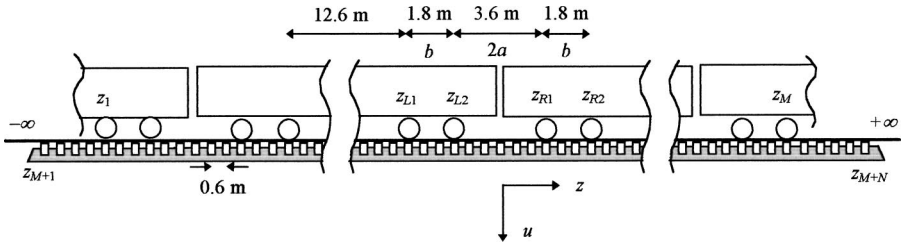


FIG. 1. Schematic diagram of a railway track model with a train (multiple wheels) on the rail.

lationship. Finally, a discretely supported Timoshenko beam model is employed for the dynamic behavior of the track. The receptances and the vibration decay rate for vertical rail vibration are calculated in the frequency range 100–5000 Hz with preloaded pad and ballast stiffness and multiple wheel/rail interactions.

II. STIFFNESS OF TRACK FOUNDATION UNDER MULTIPLE WHEEL PRELOAD

A. Nonlinear load-deflection relationship of track foundation

A railway track is composed of rails, sleepers (crossties), railpads, and ballast. The rails are fixed on the sleepers by clips with the railpad inserted between the rail and sleeper, and the sleepers are laid in the ballast. The load-deflection relations for the railpad and ballast have been found to be nonlinear, and thus their dynamic stiffness varies with the preload in them. The load-deflection relation and dynamic stiffness of the railpad considered here have been measured by Thompson *et al.*[9] The load-deflection relation of the ballast can be assumed approximately to follow the Hertz law (see Ref. 8). As the track foundation is formed by both the pad and the ballast, its load-deflection law may be obtained by combining the pad and ballast deflections under the load (see Ref. 8).

To determine the pad and ballast dynamic stiffness, the preload in them should be known. This can be calculated according to the wheel loads due to the train weight, by treating the rail as an infinite beam on an elastic foundation with nonlinearity. However, this is not easy because the wheel loads are applied to the rail at multiple points and the superposition principle does not hold here due to the nonlinear track foundation. A new method using an alternative linear track foundation to represent the nonlinear one is developed in this section.

B. Preload in track foundation caused by multiple wheel loads

In Ref. 8 the preload in the track foundation caused by a single wheel load is calculated by simplifying the railway track to an infinite uniform beam supported by a continuous nonlinear elastic foundation. This leads to a nonlinear boundary value problem with boundary conditions at an infinite distance from the wheel load. The solution is obtained in Ref. 8 by estimating two parameters, the deflection and its second derivative at the wheel load point $z=0$, to satisfy the boundary conditions at infinity. However, this approach is not appropriate for the case of multiple wheel loads acting on the rail because many parameters at the wheel load points need to be estimated to satisfy the boundary conditions so that the solution searching becomes extremely difficult.

An alternative approach is to use an equivalent linear track foundation to represent the nonlinear foundation. It was found in Ref. 8 that if a constant modulus k of a linear elastic foundation is chosen to give the same maximum deflection at the load point as that from the nonlinear model, the foundation deflection and reaction force of the linear model are very close to those of the nonlinear model. This is also expected to be valid here for the multiple wheel load. The difficulty is to find the appropriate modulus k.

Since the reaction force in the foundation only appears in a local area near the wheel load, the influence of the wheel load on the distant area, for example, more than 3 m away from the load, may be neglected. Thus for a train of bogied wagons, the preload in the track foundation can be calculated by taking only four wheel loads into account which are from a pair of bogies at the adjacent ends of two wagons. The preload of the foundation at other places along the train can be seen to repeat the same variation periodically according to the wagon length.

Figure 1 schematically shows a typical train on a discretely supported railway track. In order to calculate the deflection and reaction force in the track foundation, the foundation is simplified to be continuous and only four wheel loads are considered here. By using a linear model, the deflection $u(z)$ of the rail (or foundation) may be given using the superposition principle:[10]

$$\begin{aligned} u(z)=\frac{P\beta}{2k}[&e^{-\beta|z-a|}(\cos\beta|z-a|+\sin\beta|z-a|)\\ &+e^{-\beta|z-a-b|}(\cos\beta|z-a-b|+\sin\beta|z-a-b|)\\ &+e^{-\beta|z+a|}(\cos\beta|z+a|+\sin\beta|z+a|)\\ &+e^{-\beta|z+a+b|}(\cos\beta|z+a+b|+\sin\beta|z+a+b|)], \end{aligned} \tag{1}$$

where P is the single wheel load, k is the constant modulus of the equivalent elastic foundation, and

$$\beta=\sqrt[4]{\frac{k}{4EI}} \tag{2}$$

with EI the bending stiffness of the rail.

Now the key point is how to choose the modulus k. An optimization method is used here for determining k by minimizing the following objective function y:

$$y=\int_{-L}^{L}(f_l-f_{nl})^2\,dz, \tag{3}$$

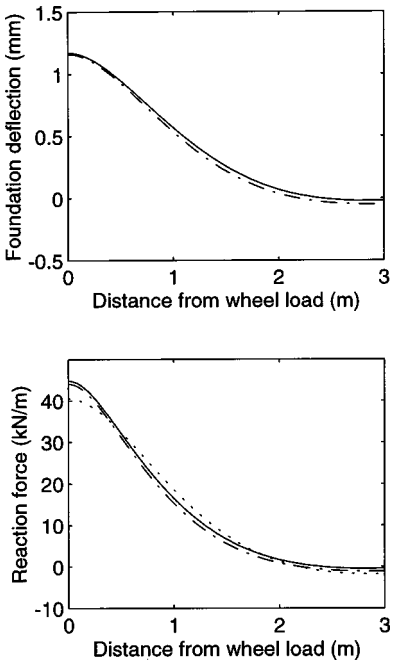


FIG. 2. Foundation deflection and reaction force under a single wheel load of 75 kN. ——— Numerical results from the nonlinear model, —·—·— results from the equivalent linear model and reaction force from the nonlinear load-deflection relationship, ··· reaction force from the linear load-deflection relationship.

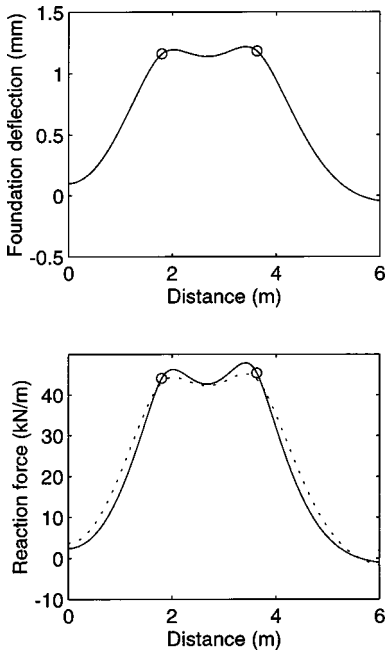


FIG. 3. Foundation deflection and reaction force under four wheel loads of 75 kN each. ——— Results from the equivalent linear model and reaction force from the nonlinear load-deflection relationship, ··· reaction force from the linear load-deflection relationship, ◦ wheel load position.

where f_l is the reaction force in the foundation and is calculated by $f_l = ku$, and u is the result from Eq. (1). Here f_{nl} also is the reaction force but from the nonlinear relationship $f(u)$. Thus f_l represents the result from the linear model, whereas f_{nl} approximately represents the result for the nonlinear model although the deflection u for determining f_{nl} is also from Eq. (1). The objective function y is calculated within a reasonable length of the foundation, for example, $L = 7$ m. By minimizing y the most appropriate modulus k for the equivalent linear foundation can be determined and it is expected that the result from the equivalent linear model will be very close to the result from the nonlinear model in terms of the reaction force. When k is found, so is the deflection of the track foundation. The reaction force in the nonlinear foundation is represented by f_{nl}.

Figure 2 shows the comparison between the results from the alternative linear foundation and the nonlinear foundation for a single wheel load of 75 kN on the rail. The nonlinear model is calculated using the numerical method developed in Ref. 8. The result from the alternative linear model is very close to the result from the nonlinear model in terms of both the deflection and the reaction force. However, if the reaction force is calculated by $f_l = ku$, some deviation appears in the loaded area. This shows the benefit of using the nonlinear load-deflection relationship.

Figure 3 shows the results from the multiple wheel loads. Four wheel loads are applied, each equal to 75 kN. The distances between the wheel loads are $2a = 3.6$ m and $b = 1.8$ m. Because of symmetry the results are shown for only one-half of the foundation. The deflection and reaction force of the foundation can be seen to be very high between the two closer wheels with their peaks near the forcing points. On the other hand, they decrease quite sharply with increasing distance from the load. Beyond about 2 m from the loads, the deflection and reaction forces become approximately zero. Between the two bogies they also become close to zero. Thus the foundation preload caused by the wheel load only appears in a limited local area.

C. Dynamic stiffness of pad and ballast under preload

When the distribution of the foundation preload is known, the preload in each pad and the ballast under each sleeper can be calculated, as in Ref. 8. Two situations are considered here, the wheel load acting either at mid-span or above a sleeper. The span length d is assumed to be uniform and chosen as $d = 0.6$ m. For the wheel distances considered above, if one wheel is at mid-span or above a sleeper, so is each wheel. This is not essential in the model but is chosen for convenience in presenting the results. The preload in each pad includes 20 kN clip load and the rail weight of about 0.36 kN (UIC 60 rail, 60 kg/m). The preload in the ballast for each support includes both the rail weight and the sleeper weight. A monobloc concrete sleeper weighs about 3 kN. Half a sleeper and a 0.6-m rail weigh about 1.86 kN. Far from the wheel load the preload in the pad is formed by the clip load and the rail weight and the preload in the ballast only by the rail and sleeper weight.

The preloaded pad and ballast incremental static stiffness can be determined from their load-stiffness relationship. At a given frequency the dynamic stiffness is in a fixed ratio to this static stiffness. The pad dynamic stiffness varies only slightly with frequency (Ref. 8). At 200 Hz the dynamic stiffness to static stiffness ratio is between about 3.5 and 3.8, corresponding to different preloads and this increases

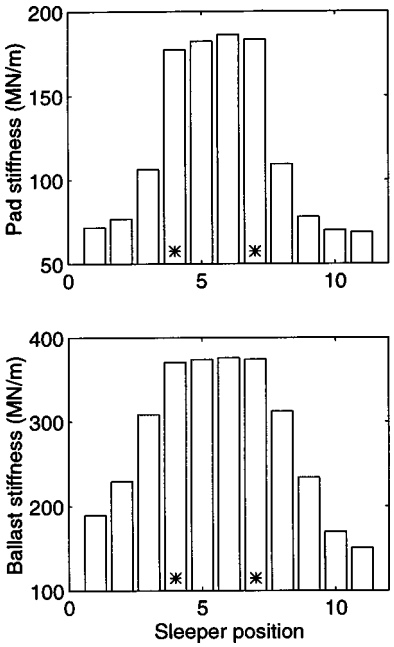


FIG. 4. Preloaded pad and ballast stiffness under four wheel loads above sleepers of 75 kN each. *Location of the wheel load.

slightly with frequency. Here a ratio 3.6 is used for the pad dynamic stiffness for all preloads and frequencies. The ballast stiffness is also frequency dependent, but less data are available; the values used correspond to 200 Hz from Ref. 11. In order to allow for the larger area of ballast in contact with a monobloc sleeper, the ballast stiffness, which was measured with bibloc sleepers (Ref. 11), has been multiplied by 1.3.

Figure 4 shows the pad and ballast dynamic stiffness at each support point for wheel loads above a sleeper. Again only one side of the foundation is shown due to symmetry. The stiffness variation reflects the preload variation in Fig. 3. Between the two closer wheels, the pad and ballast become very stiff due to the large preload in them. Away from the wheel load, they quickly reduce to stable values. The pad and ballast stiffness is affected by the wheel load only at about ten supports for each bogie because the preload in the foundation is local.

III. MODELING OF RAIL VIBRATION

To represent the rail vibration, the track is modeled with an infinite rail and a finite number of discrete supports (see Fig. 1). This is because the supports at large distances from the point at which the receptance is to be calculated can be neglected due to the wave propagation decay. The number of supports should be chosen large enough to guarantee an acceptable approximate solution. The distances between wheels are chosen approximately to represent a bogie container wagon. The model is similar to that used in Refs. 3 and 8, and based on a track model first used by Heckl.[12]

To calculate the response of the rail, both the discrete supports and the wheels can be replaced by point forces. Owing to the symmetrical arrangement of the wheels, the point receptance need be calculated only at two positions, z_{L1} and z_{L2} (or z_{R1} and z_{R2}). To calculate the point receptance at z_{L1} or z_{L2} a unit force should be applied at this position and the corresponding wheel should be omitted. Based on the superposition principle, the responses of the rail at the support points and the wheel/rail contact points to a unit force acting at z_{Li} can be written as

$$u(z_m)=\sum_{\substack{n=1\\ n\neq Li}}^{M} P_{wn}\alpha^{RT}(z_m,z_n)+\sum_{n=M+1}^{M+N} P_{sn}\alpha^{RT}(z_m,z_n) + \alpha^{RT}(z_m,z_{Li}), \quad m=1,2,...,M+N, \tag{4}$$

where $\alpha^{RT}(z_i,z_j)$ is the transfer receptance of an unsupported rail at z_i with the force acting at z_j, M is the number of wheel/rail interactions under consideration, N is the number of the discrete supports, P_{wn} and P_{sn} are the wheel/rail interaction force and the reaction force of the support, respectively, and given as

$$P_{wn}=-Z_w u(z_n), \quad n=1,2,...,M \quad \text{and} \quad n\neq L_i, \tag{5}$$

$$P_{sn}=-Z_{sn}u(z_n), \quad n=M+1,M+2,...,M+N, \tag{6}$$

where Z_{sn} is the dynamic stiffness of the support (see Ref. 8):

$$Z_{sn}=\frac{K_{pn}(K_{bn}-M_s\omega^2)}{K_{pn}+K_{bn}-M_s\omega^2}, \quad n=M+1,M+2,...,M+N, \tag{7}$$

where K_{pn} and K_{bn} are the dynamic stiffness of the pad and ballast, respectively, and are related to the preload in them (see Fig. 4). Here Z_w is the dynamic stiffness combination of the wheel and contact spring and may be calculated using the following formula:

$$Z_w=\frac{1}{\alpha^C+\alpha^W}, \tag{8}$$

where α^C is the receptance of the wheel/rail contact spring, the inverse of the contact stiffness K_H. Since the dynamic displacement is small, the nonlinear contact spring may be treated as a constant linear stiffness. It is calculated as in Ref. 1 and given a value $K_H=1.14\times10^9$ N/m. Here α^W is the wheel receptance at the contact point in the radial direction and can be approximately calculated using a simple model used in Ref. 3. In this model the wheel is composed of a mass M_w and a modal spring K_M,

$$\alpha^W=\frac{1}{K_M}-\frac{1}{M_w\omega^2}. \tag{9}$$

These parameters are chosen as $K_M=4.38\times10^9$ N/m and $M_w=600$ kg, from Ref. 3, and represent a UIC 920-mm standard freight wheel.

Substituting Eqs. (5) and (6) into Eq. (4), Eq. (4) can be solved for $u(z_m)$. Then the displacement at any point on the rail to a unit force at z_{Li} can be obtained by substituting $u(z_n)$ into the following equation:

$$u(z) = -\sum_{\substack{n=1 \\ n \neq Li}}^{M} Z_{wn} u(z_n) \alpha^{RT}(z, z_n) - \sum_{n=M+1}^{M+N} Z_{sn} u(z_n) \alpha^{RT}(z, z_n) + \alpha^{RT}(z, z_{Li}). \quad (10)$$

For $z = z_{Li}$ this gives the point receptance, for other values of z the transfer receptance.

IV. RESULTS

A. Parameters for the calculations

The following parameters are used in the numerical calculations for the railway track vibration model:

$$E = 2.1 \times 10^{11} \text{ N/m}^2, \quad G = 0.77 \times 10^{11} \text{ N/m}^2,$$

$$\rho = 7850 \text{ kg/m}^3, \quad \eta_r = 0.01,$$

$$A = 7.69 \times 10^{-3} \text{ m}^2, \quad I = 30.55 \times 10^{-6} \text{ m}^4, \quad \kappa = 0.4,$$

$$M_s = 150 \text{ kg}, \quad d = 0.6 \text{ m}, \quad \eta_p = 0.25, \quad \eta_b = 0.6,$$

where E is the Young's modulus, G is the shear modulus, and ρ is the density. The rail damping is introduced by making E and G complex with the factor $(1 + i\eta_r)$, where η_r is the rail loss factor. The geometric properties of the cross section are characterized by A, the cross-sectional area, I, the area moment of inertia, and κ, the shear coefficient. M_s and d are the sleeper mass and the span length, respectively. In the loaded area the pad and ballast stiffness K_{pn} and K_{bn} are chosen according to the preload caused by the wheel load (see Fig. 4). In the unloaded area they are chosen as constant. The damping of the pad and ballast is also introduced by making K_{pn} and K_{bn} complex with the pad and ballast loss factors η_p and η_b, respectively. The ballast loss factor is chosen lower than usual in order to make effects in the response more visible.

The point receptance of the track for vertical vibration is calculated for the excitation acting at mid-span in the frequency range 100–5000 Hz. The wave propagation decay rate along the rail is examined in the same frequency range. The number of the discrete supports is chosen as $N = 100$. Three groups of wheel loads and/or wheel/rail interactions are applied in the considered length of the supported rail. In each group there are four wheels with each single wheel load equal to 75 kN.

B. Point receptance

The point receptances of the track are calculated for the excitation acting at z_{L1} and z_{L2}, which are chosen to be at mid-span. Three cases are calculated for comparison, i.e., considering both the wheel/rail interactions and the preloaded foundation, considering the wheel/rail interactions only, or considering the preloaded foundation only. Only the results for the excitation at z_{L2} are shown. The results for the excitation at z_{L1} are similar.

Figure 5 shows the results. It can be seen that the point receptance of the preloaded track at low frequencies is lower than that of the unloaded track, and the two low-frequency resonances, at which the whole rail bounces on the ballast stiffness and on the pad stiffness, appear at higher frequencies for the preloaded track (approximately 200 and 400 Hz). At high frequencies (above 1000–2000 Hz), however, the preload in the foundation caused by the wheel load has no apparent effects on the track vibration in terms of the point receptance.

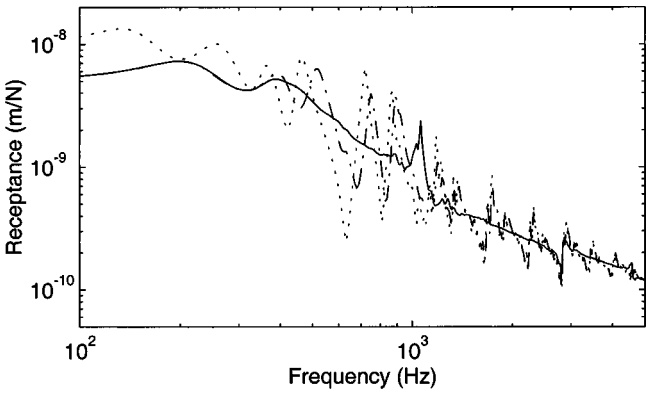


FIG. 5. Amplitude of the point receptances with excitation at z_{L2} and at mid-span. ——— For preloaded foundation only, —·—·— for both preloaded foundation and wheel/rail interactions, ··· for wheel/rail interactions only.

The effects of the wheel/rail interaction can be seen to be significant mainly at higher frequencies, above about 400 Hz. At low frequencies the point receptance of the track with multiple wheels on the rail is the same as without wheels on the rail. This is because the wave propagation decay rate is high at low frequencies so that the generated wheel/rail interaction force at the wheel positions is weak and the wave reflected by the wheel is also weak and, moreover, it is quickly attenuated before arriving back at the excitation point due to the high decay rate. At higher frequencies the point receptance of the rail with wheel/rail interactions fluctuates around that of the rail without interactions because the wheel/rail interaction forces are greater due to the low decay rate of the wave propagation. For the track with multiple wheels on the rail the first pinned-pinned resonance peak at just above 1000 Hz can be seen to disappear for the excitation acting at mid-span. A detailed analysis concerning the cancellation of the pinned-pinned resonance can be found in Ref. 3 for a single additional wheel. The fluctuation of the point receptance due to the wheel/rail interaction can be seen to be greatest around 500 to 900 Hz. The reason for this is that in this frequency range the point receptance sum of the wheel, contact spring, and rail tends to reach its minimum (see Ref. 3), whereas the wave propagation decay rate is fairly low, so that the wheel/rail interaction force P_{wn} reaches its maximum. With increasing frequency above 1000 Hz, the fluctuation of the point receptance can be seen to decrease gradually. Since the wave propagation decay rate is lower for the unloaded track than the preloaded track, the effects of the wheel/rail interaction in terms of the fluctuation of the point receptance are more apparent for the unloaded track.

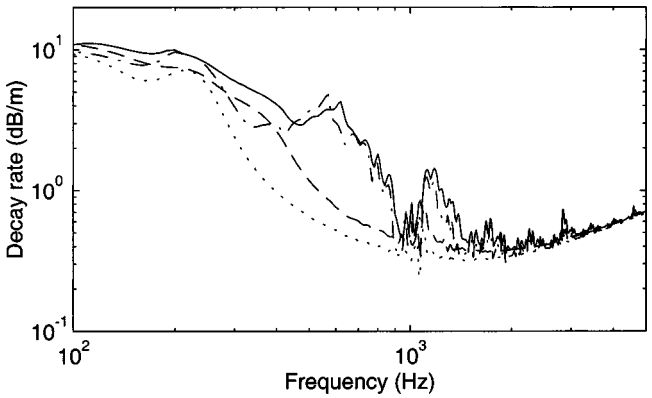


FIG. 6. Vibration decay rates along the track for different track models with excitation at z_{L2} and above a sleeper. ——— For both preloaded foundation and wheel/rail interactions, – – for preloaded foundation only, —·—·— for wheel/rail interactions only, ··· for unloaded foundation and without wheel/rail interactions.

C. Transfer receptance and decay rate

The transfer receptances of the track for vertical vibration are calculated at some frequencies to show some specific features of the vibration transmission along the track. The vibration decay rate in the track is calculated in the frequency range 100–5000 Hz from the decay in vibration level over a wagon length from the excitation point divided by the distance (19.8 m). From the decay rate a general view of the vibration decay can be obtained in the considered frequency range, whereas from the transfer receptances along the track detailed information about the vibration transmission can be gained.

The vibration decay rates along the track are calculated for different track models. These models include the unloaded track with and without wheel/rail interactions and the preloaded track with and without such interactions. Figure 6 shows the decay rate to the left side of the excitation which acts at z_{L2} and above a sleeper. The decay rate to the right side is almost the same. It can be seen that the vibration decay rate is high at low frequencies and low at high frequencies. At low frequencies up to about 200 Hz and high frequencies from about 2000 Hz the decay rates for the different track models are very similar. The main differences in decay rate for the different models are between about 300 and 900 Hz and near the first pinned-pinned resonance area above 1000 Hz. For the tracks without wheel/rail interactions the decay rate in the preloaded track is generally higher than in the unloaded track because the stiffness of the preloaded pad and ballast is higher than in the unloaded case. The maximum difference in the decay rate between the preloaded and unloaded tracks may reach about 3 dB/m at about 400 Hz. For the tracks with wheel/rail interactions the vibration decay rates are much higher than for those without such interactions in the frequency range 350–850 Hz and around the first pinned-pinned resonance. The largest difference in the decay rate can be seen to be about 6 dB/m for the preloaded track or 8 dB/m for the unloaded track due to the wheel/rail interactions.

The transfer receptances of the track at frequencies 200, 400, and 1050 Hz are shown in Figs. 7–9. They are calculated at each support point with the excitation at z_{R1} and above a sleeper. For comparison four cases are calculated for the tracks with and without the preload and with and without wheels on the rail. Compared with the point receptance, variations in the transfer receptance for the track with multiple wheels on it are more complicated because both frequency and spatial position affect the transfer receptance.

At 200 Hz (Fig. 7) the transfer receptances at the wheel/rail contact points for the track with wheel/rail interactions can be seen to be locally minimum. The reason for this is that at about 200 Hz the receptance of the wheel is approximately equal to the receptance of the wheel/rail contact spring in amplitude but out of phase with it. Thus the sum of the wheel and contact spring receptances is extremely small so that the rail seems to be pinned by the wheel at the wheel/

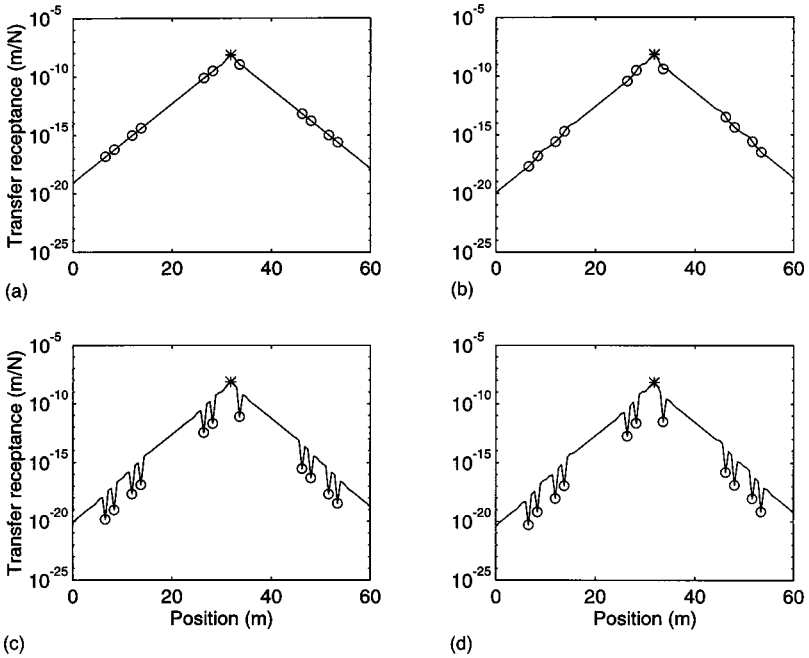


FIG. 7. Amplitude of transfer receptances at 200 Hz for different models with excitation at z_{R1} and above a sleeper. (a) Unloaded foundation and without wheel/rail interactions, (b) preloaded foundation only, (c) wheel/rail interactions only, and (d) both preloaded foundation and wheel/rail interactions. * Excitation position, ○ wheel position.

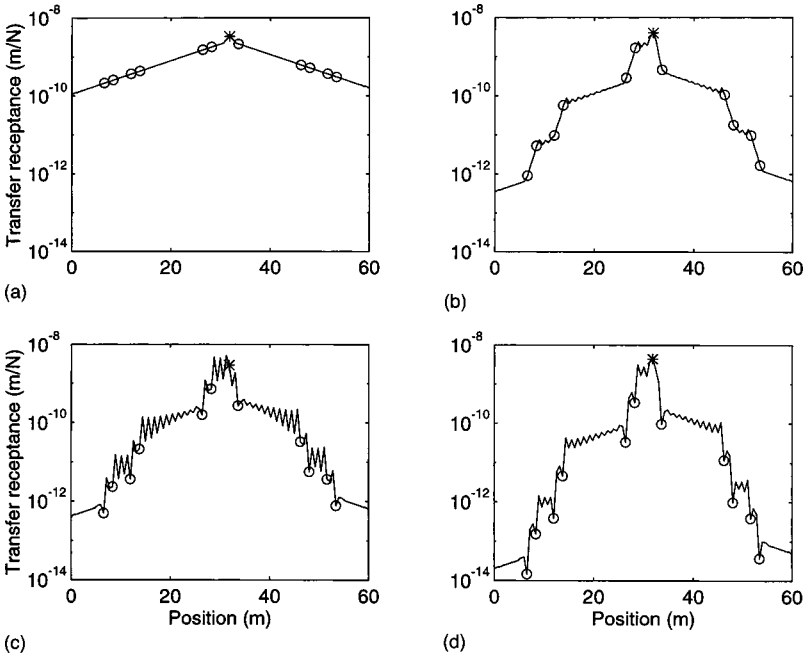


FIG. 8. Amplitude of transfer receptances at 400 Hz for different models, key as for Fig. 7.

rail contact point and the transfer receptance there becomes very small. Comparing the transfer receptance of the preloaded track at 400 Hz (Fig. 8) to the unloaded track, the vibration decay is enhanced in the loaded area of the track. What is more, the vibration decay at 400 Hz is also enhanced in the wheel regions by wheel/rail interactions. Furthermore, reflections can be seen to generate standing wave patterns. As a result, the average vibration decay rate over a wagon length for the track with preload and/or wheel/rail interactions is higher than without them, as shown in Fig. 6. The first pinned-pinned resonance occurs at about 1050 Hz. From the transfer receptances at this frequency (Fig. 9) the vibration decay can be seen to increase for the preloaded track or due to the wheel/rail interactions. However, the decay increase is found to happen only in the unloaded area or the area without wheel/rail interaction in terms of the transfer receptances at support positions (solid lines) but in the loaded area or the wheel/rail interaction area in terms of the transfer receptances at mid-spans (dotted lines). In addition, the transfer receptances at the wheel/rail contact points for the track with wheel/rail interactions are locally maximum compared to those at other support positions.

V. CONCLUSIONS

The influences on the track vibration of the preloaded foundation stiffness combined with the wheel/rail interactions have been investigated. The static reaction force in the track foundation has been calculated using an equivalent linear model representing the nonlinear model. The pad and

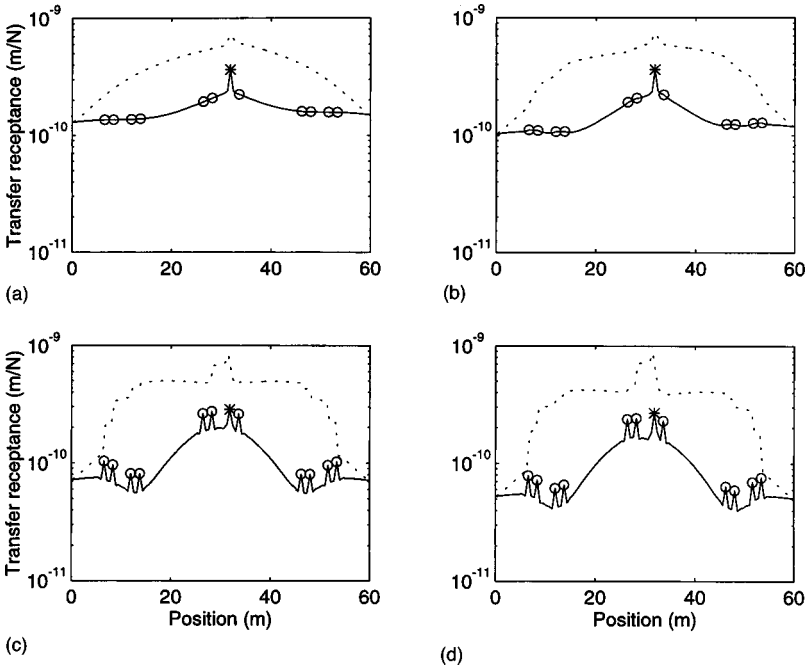


FIG. 9. Amplitude of transfer receptances at 1050 Hz for different models, key as for Fig. 7. ··· Transfer receptances at mid-span.

ballast stiffness has been determined according to their load-stiffness relationship and the preload in them. A Timoshenko beam model has been employed to represent the infinite rail with the variable stiffness supports and multiple wheel/rail interactions. The point and transfer receptances and the vibration decay rates have been calculated in the frequency range 100–5000 Hz for different models. The results from these models have been analyzed by comparison.

The preload in the track foundation is local and near the wheel loads. At distances more than about 2 m away from the wheels the preload in the foundation can be considered negligible. Thus only a limited number of pads and ballast springs are preloaded by the wheel loads and become stiffer. Due to the stiffened pad and ballast the point receptance at low frequencies up to about 300 Hz is reduced and the vibration decay rate in the frequency range up to the first pinned-pinned resonance is enhanced. Especially, the wave propagation decay in the loaded area is significantly higher than in the unloaded area in the frequency region 300–800 Hz. The first two resonances related to the foundation stiffness appear at higher frequencies for the preloaded track due to the stiffened pad and ballast.

The wheel/rail interaction is weak at low frequencies and thus the point receptance and the vibration decay rate for the track with wheel/rail interactions are similar to those without them. The effects of the wheel/rail interaction are mainly at frequencies from 500 to 2000 Hz for the point receptance and from 400 to 1400 Hz for the decay rate. Because of the presence of the wave reflections the point receptance of the track with wheels on it fluctuates from about 400 Hz. The fluctuation in the point receptance decreases gradually with increasing frequency above 1000 Hz. The first pinned-pinned resonance peak is cancelled due to the wheel/rail interactions when an excitation acts at mid-span. The vibration decay is enhanced in the wheel/rail interaction areas at frequencies from 400 to 800 Hz and around the first pinned-pinned resonance.

Finally, the main effects on track dynamics of the multiple wheels on the rail may be summarized: Wheel preloads reduce the rail receptance and increase the vibration decay in the loaded region at low frequencies, and multiple wheel/rail interactions cause fluctuation in the rail receptance, increasing vibration decay in the region around the wheels and cancellation of the first pinned-pinned resonance.

[1] D. J. Thompson, ''Wheel/rail noise: theoretical modelling of the generation of vibrations,'' Ph.D. Thesis, University of Southampton, 1990.

[2] K. L. Knothe and S. L. Grassie, ''Modelling of railway track and vehicle/track interaction at high frequencies,'' Vehicle System Dynamics **22**, 209–262 (1993).

[3] T. X. Wu and D. J. Thompson, ''Effects of multiple wheels on a rail on the rail vibration,'' in Proceedings of Sixth International Congress on Sound and Vibration, Lyngby, Denmark (1999), pp. 2629–2636.

[4] S. L. Grassie, R. W. Gregory, D. Harrison, and K. L. Johnson, ''The dynamic response of railway track to high frequency vertical excitation,'' J. Mech. Eng. Sci. **24**, 77–90 (1982).

[5] P. J. Remington, ''Wheel/rail rolling noise I: theoretical analysis,'' J. Acoust. Soc. Am. **81**, 1805–1823 (1987).

[6] D. J. Thompson, ''Wheel-rail noise generation, part I: introduction and interaction model,'' J. Sound Vib. **161**, 387–400 (1993).

[7] A. Igeland, ''Railhead corrugation growth explained by dynamic interaction between track and bogie wheelsets,'' Proc. Inst. Mech. Eng. Part F **210**, 11–20 (1996).

[8] T. X. Wu and D. J. Thompson, ''The effects of local preload on the foundation stiffness and vertical vibration of railway track,'' J. Sound Vib. **219**, 881–904 (1999).

[9] D. J. Thompson, W. J. van Vliet, and J. W. Verheij, ''Developments of the indirect method for measuring the high frequency dynamic stiffness of resilient elements,'' J. Sound Vib. **213**, 169–188 (1998).

[10] S. Timoshenko, *Strength of Materials, Part II Advanced Theory and Problems*, 3rd ed. (Van Nostrand Reinhold, New York, 1958).

[11] N. Frémion, J. P. Goudard, and N. Vincent, ''Improvement of ballast and sleeper description in TWINS—step 1: experimental characterisation of ballast properties,'' VIBRATEC report ref. 072.028a (1996).

[12] M. A. Heckl, ''Railway noise—Can random sleeper spacing help?'' Acustica **81**, 559–564 (1995).

European methodology for testing the airborne sound insulation characteristics of noise barriers *in situ*: Experimental verification and comparison with laboratory data

Massimo Garai[a] and Paolo Guidorzi
Department of Energetic, Nuclear and Environmental Control Engineering (DIENCA), Università di Bologna, Viale Risorgimento, 2, 40136 Bologna, Italy

(Received 3 December 1999; accepted for publication 9 May 2000)

In the frame of the 1994–1997 *Standard, Measurement and Testing* program, the European Commission funded a research project, named *Adrienne*, to define new test methods for measuring the intrinsic characteristics of road traffic noise reducing devices *in situ*. The research team produced innovative methods for testing the sound reflection/absorption and the airborne sound insulation characteristics of noise barriers. These methods are now under consideration at CEN (European Committee for Standardization), to become European standards. The present work reports a detailed verification of the test method for airborne sound insulation over a selection of 17 noise barriers, representative of the Italian and European production. The samples were tested both outdoors, using the new *Adrienne* method, and in laboratory, following the European standard EN 1793-2. In both cases the single number rating for airborne sound insulation recommended by the European standard was calculated. The new method proved to be easy to use and reliable for all kinds of barriers. It has been found sensitive to quality of mounting, presence of seals, and other details typical of outdoor installations. The comparison between field and laboratory results shows a good correlation, while existing differences can be explained with the different sound fields and mounting conditions between the outdoor and laboratory tests. It is concluded that the *Adrienne* method is adequate for its intended use.
[S0001-4966(00)03208-2]

PACS numbers: 43.50.Gf, 43.50.Sr, 43.50.Lj [MRS]

INTRODUCTION

The role of noise barriers in reducing the impact of road and train noise is widely acknowledged by the international community.[1] The effectiveness of noise barriers depends on several factors such as

(i) barrier geometry: height, length, thickness, shape;
(ii) barrier acoustical characteristics: sound reflection and diffusion of the surface exposed to noise, airborne sound insulation;
(iii) installation-related factors: quality of workmanship, way of mounting (panel–post and panel–panel assembling mechanism), presence of seals;
(iv) deterioration with time;
(v) site geometry: ground profile, screen position relative to noise sources and receivers;
(vi) site acoustical characteristics: ground impendance; and
(vii) meteorological conditions: wind, temperature gradients, etc.

Among these, airborne sound insulation is important to attenuate the noise transmitted directly through the screen so that it is not significant compared with the sound diffracted over the top edge. Airborne sound insulation of installed screens depends on two classes of factors:

(i) intrinsic characteristics: airborne sound insulation, quality of workmanship, way of mounting, presence of seals; and
(ii) extrinsic characteristics: all the other site-related characteristics.

Only the intrinsic characteristics are relevant to qualify the product itself. Therefore, a method should be available to test on the installation site the compliance of the intrinsic characteristics of the installed product to the design specifications. Moreover, installed noise barriers may change their characteristics with time; therefore, existing noise barriers should be regularly tested *in situ*, i.e., along roads and railways where they are installed, to check the possible decay of their acoustical properties.

For all these reasons, there is a strong need of a method to test the intrinsic characteristics of noise barriers *in situ*. On the other hand, no such method exists at international level. In most European countries only laboratory tests are performed, accordingly with the new EN 1793-1, 1793-2, and 1793-3[2–4] (European standards are identified by the two letters EN before their number). The European standards rely on the ISO 354[5] for sound absorption and the ISO 140-3[6] for airborne sound insulation, with the additional requirement that a post must be included in the laboratory specimen (when applicable for the specific kind of barrier). The abovementioned ISO standards have been designed for building products and therefore require the tests to be performed inside special rooms under a diffuse sound field, which is very

[a]Electronic mail: massimo.garai@mail.ing.unibo.it

different from the sound field in front of an installed noise barrier. As a result, the airborne sound insulation of noise barriers in their intended use is underestimated.[7] Furthermore, the most common defects of actual installations along roads and railways, like sound leakage due to poor workmanship, bad connections between panels and posts, absence of seals, etc., cannot be detected using laboratory methods.

At national level, there is only one standard regarding *in situ* measurements, used in France.[8] However, this method relies on the use of a gunshot as sound source, which has a poor repeatability, and a short time window for the analysis (3 ms in length), which limits the lowest reliable frequency to the one-third octave band of 400 Hz. Moreover, the placement of the analysis window is left to the operator's choice. For these reasons, the French method has not been accepted by the European Committee for Standardization (CEN).[7,9,10]

Recognizing that this situation is an obstacle to the free circulation of noise barriers on the European market, the European Commission, in the frame of the 1994–1997 *Standard, Measurement and Testing* (SMT) program, funded a research project, named *Adrienne*, to define new test methods for measuring the intrinsic characteristics of road traffic noise reducing devices *in situ*. After a three-year work, the research team produced innovative methods for testing the sound reflection/absorption and the airborne sound insulation characteristics of noise barriers. These methods are now under consideration at CEN, in CEN/TC 226/WG6 (Technical Committee 226, Working Group 6), to become European standards. The present work reports a detailed verification of the *Adrienne* airborne sound insulation test method on a selection of 17 noise barriers, representative of the Italian and European market. The samples were tested both outdoors, using the new *Adrienne* method in controlled conditions, and in laboratory, following the EN 1793-2 standard. In both cases the single number rating for airborne sound insulation recommended by the European standard was calculated.[3] The main aims of the work were at least three:

(i) to test the practicability and the reliability of the new *Adrienne* method for different kinds of barriers (composed mainly of concrete, metal, acrylic or wood; having a flat or nonflat surface; having an absorbing or reflecting surface; etc.);
(ii) to test the sensitivity of the new method to quality of workmanship, way of mounting, and other details typical of real outdoor installations; these sources of possible problems are present in real situations but are carefully eliminated when preparing laboratory specimens; and
(iii) to compare the outdoor and laboratory airborne sound insulation values obtained on the same set of barrier samples and to investigate their possible correlation, which would be useful for predicting the expected field performance from laboratory data.

I. THE SAMPLES

All samples had the same global size: about $3.0 \times 3.5\,m^2$ for the laboratory test (the size of the test opening between the coupled rooms) and $18.0 \times 4.0\,m^2$ for the outdoor test. The basic characteristics of the 17 barrier samples are summarized in Table I. The barrier samples are presented with conventional names in order to not disclose the producer names. The barriers submitted to the test can be grouped in six classes:

(1) Concrete barriers (five samples): a typical barrier element is made of a heavy concrete back panel supporting a front panel made with lighter concrete and with a nonflat shape, in order to increase the sound absorption of the surface. The posts are large and strong to support the considerable weight of the structure.
(2) Metallic barriers (seven samples): a typical barrier element is a box made with a metallic sheet having a thickness ranging from 1 to 2 mm. The box is perforated on one face and partially filled with a high-density rock wool. In two cases a high-density synthetic damper was added (MET2 and MET7). In two cases the metallic sheet was not folded to form a box and not perforated (MET4 and MET5). The posts are metallic beams with a ''*H*'' section and 160 mm wide.
(3) Resin barriers (one sample): the barrier elements are boxes made with a 3-mm-thick polyether resin sheet reinforced using glass fibers. The boxes are perforated on one face and partially filled with a glass fiber blanket. The posts, 100 mm wide, are made of the same polyether resin, using a layer 10 mm thick.
(4) Acrylic barriers (one sample): the barrier elements are transparent polymethylmethacrylate (PMMA) sheets, 20 mm thick, supported by a light metallic frame.
(5) Mixed barriers (one sample): the half-barrier close to the ground is made of metallic panels, like those described above (point 2); the upper half is made of transparent polymethylmethacrylate (PMMA) sheets, 15 mm thick, supported by a light metallic frame. The posts are metallic beams with a ''*H*'' section and 160 mm wide.
(6) Wood barriers (one sample): the barrier is made of four layers i.e., from front to back, wood tiles made of spaced laths; rock wool blanket; fiber–concrete aggregate board; wood board. The posts are metallic beams with a ''*H*'' section and 160 mm wide.

II. LABORATORY MEASUREMENTS

The laboratory test method specified in EN 1793-2[3] fully conforms to the well-known ISO 140-3,[6] with some additions particularly relevant for noise barriers. The most important points of the procedure are summarized in the following.

The measurements are performed in laboratory test facilities consisting of two adjacent reverberant rooms, called the source room and the receiving room, respectively, with an opening between them in which the test specimen is inserted. Transmission of sound on flanking paths must be suppressed. The measurements reported in the present work were performed in a laboratory where the test aperture is about $3.0 \times 3.5\,m^2$ (width×height). When necessary, a filler wall with a significantly better sound insulation than the test specimen was used to fit the specimen in the aperture.

Following EN 1793-2, the test specimen must be mounted in the test opening and assembled in the same man-

TABLE I. Basic characteristics of the tested barriers.

Sample	Type	Basic elements composition	Element thickness (mm)	Element height (m)	Element width (m)
CON1	Concrete	Back: concrete 2500 kg/m^3 Front: light clay aggregate 1250 kg/m^3	275	2.00	4.50
CON2	Concrete	Back: concrete 2500 kg/m^3 Front: light clay aggregate 1250 kg/m^3	250	2.00	3.00
CON3	Concrete	Back: concrete 2500 kg/m^3 Front: light clay aggregate 1250 kg/m^3	250	2.00	3.00
CON4	Concrete	Back: concrete 2050 kg/m^3 Front: light clay aggregate blocks (1000 kg/m^3) with holes (resonators)	220	2.00	3.26 (last: 1.70)
CON5	Concrete	Back: concrete 2500 kg/m^3 Front: light clay aggregate 1200 kg/m^3	190	4.00	2.40 (last: 1.20)
CON6	Concrete	Back panel: concrete 2500 kg/m^3 Front panel: grooved light concrete	240	1.00	4.00 (last: 2.00)
MET1	Metal	Box made with a 1.5-mm metallic sheet, perforated on one face and filled with a 100-kg/m^3 rock wool blanket, 80 mm thick	119	0.50	4.00 (last: 2.00)
MET2	Metal	Curved box made with a 1.5-mm metallic sheet, perforated on one face and filled with an 85-kg/m^3 rock wool blanket, 50 mm thick, and a high-density synthetic damper	130	0.50	4.00 and 3.00
MET3	Metal	Box made with a 1.5-mm metallic sheet, perforated on one face and filled with a 95-kg/m^3 rock wool blanket, 60 mm thick	115	0.50	3.00
MET4	Metal	Single 1.0-mm metallic sheet, folded to form protruding supports 60 mm wide	60	0.50	4.00 (last: 2.00)
MET5	Metal	Single 1.0-mm metallic sheet, folded to form protruding supports, 60 mm wide, containing panels in polyesther fibre wool, thickness 50 mm, density 50 kg/m^3. Frontal protection grid.	60	0.50	4.00 (last: 2.00)
MET6	Metal	Box made with a 1.0-mm metallic sheet, perforated on one face and filled with a 100-kg/m^3 rock wool blanket, 40 mm thick	122	0.50	3.00
MET7	Metal	Box made with a 1.8-mm metallic sheet, perforated on one face and filled with a 85-kg/m^3 rock wool blanket, 40 mm thick, and a high-density synthetic damper	80	0.50	4.00 and 3.00
RES1	Resin	Box made with a 3-mm-thick polyether resin sheet reinforced using glass fibers, perforated on one face and filled with a 60-kg/m^3 glass fiber blanket, 40 mm thick	86	0.50	3.00
ACR1	Acrylic	Polymethylmethacrylate (PMMA) sheets 20 mm thick, supported by a metallic frame, 130 mm thick	20	2.00	3.00
MIX1	Bottom (0 to 2 m): metal; top (2 to 4 m): acrylic	Box made with a 1.5-mm metallic sheet, perforated on one face and filled with a 90-kg/m^3 rock wool blanket, 60 mm thick. Polymethyl methacrylate (PMMA) sheets 15 mm thick, supported by a metallic frame, 120 mm thick	120/15	0.50/2.00	3.00
WOOD	Wood	From front to back: wood tiles 1×1 m^2, made of spaced laths 19 mm thick; 90-kg/m^3 rock wool blanket, 60 mm thick; fiber-concrete aggregate board, 4 mm thick; wood board, 19 mm thick	127	1.00	3.00

ner as the manufactured device used in practice, with the same connections and seals between component parts. The edge supports must not overlap the sample by more than 70 mm and must be sealed to prevent leakage of sound. Where posts are employed in construction, at least one post must be included in the specimen, with panels attached on both sides. The length of the panels on one side of the post must be greater than or equal to 2 m. The side that would face the traffic noise source must face the source room.

In the source room a loudspeaker produces a continuous

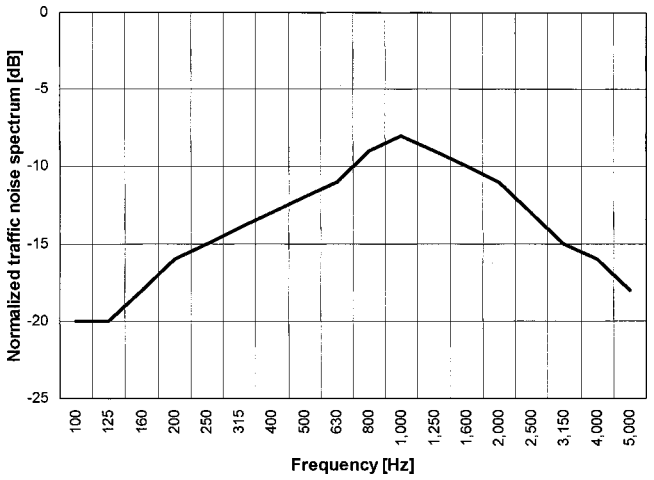


FIG. 1. Normalized A-weighted traffic noise spectrum from EN 1793-3.

broadband noise (pink noise was used in the present work) in the one-third octave bands from 100 Hz to 5 kHz. Moving microphones are used to obtain the average sound pressure levels in each room. Measurements are repeated moving the loudspeaker in different positions and averaged. The values of the airborne sound reduction index R_j, in each frequency band, is then calculated from the equation

$$R_j = L_{p,1,j} - L_{\rho,2,j} + 10 \cdot 1g \frac{S}{A_j} (\mathrm{dB}), \tag{1}$$

where j is the index of the one-third octave bands from 100 Hz to 5 kHz; $L_{p,1,j}$ is the average sound pressure level in the source room (dB); $L_{p,2,j}$ is the average sound pressure level in the receiving room (dB); S is the area of the test specimen (m^2); and A_j is the equivalent absorption area in the receiving room (m^2).

The sound generation chain was assembled with a WR-Electronics loudspeaker, a Masters PWA-202/4 power amplifier, a Brüel & Kjær type 1405 noise generator, and an Applied Research & Technology, Inc. HD-31 equalizer. The sound field was sampled using two microphones, Brüel & Kjær type 4192, with pre-amplifiers, Brüel & Kjær type 2669, connected to a Brüel & Kjær analyzer (a type 2144 and a type 2123 were used).

According to EN 1793-2, a single number rating of sound insulation, DL_R, must be calculated to give an overall indication of the performance of the product. It is defined as

$$DL_R = -10 \lg \left[\frac{\Sigma_{j=1}^{18} 10^{0.1L_j} 10^{-0.1R_j}}{\Sigma_{j=1}^{18} 10^{0.1L_j}} \right] (\mathrm{dB}), \tag{2}$$

where L_j is the normalized A-weighted sound pressure level of traffic noise in the jth one-third octave band (dB), as defined in EN 1793-3,[4] (see Fig. 1). DL_R must be given rounded to the nearest integer. If there is a need to categorize airborne sound insulation, this is done on the basis of the DL_R value, using the categories listed in Table II. It is worth noting that the normalized traffic noise spectrum comes from the average of road traffic noise spectra taken in Europe; consequently, greatest weight is given to performance at frequencies that are important in European road traffic noise. It is also worth noting that DL_R is different, for the calculation procedure and the reference spectrum, from the well-known rating R_w used in building acoustics.[11]

TABLE II. Categories of airborne sound insulation (EN 1793-2).

Category	DL_R (dB)
B0	Not determined
B1	<15
B2	15 to 24
B3	>24

Examples of measured values of the sound reduction index, R_j, in the one-third octave frequency bands from 100 Hz to 5 kHz, are reported in graphical form in Figs. 10–17 together with the corresponding outdoor measurement results (see Sec. IV). All values of the ratings R_w and DL_R are reported in Table III. The DL_R values were calculated both on the full frequency range 100 Hz to 5 kHz, in one-third octave frequency bands, and in the ''restricted'' frequency range 200 Hz to 5 kHz; the latter calculation was made in view of the comparison with the single number rating values resulting from the outdoor measurements, which have a low-frequency limit of about 160 Hz and then a valid frequency range going from 200 Hz to 5 kHz, in one-third octave frequency bands (see Sec. IV C).

III. SUMMARY OF THE *ADRIENNE* TEST METHOD FOR AIRBORNE SOUND INSULATION

As the *Adrienne* test method is rather new, it will be summarized here in more detail than the laboratory method. However, this summary is not intended to explain the reason of each choice made during the development of the *Adrienne* research; for this, reference should be made to Refs. 7 and 10.

A. Review of previous literature

The *Adrienne* method is based on the recovering of an acoustic impulse response behind the barrier under test; to this extent it has an ideal connection with transient methods proposed for testing the transmission loss of partitions in buildings. The first attempt was made by Raes,[12] who used sinusoidal waves with exponential modulated amplitude as test signal and compared the maximum amplitude of the incident and transmitted sound signals in the time domain. In a subsequent paper, Raes[13] distinguished between ''static'' and ''dynamic'' transmission loss, the first being measured using a stationary sound field, the latter using an impulsive excitation. He pointed out that measuring the transmission loss using transient sound signals is closer to real life situations, where most noise sources are not stationary and do not produce a diffuse sound field. De Tricaud,[14] using a pistol shot as sound source, obtained a fairly good agreement between impulsive and classical measurements of transmission loss in buildings. He remarked that the pistol shot should have the same power spectrum of the loudspeaker used in the conventional method (at least in octave bands) for the results of the two tests be comparable. He also claimed that impulsive methods are more robust against extraneous noise. Roland[15] improved the de Tricaud technique, pointing out

TABLE III. Single number ratings of airborne sound insulation.

Sample	Type	R_w (dB) laboratory 100 Hz to 5 kHz	DL_R (dB) laboratory 100 Hz to 5 kHz	DL_R (dB) laboratory 200 Hz to 5 kHz	DL_{SI} (dB) outdoors elements 200 Hz to 5 kHz	DL_{SI} (dB) outdoors posts 200 Hz to 5 kHz
CON1	Concrete	56	52	54	63	61
CON2(Q)	Concrete	46	44	44	57	38
CON2(A)	Concrete	56	52	53	57	38
CON3	Concrete	53	48	50	62	54
CON4	Concrete	55	50	51	60	64
CON5	Concrete	48	45	45	55	57
CON6	Concrete	53	50	51	59	55
MET1	Metal	36	31	33	39	33
MET2	Metal	33	29	31	32	35
MET3	Metal	36	31	34	37	33
MET4	Metal	26	23	23	31	26
MET5	Metal	30	26	26	32	32
MET6	Metal	34	28	31	30	34
MET7	Metal	30	28	28	33	36
RES1	Resin	27	23	25	25	23
ACR1	Acrylic	36	33	33	40	40
MIX1	Metal/acrylic	32	30	31	37	29
WOOD	Wood	34	30	30	34	27

that, for best comparability of the classical and transient methods, (i) the source used to emit the test signal should be the same in both cases or, at least, should have the same acoustical characteristics (directivity, volume, etc.) and (ii) peaks in the transient test signal should be not so high as to violate the linear acoustic assumptions. Davies and Gibbs[16] applied the impulse method to a freely suspended Perspex panel. They refined the measurement technique using a digitally generated pulse (square wave), inverted each two cycles and synchronously averaged 256 times. This test signal is much closer to a maximum length sequence (MLS) than to a pistol shot. Balilah and Gibbs,[17] continuing the work of Davies and Gibbs with a better instrumentation, tested different samples, in laboratory and *in situ*. Systematic computations of time of flights permitted the identification of the sound radiation due to bending wave propagation in the plates under test. Zuomin and Chu[18] proved that transmission loss measurements can be made in the laboratory using a MLS signal. They also provided an empirical formula to estimate the number of averages required to reach a desired accuracy in the presence of a given signal-to-noise ratio. All the above-mentioned authors were mainly concerned with building acoustics measurements; no tests were performed on noise barriers installed outdoors.

B. General principle

A loudspeaker is placed facing the traffic side of the barrier under test, a microphone is placed on the opposite side. The loudspeaker emits a test sound wave that is partly reflected, partly transmitted, and partly diffracted by the noise barrier (see Fig. 2). The microphone receives a signal that, suitably postprocessed, gives an overall impulse response. This includes the transmitted component, traveling from the sound source through the noise barrier to the microphone; the component diffracted by the top edge of the screen; and other ''parasitic'' components (reflected from the ground, both on the source and the receiver side; diffracted from the lateral edges, etc.). In particular, for the test to be meaningful, the diffraction from the lateral edges should be sufficiently weak and delayed. The transmitted sound pressure wave can be extracted by the global impulse response applying a suitable time window. If the measurement is repeated without the noise barrier between the loudspeaker and

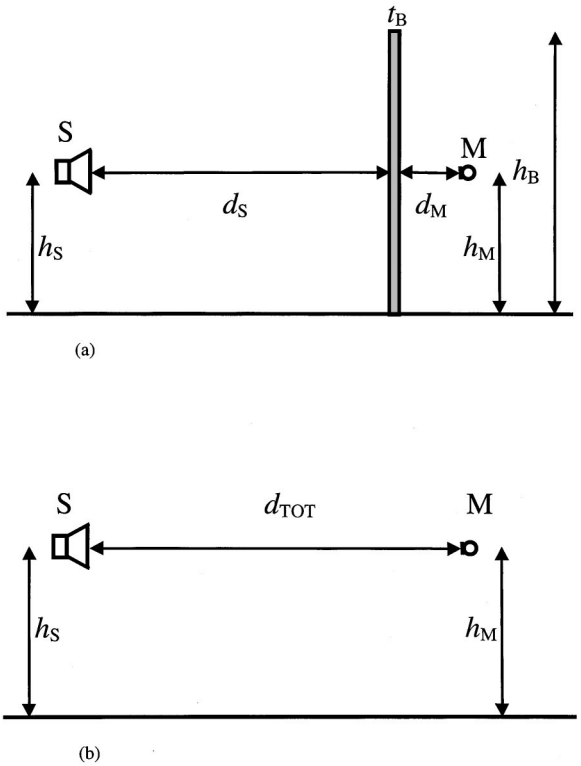


FIG. 2. Sketch of the setup for the sound insulation index measurement. Normal sound incidence of sound on the sample. S: loudspeaker. M: microphone. (a) Transmitted sound measurement. (b) Free-field (incident) sound measurement.

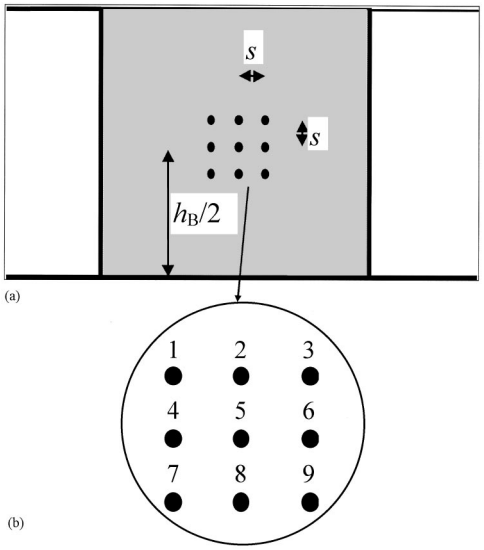


FIG. 3. (a) Grid of the microphone positions for the measurement by scanning at discrete points on a test section of a noise barrier (front view, receiver side). (b) Numbering of the measurement points.

the microphone (this is called ''free-field'' measurement in Ref. 7), the direct component alone can be sampled. The power spectra of the direct ''free-field'' component and the transmitted component, corrected to take into account the path length difference of the two signals, gives the basis for calculating the outdoor transmission loss [see Eq. (3)], which has been called *sound insulation index*.[10]

The measurement in front of the barrier is repeated at nine points placed on an ideal grid (scanning points). The final sound insulation index is the logarithmic average of the results in these nine positions (see Fig. 3).

C. Measured quantity

The expression used to compute the *sound insulation index SI* as a function of frequency is[7,10]

$$\mathrm{SI}_j = -10\,\lg\left\{\frac{\Sigma_{k=1}^{n}\int_{\Delta f_j}\left|\mathbf{F}[p_{tk}(t)w_{tk}(t)]\right|^2 df (d_k/d_i)^2}{n\cdot\int_{\Delta f_j}\left|\mathbf{F}[p_i(t)w_i(t)]\right|^2 df}\right\}\ (\mathrm{dB}), \tag{3}$$

where $p_i(t)$ is the reference free-field component; $p_{tk}(t)$ is the transmitted component at the kth scanning point; d_i is the geometrical spreading correction factor for the reference free-field component; d_k is the geometrical spreading correction factor for the transmitted component at the kth scanning point ($k=1,...,n$); $w_i(t)$ is the time window (*Adrienne* window) for the reference free-field component; $w_{t,k}(t)$ is the time window (*Adrienne* window) for the transmitted component at the kth scanning point; **F** is the symbol of the Fourier transform; Δf_j is the jth one-third octave frequency band (from 100 Hz to 5 kHz); and $n=9$ is the number of scanning points.

The geometrical spreading correction factors d_i (incident wave) and d_k ($k=1,...,9$) are (see Figs. 2 and 3)

$$d_i = d_5 = d_{SM} = d_s + t_B + d_M = 1{,}25 + t_B, \tag{4}$$

$$d_2 = d_4 = d_6 = d_8 = \sqrt{d_i^2 + s^2}, \tag{5}$$

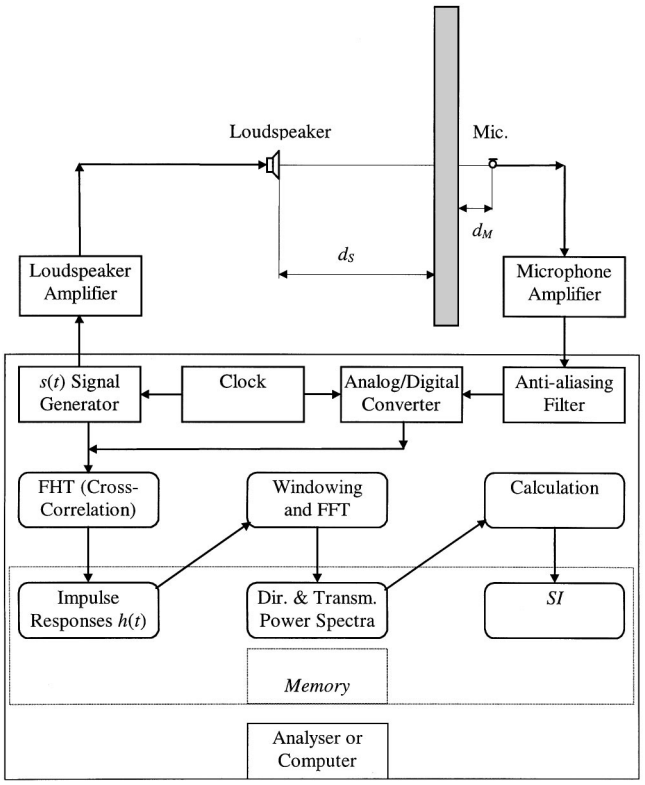


FIG. 4. Sketch representing the essential components of the measuring system.

$$d_1 = d_3 = d_7 = d_9 = \sqrt{d_i^2 + 2s^2}, \tag{6}$$

where t_B is the barrier thickness (m) and s is the measurement grid step (m) (see Fig. 3).

The frequency domain integration in Eq. (3) indicates the reconstruction of the one-third octave band values SI_j from narrow-band data. The European standards require results be presented in the one-third octave bands from 100 Hz to 5 kHz.

D. Measuring equipment and procedure

1. Measuring system

The measuring equipment comprises an electro-acoustical source, consisting of an electrical signal generator, a power amplifier and a loudspeaker; a microphone with its microphone amplifier; and a signal analyzer capable of performing transformations between the time domain and the frequency domain. Part of these devices can be integrated into a frequency analyzer or a personal computer equipped with specific add-on board(s). The essential components of the measuring system are shown in Fig. 4.

The impulse response is acquired using the well-known MLS technique[19–22] that allows the measurement be done without impulse excitation. A MLS is a deterministic sequence of bipolar pulses. Its frequency spectrum is white and its circular autocorrelation function is approximately a periodic unit-sample, with a negligible dc component. The loudspeaker, driven with an input electrical signal consisting of a MLS, excites the device under test (Fig. 4). The system impulse response is derived from the signal picked up by the microphone applying a fast Hadamard transform (FHT). This

is equivalent to cross correlating the microphone signal with the MLS fed to the loudspeaker.[20,21,23] The sample rate of the acquisition device must be equal to the clock rate of the MLS signal generator. The MLS technique is highly noise rejecting compared to more conventional ones;[21,24] the signal-to-noise ratio can be further improved by synchronous averaging.[25] An important prerequisite of the MLS technique is that the system under test must be linear and time invariant (LTI system).

An absolute calibration of the measurement chain with regard to the sound pressure level is not needed, because the measurement procedure described here is based on the ratio of the power spectra of signals extracted from impulse responses sampled with the same equipment in the same place under the same conditions.

For the *Adrienne* test method, the sample rate f_s must have a value greater than 43 kHz.[7] Although the signal is already unambiguously defined when the Nyquist criterion is met, higher sample rates facilitate a clear reproduction of the signal. Errors can be detected and corrected more easily, such as corrections needed to account for time shifts due to temperature changes. It is worth noting that *Adrienne* requirements are the same for sound insulation index and sound reflection index measurements. The *sound reflection index* is the quantity introduced to characterize the reflecting and diffusing property of noise barriers (not treated in the present article).[7,10] The *sound reflection index* is calculated using the signal subtraction technique,[26] which necessitates of an exact reproduction of time signals and therefore of high sample rate values.

2. Positioning of the measuring equipment

The measuring equipment must be placed near the noise reducing device to be tested according to the following rules.

A *source reference plane* is defined as the vertical plane tangential to the most protruding edges of the sample, on the traffic side. A *reference position* is located on this reference plane at a height equal to half the height h_B of the barrier under test. The loudspeaker is placed in front of the reference position at a horizontal distance d_S of 1 m from the source reference plane and at a height h_S equal to half the barrier height [Fig. 2(a)].

On the opposite side of the barrier under test *a microphone reference plane* is located and an *ideal measurement grid* is defined on it. The measurement grid must be squared, with a side length of 0.80 m. The measurement grid shall be at a distance d_M of 0.25 m from the plane of reference for the microphone. On the measurement grid, nine measurement points are located, each of them having horizontal and vertical distance from the neighboring points on the same alignment s of 0.40 m (see Fig. 3). The microphone is subsequently placed in each of the nine measurement points, at a horizontal distance d_M of 0.25 m from the microphone plane of reference, and an impulse response is sampled in each measurement point. When the microphone is in the central position of the measurement grid, the acoustic center of the sound source and the acoustic center of the microphone must lie on the same line normal to the two planes of reference and must have the same height, equal to half the barrier height [Fig. 2(a)].

A free-field measurement is taken displacing the loudspeaker and the microphone far from any nearby object. In particular, the distance of the microphone from the sound source must be kept equal to the distance when the equipment is in the position 5 [see Fig. 2(b)]:

$$d_{\mathrm{TOT}} = d_S + t_B + d_M = 1{,}25 + t_B\,. \tag{7}$$

The nine measurements taken on the measurement grid plus the corresponding free-field measurement shall be processed and averaged according to the sound insulation index formula, Eq. (3). Figure 8 shows the single results of the nine measurements taken in front of the barrier elements and the final logarithmic average for sample MET1.

For noise barriers constituted by one or several acoustic elements sustained by vertical posts at fixed distances, a set of nine measurements on the measurement grid plus a free-field measurement must be performed both close to the middle of a representative element and facing a representative post. This permits the detection of the two most common kinds of sound leaks, which are usually located at panel–panel and panel–post connections.

If it is suspected that sound leaks may exist at a different position, e.g., at the bottom edge of the barrier under test, a further set of nine measurements on the measurement grid plus a free-field measurement can be performed placing the measuring equipment close to that position. In this particular case, the sound signal coming from the bottom edge is no longer a ''parasitic'' signal: it is of course, the ''transmitted'' signal one is looking for. The *Adrienne* window must then be enlarged so as to include this signal and to avoid other parasitic signals, on the basis of a geometrical computation to be shown on the test report. Among the possible parasitic signals, the ground reflection on the receiver side is not of concern, as the apparent sound source, i.e., the leak, is located on the ground.

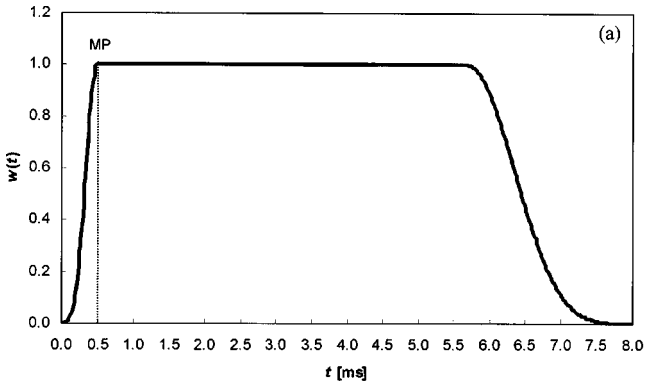
Any object other than the device under test shall be considered a reflecting object which could cause parasitic reflections (e.g., safety rails, fences, rocks, parked cars, etc.). These objects must remain far from the microphone.

3. Temporal window

For the *Adrienne* test method, the analysis window must be uniquely defined in shape, length, and position. The analysis window must be the new *Adrienne* window, described in Refs. 7 and 10. It is a composite analysis window, built as follows (see also Fig. 5):

(i) a leading edge having a left-half Blackman-Harris shape and a total length of 0.5 ms;
(ii) a flat portion having a total length of 5.18 ms (main body); and
(iii) a trailing edge having a right-half Blackman-Harris shape and a total length of 2.22 ms.

The total length of the *Adrienne* window is 7.9 ms; if the window length has to be varied (this occurs only in exceptional cases), the lengths of the flat portion and the right-half



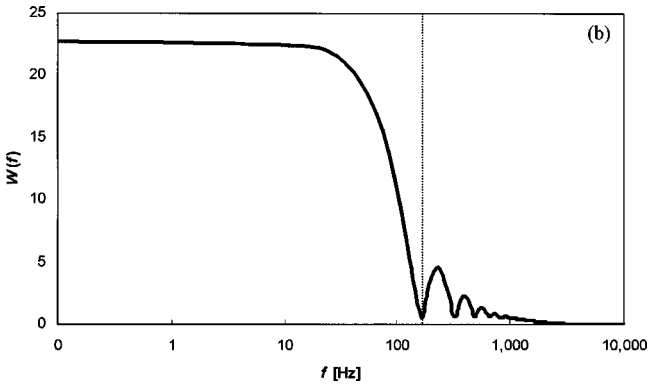


FIG. 5. The *Adrienne* analysis window. (a) In the time domain, the dotted line shows the marker point MP. (b) In the frequency domain (positive frequencies only), the dotted line shows the end of the main lobe.

Blackman-Harris portion must have a ratio of 7/3. The point where the flat portion of the Adrienne window begins is called the marker point (MP).

For the direct ''free-field'' component, the window is placed as follows (Fig. 6):

(i) the maximum peak of the impulse response is detected;
(ii) a time instant preceding the direct component peak of 0.2 ms is located; and
(iii) the *Adrienne* window is placed so as its marker point corresponds to this time instant.

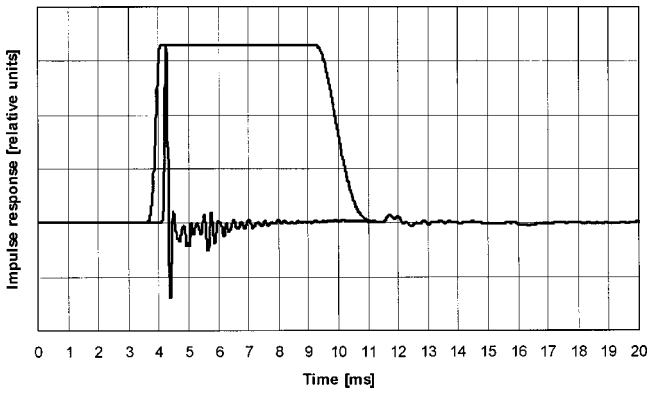


FIG. 6. The *Adrienne* window applied at the free-field component (sample MET1).

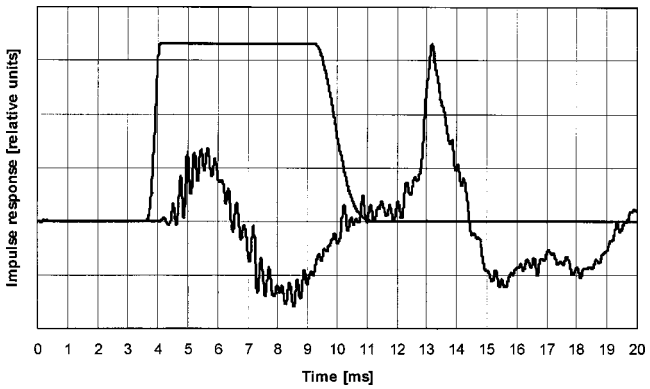


FIG. 7. The *Adrienne* window applied at the transmitted component; position 5 in front of an acoustic element (sample MET1).

In other words, the direct component *Adrienne* window is placed so as its flat portion begins 0.2 ms before the direct component peak.

For the transmitted component, the window shall be placed as follows (Fig. 7):

(i) the time instant when the transmission begins is located, possibly with the help of geometrical computation (conventional beginning of transmission);
(ii) a time instant preceding the conventional beginning of transmission of 0.2 ms is located;
(iii) the transmitted component *Adrienne* window is placed so as its marker point corresponds to this time instant;
(iv) the time instant when the diffraction begins is located, possibly with the help of geometrical computation (conventional beginning of the diffraction); and
(v) the transmitted component *Adrienne* window stops 7.4 ms after the marker point or at the conventional beginning of the diffraction, whichever of the two comes first.

In other words, the transmitted component *Adrienne* window is placed so that its flat portion begins 0.2 ms before the first peak of the transmitted component and its tail stops before the beginning of the diffraction.

Using these rules, the placement of the analysis window can be done automatically, without relying on the operator's skills (actually, the rules are implemented in the program ALFA©, developed by the authors and used for processing the data presented in this article). In computations involving the sound speed *c*, its temperature-dependent value must be assumed.

It is worth noting that the first peak of the direct ''free-field'' component is also the maximum peak. This is not necessarily true for the transmitted component, whose shape depends on the structural and vibrational characteristics of the barrier under test. For example, Fig. 7 shows the impulse response measured at position 5 in front of the panels of the metallic barrier MET1. Geometrical computation confirms that the windowed component is the transmitted one, while the maximum peak coming after it corresponds to the diffraction over the top edge of the screen.

The window length of 7.4 ms after the marker point can

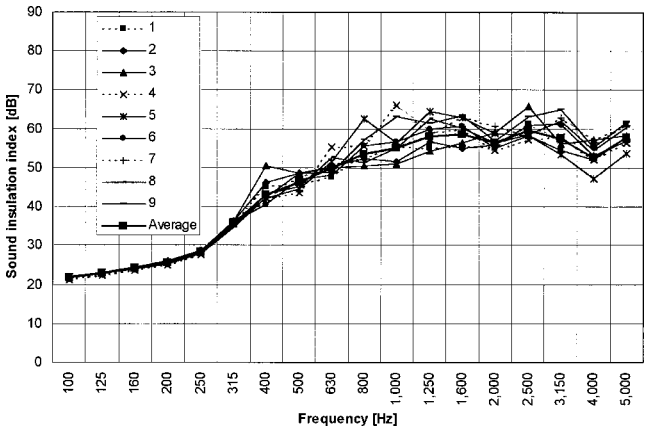


FIG. 8. Sound insulation index values measured at the nine scanning positions in front of the acoustic elements and final logarithmic average (sample MET1).

be explained with an example, assuming for the sake of simplicity an infinitely thin screen ($t_B \approx 0$) with a height h_B of 4 m. Let d_{SE} be the distance between the loudspeaker and the diffracting to edge and d_{EM} the distance between the diffracting top edge and the microphone in the central top position, position 2 (Figs. 9 and 3). The minimum path length of the diffracted signal is

$$d_{SEM} = d_{SE} + d_{EM} = \sqrt{d_S^2 + \left(\frac{h_B}{2}\right)^2} + \sqrt{d_M^2 + \left(\frac{h_B}{2} - s\right)^2} = 3.86 \ \text{(m)}, \tag{8}$$

where h_B is the barrier height. The minimum path length of the transmitted signal is

$$d_{SM} = \sqrt{(d_S + d_M)^2 + s^2} = 1.31 \ \text{(m)}. \tag{9}$$

The time delay gap corresponding to the path length difference between the transmitted and diffracted signals therefore is

$$\tau = \frac{d_{SEM} - d_{SM}}{c} = 7.4 \ \text{(ms)}, \tag{10}$$

exactly the *Adrienne* window length, apart from the ''prewindow'' (left Blackman-Harris edge of 0.5 ms).

4. Low-frequency limit

The low-frequency limit of sound insulation index measurements is inversely proportional to the width of the analysis window and depends also on its shape; taking the first notch in the magnitude spectrum of the window as an indicator of the low-frequency limit,[27,28] for an *Adrienne* window 7.9 ms wide, this limit is about 160 Hz [see Fig. 5(b)].[29] It can be reduced increasing the window width, which is possible when the height of the noise barrier under test is greater than 4 m. Strictly speaking, the outdoor measured values shown in the following for 4-m-tall barriers are therefore valid only starting from the 200-Hz one-third octave band. In spite of this, on the graphs (Figs. 10–17) also the values measured in the 100- to 160-Hz one-third octave bands have been retained, just to have an idea of their consistency with laboratory results. The correlations between outdoor and laboratory results presented in Sec. V have of course been obtained on the valid frequency range 200 Hz to 5 kHz.

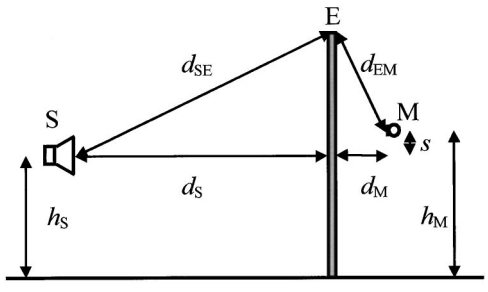


FIG. 9. For the computation of the *Adrienne* window length (see text).

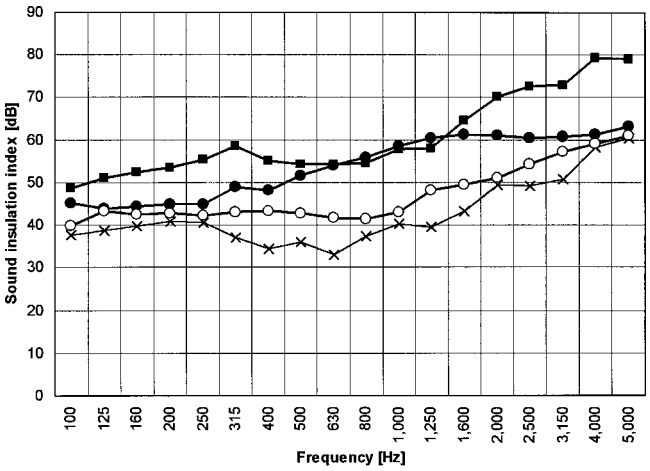


FIG. 10. Sound insulation index values for barrier CON2: (●) laboratory measurements—accurate seal; (○) laboratory measurements—quick seal; (■) outdoor measurements—elements; (×) outdoor measurements—post.

IV. OUTDOOR MEASUREMENTS

A. Measuring equipment

The measurement system was assembled using the following devices:

(1) Sound source: prototype diffusor built inserting a single loudspeaker driver (JBL type 2123H, diameter 250 mm) in a closed cabinet with parallelepiped shape.
(2) Loudspeaker amplifier: QSC USA 1300.
(3) Parametric equalizer: BSS Audio FCS-926 ''varicurve'' with 12 parametric filters.
(4) Microphone Brüel & Kjær type 4190, with Brüel & Kjær type 2669 preamplifier and Brüel & Kjær type 2807 power supply.
(5) Sampling digitizer board: A2D-160 board, containing the MLS generator, the programmable eight-pole antialiasing filter and the A/D converter (having an effective resolution of 16 bits).
(6) Toshiba T6600C portable computer, containing the A2D-160 board.

The A2D-160 board is driven by the MLSSA© software[30] for the generation of the MLS signal and the acquisition of the impulse response; the rest of the signal processing is done using ALFA©, a special-purpose software developed by the authors. The test signal was a MLS sequence of order 16

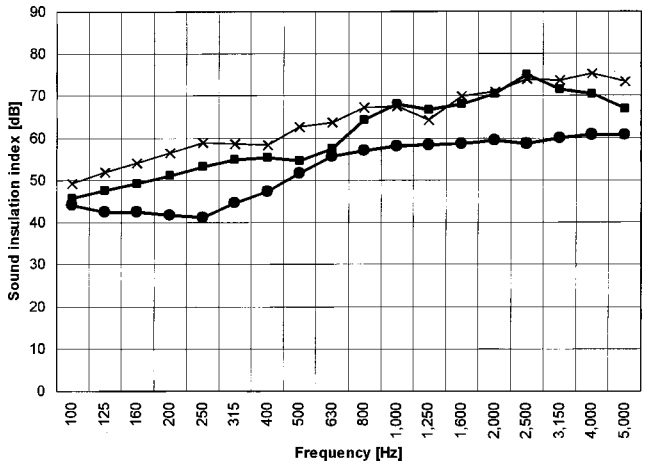


FIG. 11. Sound insulation index values for barrier CON4: (●) laboratory measurements; (■) outdoor measurements—elements; (×) outdoor measurements—post.

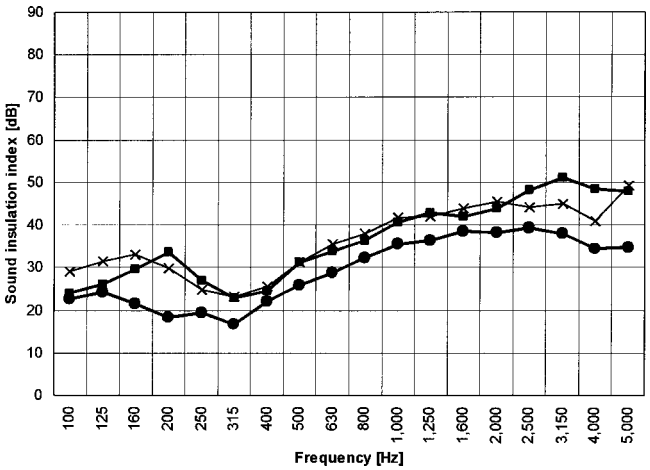


FIG. 14. Sound insulation index values for barrier MET5: (●) laboratory measurements; (■) outdoor measurements—elements; (×) outdoor measurements—post.

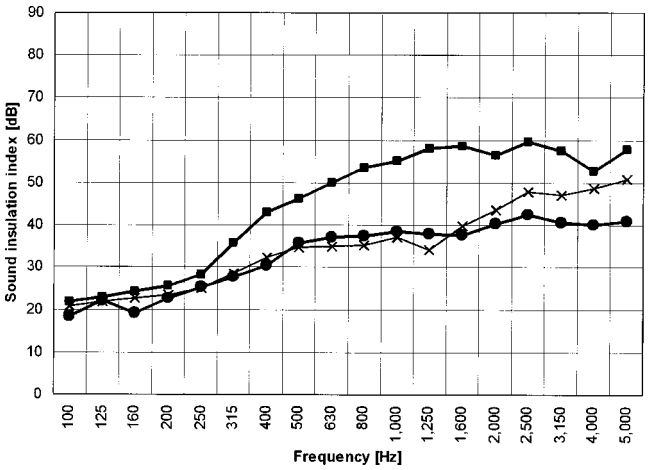


FIG. 12. Sound insulation index values for barrier MET1: (●) laboratory measurements; (■) outdoor measurements—elements; (×) outdoor measurements—post.

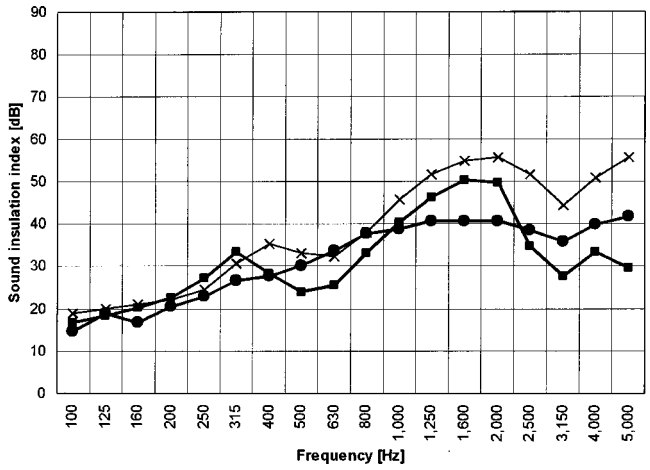


FIG. 15. Sound insulation index values for barrier MET6: (●) laboratory measurements; (■) outdoor measurements—elements; (×) outdoor measurements—post.

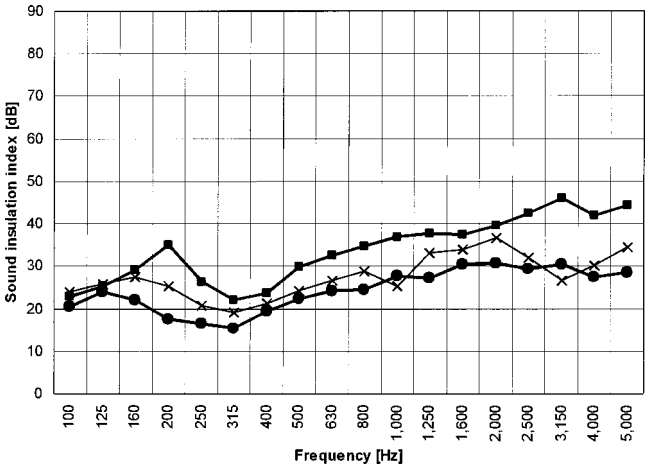


FIG. 13. Sound insulation index values for barrier MET4: (●) laboratory measurements; (■) outdoor measurements—elements; (×) outdoor measurements—post.

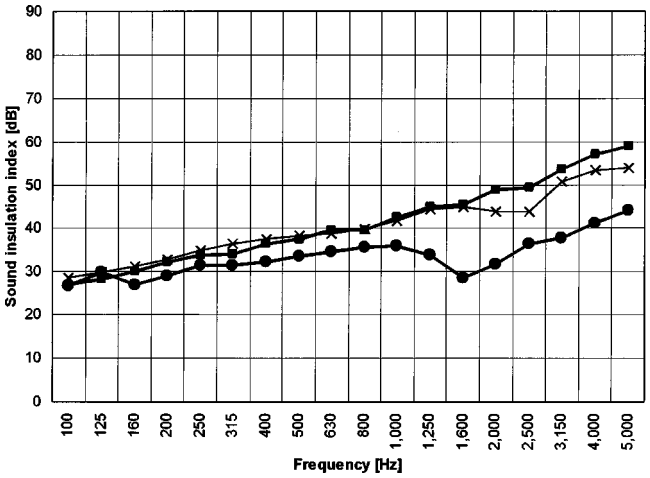


FIG. 16. Sound insulation index values for barrier ACR1: (●) laboratory measurements; (■) outdoor measurements—elements; (×) outdoor measurements—post.

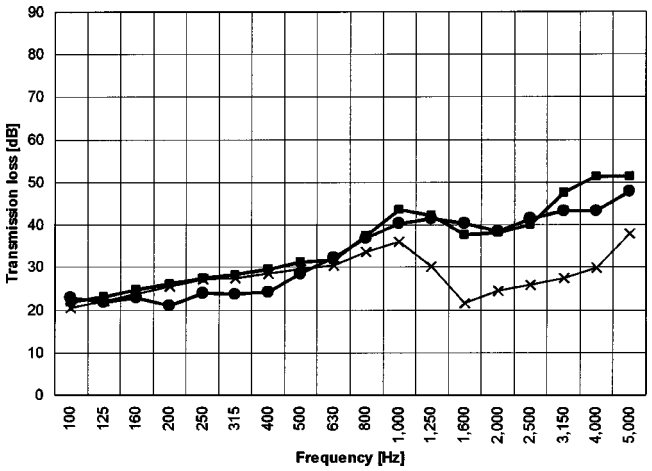


FIG. 17. Sound insulation index values for barrier WOOD: (●) laboratory measurements; (■) outdoor measurements—elements; (×) outdoor measurements—post.

generated by the A2D-160 board;[30] 64 averages were performed for each impulse response acquisition.

B. Outdoor test site

The test site is a flat, grass-covered ground. The grass was cut before the beginning of the tests. No trees, stones, or other reflecting objects were present in a $50\times50\,\mathrm{m}^2$ area around the noise barrier under test. All samples were built in the same place and removed after the test, one after the other. Measurements were taken in good meteorological conditions, with no rain or strong wind (wind speed always <4 m/s).

Background noise was very limited, the only noise source being a local road, where few cars pass every day, 50 m away from the test samples. Moreover, it is well known that MLS technique is strongly noise rejecting.[19,21,23] In any case, when some occasional noise occurred, like airplane flyovers, measurements were repeated and it was verified that the extraneous noise had no influence.

C. Outdoor measurement results

Examples of outdoors measurements are reported in graphical form in Figs. 10–17 together with the corresponding laboratory measurements (see Sec. II). In each graph, three curves are shown, because, as stated in Sec. III D 2, for each noise barrier the outdoor measurement procedure was repeated two times, placing the measuring equipment first close to the acoustic elements and then close to a representative post. This permitted the investigation of the two most common kinds of sound leak, which are usually located at panel–panel and panel–post connections.

For each barrier sample tested outdoors, the single number rating DL_{SI} was computed, using the formula (2) with SI_j in place of R_j. Due to the above-mentioned low-frequency limit of outdoor measurements (see Sec. III D 4), the calculations were performed in the one-third frequency bands from 200 Hz to 5 kHz. The results are reported in Table III.

V. COMPARISON BETWEEN LABORATORY AND OUTDOOR DATA

There is a general tendency for the laboratory results to be lower than the outdoors results for the same kind of barrier. This may be due to the different sound fields in front of the test specimen: diffuse field in laboratory, frontal free-field outdoors. In fact, the oblique components of the diffuse field generate the coincidence effect, which of course is not possible outdoors. This is particularly clear in Fig. 16 for sample ACR1, constituted by a simple homogeneous acrylic sheet: the laboratory curve exhibits an evident coincidence dip in the 1600-Hz one-third octave band; for outdoors measurements this effect is absent. Moreover, the steady-state signal recorded in the laboratory is very different from the impulse response recorded outdoors (Raes[13] distinguished between ''static'' and ''dynamic'' transmission loss). Finally, the boundary conditions for the laboratory and the outdoors test samples are very different: rigidly clamped on four sides in the laboratory, relatively free on three sides outdoors. Therefore, differences between laboratory and outdoor values were expected. Actually, one of the aims of this work was to find a correlation between airborne sound insulation values measured using the laboratory and *Adrienne* procedures (see later in this article).

In most cases, sound insulation index values measured outdoor in front of a post are worse than values measured in front of the acoustic elements, especially at high frequency (see, for example, Figs. 10, 12, and 13). This happens when the connections between the acoustic elements and the posts are not perfect and may depend not only on the workmanship, but also on the design of connections and the lack of good seals. In these cases, the laboratory performance is influenced by the element/post connections and is closer to the outdoor performance in front of a post. This confirms the importance of including a post in the test, both in laboratory and outdoors. For the concrete barrier CON2 (Fig. 10), the laboratory test was repeated two times, the first with a

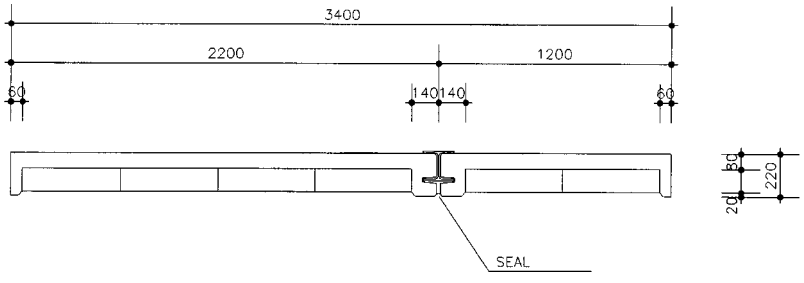


FIG. 18. Barrier CON4: plan view of typical components. Dimensions in millimeters.

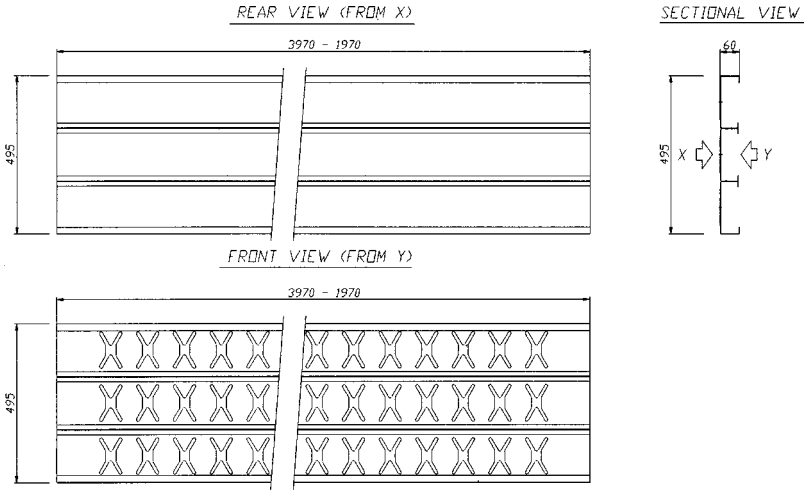


FIG. 19. Barrier MET4: front, rear, and sectional view of the sample for outdoor measurements. Dimensions in millimeters.

''quick'' seal at posts—similar to those used outdoors—and the second with an accurate seal. As can be seen, with the quick seal the laboratory performance is closer to the outdoor performance measured in front of a post, while with the accurate seal the laboratory performance is closer to the outdoor performance in front of barrier panels. The two different cases are indicated in Table III with CON2(Q) and CON2(A), respectively.

The concrete barrier CON4 gives an apparently anomalous result: the post performance is better than the element performance, and both are better than the laboratory result (Fig. 11). This can be easily explained looking at the barrier structure: posts are divided in two halves, each one being a massive concrete beam, 220 mm thick; each half-post forms a single block with the back panel of a barrier element, 80 mm thick, which supports light clay aggregate blocks, 120 mm thick (Fig. 18). The blocks contain cavities connected to front holes on the exposed face and are intended to act as resonators to improve the sound absorption. Connections are formed between half-posts, blocking them on a steel beam and adding a further seal. From this design, one can expect a better outdoor performance close to posts. In the laboratory, the use of a diffuse sound field results in lower airborne sound insulation values.

The sample MET1 (Fig. 12) is a metallic barrier with evident problems of connections between elements and posts; outdoor the difference is remarkable (6 dB in DL_{SI} values); in the laboratory the result is strongly conditioned by the presence of the post.

The samples MET4 and MET5 (Figs. 13 and 14) are actually the same barrier: MET4 is constituted of 1.0-mm metallic sheets, folded to form protruding supports, 60 mm wide (Fig. 19). MET5 is MET4 plus a lining in polyester fiber wool panels (see Table I). For barrier MET5 all joints were carefully sealed. With the additional treatment the overall performance is better (in terms of single number rating, 3-dB gain in the laboratory and 6-dB gain outdoors close to posts; see Table III); for barrier MET5 the single number rating DL_{SI} obtained close to a post is equal to that obtained close to metallic panels.

For the sample MET6 (Fig. 15) the metallic sheets are relatively free to vibrate. This conditions the outdoor performance: the sound insulation index is better close to posts, made with a rigid steel beam. In the laboratory, where the test specimen is clamped on four sides, the mobility is reduced and the sound transmission can be influenced.

For the barrier WOOD, the low values at high frequency of the sound insulation index measured outdoors close to a post (Fig. 17) are due to an evident leakage at panel–post connection, detectable by visual inspection.

Comparing the values reported in Table III with the categories shown in Table II, it is worth noting that all samples, excluding MET4 (laboratory test) and RES1 (laboratory test 100 Hz to 5 kHz and outdoor post test), got a category B3 of airborne sound insulation, according to EN 1793-2; the present EN classification does not allow discrimination among barriers with a single number rating greater than 24 dB.

Looking at Figs. 10–17, it can also be noticed that outdoor measured values in the one-third octave frequency bands from 100 to 160 Hz, while outside the range of validity as discussed in Sec. III D 4, are consistent with the laboratory measured values.

The application of standard statistical theory to data of Table III permitted us to obtain linear correlation laws be-

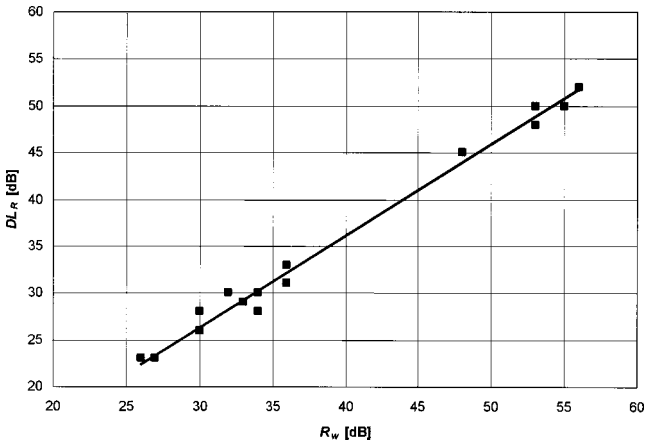


FIG. 20. Linear correlation between the single number ratings R_w and DL_R obtained from laboratory data. Frequency range: one-third octave bands from 100 Hz to 5 kHz.

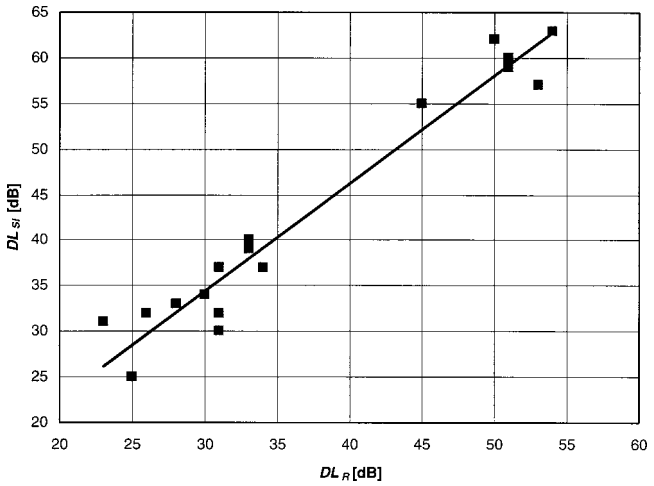


FIG. 21. Linear correlation between the single number ratings obtained in laboratory (DL_R) and outdoors in front of the barrier acoustic elements (DL_{SI}). Frequency range: one-third octave bands from 200 Hz to 5 kHz.

tween airborne sound insulation rating figures obtained using the laboratory and *Adrienne* procedures.

Figure 20 shows the correlation between the values of the two single number ratings obtained from laboratory measurements, calculated over the frequency range 100 Hz to 5 kHz, the DL_R specified in EN 1793-2 and the R_w specified in ISO 717-1:[11]

$$DL_R = 0.98R_w - 3.05 \quad (r = 0.995). \tag{11}$$

Using the same values of R_w and the values of DL_R calculated over the frequency range 200 Hz to 5 kHz, the correlation becomes

$$DL_R = 0.93R_w + 0.37 \quad (r = 0.992). \tag{12}$$

The controlled conditions of the tests and the excellent value of the correlation coefficient r support the conclusion that on average the EN single number rating DL_R is few decibels lower than the index R_w used in building acoustics.

Figures 21 and 22 show the linear correlation laws between the single number rating DL_R obtained from laboratory data and the single number rating DL_{SI} obtained from outdoor data measured close to barrier elements and to posts, respectively. All ratings are calculated over the frequency range 200 Hz to 5 kHz. For barrier CON2, submitted as previously explained to two laboratory tests with different seals at posts, the ''quick'' seal rating (44 dB) was taken for the correlation with the outdoor ratings of measurements close to a post and the accurate seal rating (53 dB) was taken for the correlation with the outdoor ratings of measurements close to the acoustic elements:

$$\text{Elements: } DL_{SI} = 1.18DL_R - 0.94 \quad (r = 0.97), \tag{13}$$

$$\text{Posts: } DL_{SI} = 1.18DL_R - 3.16 \quad (r = 0.93). \tag{14}$$

The linear correlation coefficient r is excellent for elements and very good for posts; this difference was expected, because outdoor results are less regular at posts due to the above-mentioned problems of panel–post connections. In any case, the high values of the correlation coefficient support the conclusion that Eqs. (13) and (14) can be useful for predicting the expected field performance from laboratory data measured according to EN 1793-2.

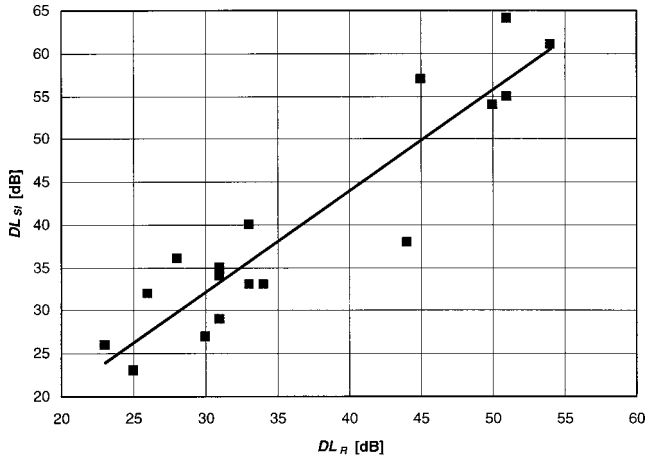


FIG. 22. Linear correlation between the single number ratings obtained in laboratory (DL_R) and outdoors in front of barrier posts (DL_{SI}). Frequency range: one-third octave bands from 200 Hz to 5 kHz.

VI. CONCLUSIONS

A detailed verification of the *Adrienne* test method for airborne sound insulation over a selection of 17 noise barriers, representative of the Italian and European production, has been done. The new method proved to be easy to use and reliable for all kinds of barrier. It has been found sensitive to quality of mounting, presence of seals, and other details typical of outdoor installations. The comparison between field and laboratory results shows a very good correlation, while existing differences can be explained with the different sound fields and mounting conditions between the outdoor and laboratory tests. In other words, results obtained using the *Adrienne* test method are consistent with laboratory results obtained using EN 1793-2. The correlation laws resulting from the present work can be useful for predicting the airborne sound insulation performance of noise barriers in the field from laboratory data. It can be concluded that the *Adrienne* method is adequate for its intended use.

ACKNOWLEDGMENTS

The measurements reported in this work were done in 1999 by the authors for *Italferr*, the engineering company of *Ferrovie dello Stato* (Italian Railways) responsible for the technical supervision of the Italian high speed railway network. The laboratory tests were performed in the facilities of *Istituto Giordano* (Bellaria, Italy), in the frame of a contract with *DIENCA*. In particular, the authors of this article would like to thank Andrea Bruschi for his invaluable help during laboratory measurements. The outdoor test method used during this work was developed in the frame of the European project *Adrienne*, funded by the European Commission (Contract No. MAT1-CT94049). The main partners of the research were *Acoustical Technologies* (B), *Fraunhofer Institut für Bauphysik* (D), *ENTPE-LASH* (F), and *DIENCA* (I). Other associated partners were *CSTB* (F), *LCPC* (F), *IA-CSIC* (E), *FIGE* (D), and *CEDIA* (B).

[1] I-INCE Publication 98-1, ''Technical assessment of the effectiveness of noise walls,'' Noise/News International **6**, 11–35 (1998).
[2] EN 1793-1, ''Road traffic noise reducing devices—Test methods for determining the acoustic performance—Part 1: Intrinsic characteristics of sound absorption'' (1997). (Information on how CEN Standards can be obtained is available on the Web site http://www.cenorm.be)
[3] EN 1793-2, ''Road traffic noise reducing devices—Test methods for determining the acoustic performance—Part 2: Intrinsic characteristics of airborne sound insulation'' (1997). (Information on how CEN Standards can be obtained is available on the Web site http://www.cenorm.be)
[4] EN 1793-3, ''Road traffic noise reducing devices—Test methods for determining the acoustic performance—Part 3: Normalized traffic noise spectrum'' (1997). (Information on how CEN Standards can be obtained is available on the Web site http://www.cenorm.be)
[5] ISO 354, ''Acoustics—Measurement of sound absorption in a reverberation room'' (1985).
[6] ISO 140-3, ''Acoustics—Measurements of sound insulation in buildings and of buildings elements. Part 3: laboratory measurements of airborne sound insulation of building elements'' (1995).
[7] *Adrienne* Research Team, ''Test methods for the acoustic performance of road traffic noise reducing devices—Final report,'' European Commission, DGXII SMT Project MAT1-CT94049 (1998).
[8] AFNOR S 31-089, ''Code d'essai pour la détermination de caractéristiques acoustiques d'écrans installés en champ libre'' (1990).
[9] J.-P. Clairbois, J. Beaumont, M. Garai, and G. Schupp, ''A new *in-situ* method for the acoustic performance of road traffic noise reducing devices,'' J. Acoust. Soc. Am. **103**, 2801(A) (1998).
[10] F. Anfosso, M. Garai, and J-P. Clairbois, ''*Adrienne*: une méthode européenne pour la qualification sur site des écrans antibruit,'' Bullettin des Laboratoires des Ponts et Chausées, n. 225, 89–104 (2000).
[11] ISO 717-1, ''Acoustics—Rating of sound insulation in buildings and of buildings elements. Part 1: airborne sound insulation'' (1996).
[12] A. C. Raes, ''A tentative method for the measurement of sound transmission losses in unfinished buildings,'' J. Acoust. Soc. Am. **27**, 98–102 (1955).
[13] A. C. Raes, ''Static and dynamic transmission losses of partitions,'' J. Acoust. Soc. Am. **35**, 1178–1182 (1963).
[14] P. de Tricaud, ''Impulse techniques for the simplification of insulation measurement between dwellings,'' Appl. Acoust. **8**, 245–256 (1975).
[15] J. Roland, ''Airborne isolation measurements by impulse technique,'' Noise Control Eng. J. **16**, 6–14 (1981).
[16] J. C. Davies and B. M. Gibbs, ''The oblique incidence measurement of transmission loss by an impulse method,'' J. Sound Vib. **74**, 381–393 (1981).
[17] Y. A. Balilah and B. M. Gibbs, ''The measurement of the transmission loss of single leaf walls and panels by an impulse method,'' J. Sound Vib. **123**, 229–245 (1988).
[18] W. Zuomin and W. T. Chu, ''Ensemble average requirement for acoustical measurements in noisy environment using the *m*-sequence correlation technique,'' J. Acoust. Soc. Am. **94**, 1409–1414 (1993).
[19] M. R. Schroeder, ''Integrated impulse method measuring sound decay without impulses,'' J. Acoust. Soc. Am. **66**, 497–500 (1979).
[20] J. Borish and J. B. Angell, ''An efficient algorithm for measuring the impulse response using pseudorandom noise,'' J. Audio Eng. Soc. **31**, 478–488 (1983).
[21] D. D. Rife and J. Vanderkooy, ''Transfer-function measurement with maximum-length sequences,'' J. Audio Eng. Soc. **37**, 419–444 (1989).
[22] J. Vanderkooy, ''Aspects of MLS measuring systems,'' J. Audio Eng. Soc. **42**, 219–231 (1994).
[23] M. Garai, ''Measurement of the sound-absorption coefficient *in situ*: the reflection method using periodic pseudo-random sequences of maximum length,'' Appl. Acoust. **39**, 119–139 (1993).
[24] C. Bleakley and R. Scaife, ''New formulas for predicting the accuracy of acoustical measurements made in noisy environments using the averaged *m*-sequence correlation technique,'' J. Acoust. Soc. Am. **97**, 1329–1332 (1995).
[25] M. Vorländer and M. Kob, ''Practical aspects of MLS measurements in building acoustics,'' Appl. Acoust. **52**, 239–258 (1997).
[26] E. Mommertz, ''Angle-dependent *in-situ* measurements of reflection coefficients using a subtraction technique,'' Appl. Acoust. **46**, 251–263 (1995).
[27] S. Gade and H. Herlufsen, ''Use of weighting functions in DFT/FFT analysis (Part I),'' Brüel & Kjær Technical Review no. 3 (1987), pp. 1–28.
[28] S. Gade and H. Herlufsen, ''Use of weighting functions in DFT/FFT analysis (Part II),'' Brüel & Kjær Technical Review no. 4 (1987), pp. 1–35.
[29] P. Cobo, ''Some calculations concerning the *Adrienne* setup and the lowest reliable frequency,'' Report to the *Adrienne* Research Team (1998).
[30] D. D. Rife, *MLSSA Reference Manual vr. 10.0A* (DRA Laboratories, Sterling, VA, 1996).

The effect of high-amplitude sound on the attenuation of perforated tube silencers

N. S. Dickey and A. Selamet
Department of Mechanical Engineering, The Ohio State University, Columbus, Ohio 43210

J. M. Novak
Ford Motor Company, Powertrain Operations, Dearborn, Michigan 48121

(Received 18 October 1999; accepted for publication 26 May 2000)

A time-domain computational approach is applied to investigate the behavior of perforated tube silencers at high sound levels. The one-dimensional computational technique employs a lumped parameter model for the perforate flows. The lumped parameter perforate model is based on time-invariant approximations for the equivalent length l_{eq} and resistance R, consistent with the use of a nonlinear perforate impedance. Empirical expressions for l_{eq} and R are developed experimentally using nondimensional scaling parameters. The model is applied to geometries representative of automotive resonators and multiple-duct mufflers. Conditions are simplified from those in an actual automotive system by considering single-frequency excitation and zero mean flow. Simulations with linear perforate behavior are compared to experimental data obtained with an extended impedance tube system. Simulations with nonlinear perforate behavior for one concentric tube resonator are compared to published experimental data. © *2000 Acoustical Society of America.* [S0001-4966(00)01409-0]

PACS numbers: 43.50.Gf, 43.25.Ed [MRS]

INTRODUCTION

Perforated tubes are common components in silencers for automotive engines, often representing the primary element(s) in exhaust system mufflers. Typically these silencers include one or more perforated tubes that are installed within an external expansion chamber. In an automotive muffler, the perforated tube section is often combined with series and parallel connections to other elements (end chambers, resonators, etc.). Despite their widespread use and a number of informative studies, the acoustic behavior of perforated tube silencers under realistic operating conditions is still not well understood.

In analyzing the acoustic behavior of a perforated tube silencer (in the absence of an engine and other system components), the primary difficulty is modeling the complex behavior of the flow through the perforations. Flow visualization and computational studies of unsteady flows through circular orifices have shown that complex vortical flow patterns are established (see, for example, Refs. 1–3). Thus, analytical treatments are only available for simple, rather unrealistic conditions, and predictive tools must rely on experimental data and empirical models. Although the behavior of an arbitrary perforate usually can not be predicted quantitatively, previous experimental studies have provided a general qualitative understanding of the factors that affect the perforate flow.[1,4–7]

In the limit of low sound-pressure levels, the unsteady flow through perforations is linearly related to the pressure difference across the interface. As the sound-pressure level across the interface is increased, corresponding to higher amplitude oscillations in the perforate velocity, the pressure/flow relationship for the perforate becomes nonlinear. For this case, the commonly used perforate impedance is no longer independent of the perforate velocity, resulting in a decreasing reactance and increasing resistance as velocity amplitudes increase. In an automotive silencer, pressure amplitudes are expected to be large enough to cause this nonlinear behavior. The point at which nonlinear perforate behavior becomes significant depends on a number of factors but, as an example, orifice nonlinearity with zero mean flow has been observed at sound-pressure levels of 120–130 dB (*re* 20 μPa).[8] In comparison, oscillating pressure amplitudes in an automotive exhaust system can easily exceed 170 dB.[9] In addition to sound-pressure level, the behavior of the perforate is influenced by local fluid properties (density, viscosity, etc.), mean flow tangential to the perforated interface (grazing flow), mean flow through the perforations (throughflow) and, in the nonlinear regime of behavior, the spectral content of the sound.

The practical significance of perforated tube silencers for reducing radiated engine noise has motivated many works focusing on their acoustic behavior. The majority of these methods have been frequency domain approaches based on the assumptions of linear duct acoustics and linear perforate behavior. Sullivan and Crocker[8] presented the first such analysis, after which several techniques were developed to solve the resulting coupled sets of ordinary differential equations.[10–12] These solution techniques have subsequently been used in numerous studies which will not be listed here. While linear analyses have proven very useful for small-amplitude disturbances, they are unable to include the effects of perforate nonlinearity.

Although quite limited, some information is available regarding the effects of nonlinear perforate behavior on perforated tube silencers. Following his original work, Sullivan[13] presented a frequency domain segmentation approach that allows for spatial changes in the perforate imped-

ance caused by nonlinearity. Good agreement with experiments was obtained with the segmentation model for a small concentric tube resonator. Results from this study clearly demonstrated that the common practice of analyzing a silencer based on linear perforate behavior can lead to large discrepancies between the expected and actual attenuation characteristics. An alternative approach was taken by Chang and Cummings,[14] who applied a numerical time-domain technique to model another small (albeit somewhat longer) concentric tube resonator. The wave propagation in ducts was assumed to be linear, but a time-varying nonlinear perforate model was used. Depending on local conditions, the perforate behavior was assumed to be either linear or quasi-steady. Although the binary nature of the orifice model is rather undesirable, this method allows for temporal and spatial changes in the perforate behavior. Good correlations between predicted and experimental pressure traces were obtained for high-amplitude disturbances.

The objective of the present study is to model and investigate the behavior of perforated tube silencers at high-amplitude sound levels (up to 160 dB) with a time-domain computational approach. The work focuses specifically on the effects of nonlinear perforate behavior, and thus considers isolated silencers with zero mean flow subjected to single-frequency excitation. An experimental investigation of the unsteady flow through circular orifices is included for addressing the effects of nonlinear perforate behavior and developing suitable perforate submodels for the time-domain computational approach. The model is applied to geometries representative of automotive resonators and multiple-duct mufflers. Computational predictions for linear perforate behavior are compared to experimental data for each silencer geometry. For nonlinear perforate behavior, computational predictions for one concentric tube resonator are compared to the published data of Sullivan.[13]

I. COMPUTATIONAL APPROACH

The computational approach assumes that the duct flows in the silencer are one-dimensional in the axial direction. A lumped parameter perforate model is used to relate the local perforate mass flux to the pressure difference across the perforated interface. The assumption of one-dimensional duct flow is justified provided that the wavelengths of interest are significantly longer than the cross dimensions of a silencer. The upper frequency limit that this imposes is generally high enough for the analysis to include the engine firing fundamental and its first few harmonics, which contain the majority of the acoustic energy. Naturally, for very complex silencer geometries, high frequencies or multidimensional phenomena (such as flow noise), the one-dimensional assumption is not justified and other methods are required.

The duct flows are simulated using a nonlinear time-domain approach for one-dimensional flows in ducts of variable cross section. The numerical technique is based on the one-dimensional finite-difference approach of Chapman, Novak, and Stein,[15–18] which has been modified to include flows through perforations. Since the numerical scheme computes the unsteady flows in the time domain, the nonlinear terms in the balance equations of mass, momentum, and energy are retained. Therefore, the approach is capable of modeling nonlinear effects that occur in automotive systems such as choked flow, shock waves, and heat transfer. Moreover, for an integrated intake/engine/silencer simulation, the spatial variations in bulk fluid velocity, mean pressure, and temperature, which affect the performance of both the engine and silencers, are included in the solution.

A. Duct flows

The numerical approach is based on a finite-difference treatment of unsteady compressible flow in ducts of variable cross section (which may or may not be perforated). For the one-dimensional approach, the balance equations for mass, momentum, and internal energy are applied to a control volume lumped over the duct cross section with axial (x-direction) conduction in the fluid neglected. Assuming that the flows through the perforate interface transfer a negligible amount of axial direction momentum, the balance equations may be written as

$$\frac{\partial}{\partial t}(\rho A)+\frac{\partial}{\partial x}(\rho(AU+A_pU_p))=0, \tag{1}$$

$$\frac{\partial}{\partial t}(\rho AU)+\frac{\partial}{\partial x}(\rho AU^2)+\frac{\partial}{\partial x}(pA)+\tau_w\mathcal{P}=0, \tag{2}$$

$$\frac{\partial}{\partial t}(\rho Ae)+\frac{\partial}{\partial x}(\rho e(AU+A_pU_p))+p\frac{\partial}{\partial x}(UA+U_pA_p) - \tau_w\mathcal{P}U+q_w\mathcal{P}=0, \tag{3}$$

where ρ is density, U is velocity, p is pressure, e is internal energy, A is cross-sectional area, τ_w is wall shear stress, $\mathcal{P}$ is perimeter, q_w is the wall heat flux, and the subscript p denotes the perforate. The ideal gas equation of state,

$$p=(\gamma-1)\rho e, \tag{4}$$

where γ is the ratio of specific heats, is used to relate the thermodynamic variables and close the system of equations.

Equations (1) through (3) are discretized using the explicit finite-difference method of Chapman, Novak, and Stein[15] that has been modified to incorporate flow through perforates. In the numerical approach, ducts are divided into computational cells by nodal points and the geometric and dynamic variables are defined on a staggered mesh. For the one-dimensional flow in the duct, vector quantities are located at node points, and scalar quantities are located at cell midpoints. Aggregate quantities for the orifices contained between two computational nodes are located at cell midpoints. Details of the numerical technique will not be elaborated here, as they may be found elsewhere.[15,17,19]

B. Lumped parameter perforate model

For the perforations that are used in automotive silencers, the perforate dimensions are usually much smaller than the wavelengths of interest. Also, the velocities encountered in and near the perforations are typically much less than the speed of sound. Under these conditions, the flow through the orifices may be treated as a locally incompressible flow that transfers mass and energy between two external regions. For

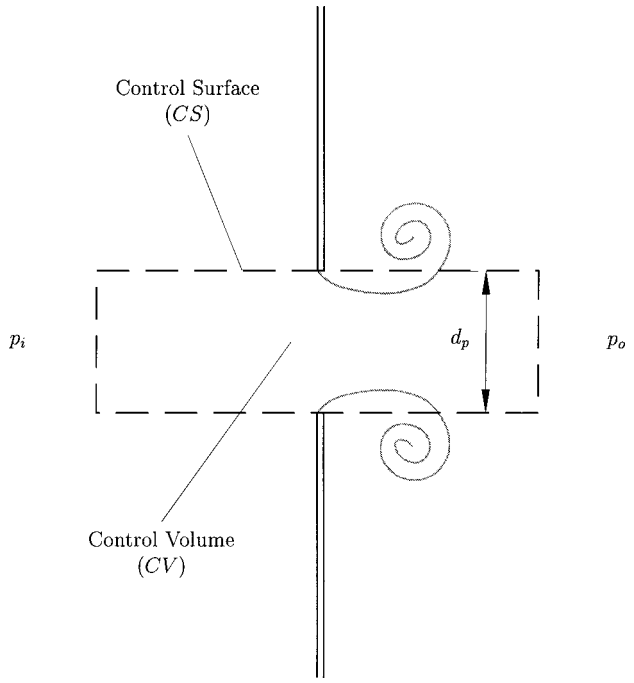


FIG. 1. Unsteady flow through a circular orifice-control volume for orifice model equation derivation.

the present case, the model problem considers oscillating flow through an isolated, circular, square-edged orifice in a flat plate of infinite extent as shown in Fig. 1. It is assumed that each orifice in a perforated plate or tube acts independently and there is no mean flow in the system.

An expression for modeling the perforate flows can be obtained by applying a momentum balance to the control volume depicted in Fig. 1. The form of the expression is simplified by introducing scaling coefficients for the inertia, momentum convection, and shear terms as (see Dickey *et al.*[20])

$$\underbrace{\Delta p}_{\text{pressure}} = \underbrace{\rho l_{\text{eq}} \frac{dU_p}{dt}}_{\text{inertia}} + \underbrace{C_M \rho U_p |U_p|}_{\text{momentum convection}} + \underbrace{C_\tau U_p}_{\text{shear}}, \tag{5}$$

where U_p is the (area averaged) perforate velocity. In Eq. (5), the coefficient l_{eq} is an equivalent length for the orifice, consistent with the common use of end corrections in acoustic analyses. The coefficient C_M relates the net rate of momentum convection to $\rho U_p|U_p|$. For the steady flow case C_M is equivalent to a flow loss coefficient $K=\Delta p/\rho U_p^2$ and may be related to the steady flow discharge coefficient as $C_M=1/2C_D^2$. Finally, C_τ is a coefficient relating the total x-direction shear force for the control volume to U_p.

Expressions with varying degrees of similarity to Eq. (5) have been obtained in analyses of Helmholtz resonators[21,22] and duct terminations.[23] In general, each of the coefficients in Eq. (5) is expected to be a time-varying function of the instantaneous pressure and velocity fields. Therefore, each of the terms on the right-hand side of Eq. (5) can contribute to the nonlinear orifice behavior. However, the primary source of nonlinearity is the convective momentum term which involves U_p^2. For small-amplitude disturbances, this term is negligible, corresponding to the linear regime of behavior. At very high amplitudes, the convective term dominates, corresponding to the quasisteady regime of behavior where $\Delta p \sim \rho U_p^2$.

When the orifice velocity is large enough for the convective term to be important, flow separation occurs and vorticity is generated. Since very little of the vortical energy is restored to the pressure field, the convective and viscous terms both represent energy losses that occur in the flow. It is therefore convenient to combine the convective and viscous terms into a single loss term as

$$C_\tau U_p + C_M \rho U_p |U_p| = R U_p, \tag{6}$$

where R is the orifice resistance. Incorporating the resistance into Eq. (5) gives the expression commonly used for perforate models as

$$\underbrace{\Delta p}_{\text{pressure}} = \underbrace{\rho l_{\text{eq}} \frac{dU_p}{dt}}_{\text{inertia}} + \underbrace{R U_p}_{\text{resistance}}, \tag{7}$$

which is also adopted in the present study. An attractive feature of Eq. (7) is that for constant values of l_{eq} and R, the expression is linear. For low-amplitude disturbances, the linear orifice behavior can be solved using the well-established methods of linear analysis. When the perforate behavior is nonlinear, the coefficients in Eq. (7) will depend on the instantaneous velocity and therefore will also vary in time. However, an approximate treatment that linearizes Eq. (7) is still possible by assuming that the coefficients are constant over a cycle (time invariant), but depend on the peak velocity or other cyclic parameters.

C. Perforate impedance and scaling parameters

In the present study, the base nonlinear perforate model assumes that the coefficients l_{eq} and R are time-invariant, corresponding to linearized treatment of the orifice model equation. For the special case of single-frequency excitation that is considered, this model is equivalent to the use of a nonlinear perforate impedance. The perforate impedance is obtained from the fundamental pressure and velocity components as

$$\mathbf{Z}_p = \frac{\Delta \mathbf{p}^{[1]}}{\mathbf{U}_p^{[1]}} = i\chi^{[1]} + R^{[1]}, \tag{8}$$

where the imaginary part of the impedance ($\chi^{[1]}$) is called the reactance and the real part of the impedance ($R^{[1]}$) is called the resistance. In Eq. (8), boldface symbols denote complex quantities and the bracketed superscript 1 designates the fundamental harmonic component. A nonbold symbol of a complex harmonic variable ($U_p^{[1]}$, for example) will be used to denote the absolute magnitude of a complex quantity. Inserting the fundamental components of pressure and velocity into Eq. (7) and dividing by $\mathbf{U}_p^{[1]}$ yields

$$\mathbf{Z}_p = \frac{\Delta \mathbf{p}^{[1]}}{\mathbf{U}_p^{[1]}} \cong i\omega\rho l_{\mathrm{eq}}^{[1]} + R^{[1]}, \tag{9}$$

where the $\cong$ sign has been used to distinguish that harmonics other than the (dominant) fundamental have been neglected. Inspection of Eqs. (8) and (9) shows that $R^{[1]}$ is simply the real part of the impedance as shown above and $l_{\mathrm{eq}}^{[1]}$ is related to the imaginary part of the impedance (reactance) as

$$l_{\mathrm{eq}}^{[1]} = \frac{\mathfrak{T}(\mathbf{Z}_p)}{\rho\omega} = \frac{\chi^{[1]}}{\rho\omega}. \tag{10}$$

Utilizing the fact that the spectra are dominated by the fundamental frequency f and introducing some additional information allows the nondimensional scaling variables described by Panton and Goldman[24] to be formed from the ratios of the terms in Eq. (5). Noting that for acoustic disturbances and thin orifices $l_{\mathrm{eq}} \sim d_p$ and $\tau_w \sim \mu U_p^{[1]}/\sqrt{\nu/\omega}$, where d_p is the orifice diameter, $\nu = \mu/\rho$ is the kinematic viscosity and $\omega = 2\pi f$ is the angular frequency, the ratios of terms in the momentum balance may be expressed as

$$\frac{\text{momentum convection}}{\text{shear}} \sim \frac{U_p^{[1]}}{\sqrt{\nu\omega}} \sim \frac{L_e}{\delta_{BL}}, \tag{11}$$

$$\frac{\text{momentum convection}}{\text{inertia}} \sim \frac{U_p^{[1]}}{\omega d_p} \sim \frac{L_e}{d_p}, \tag{12}$$

$$\frac{\text{inertia}}{\text{shear}} \sim d_p \sqrt{\frac{\omega}{\nu}} \sim \frac{d_p}{\delta_{BL}}. \tag{13}$$

The foregoing equations also include interpretations in terms of ratios of length scales for the flow, as discussed in Ref. 24. The variable L_e represents an acoustic displacement amplitude and $\delta_{BL} = \sqrt{\nu/\omega}$ is proportional to the acoustic boundary layer thickness.

Since the momentum convection term is the primary nonlinear term, the nondimensional ratios provide parameters to help differentiate the linear and nonlinear regimes of behavior. The linear regime occurs when the convective term is much smaller than one of the other two terms. Therefore, predominantly linear behavior will occur for either

$$\text{linear regime:} \quad \frac{U_p^{[1]}}{\sqrt{\nu\omega}} \ll 1 \quad \text{or} \quad \frac{U_p^{[1]}}{\omega d_p} \ll 1. \tag{14}$$

In the highly nonlinear regime, the convective term is much larger than both of the other terms. Therefore, the highly nonlinear regime exists when

$$\text{highly nonlinear regime:} \quad \frac{U_p^{[1]}}{\sqrt{\nu\omega}} \gg 1 \quad \text{and} \quad \frac{U_p^{[1]}}{\omega d_p} \gg 1. \tag{15}$$

In the present study, the highly nonlinear regime is also referred to as the quasisteady regime. While a viscous-dominated flow could also be considered quasisteady, this condition is unlikely to be of practical significance. For orifice dimensions and fluid properties typical of automotive silencers, the inertia/shear ratio is generally rather large (for example, with $d_p = 0.25$ cm and $f = 100$ Hz, $d_p/\delta_{BL} \cong 16$ for air at atmospheric conditions). Therefore, the quasisteady condition of interest is the highly nonlinear case where $\Delta p \sim \rho U_p^2$, and the primary indicator of orifice nonlinearity is the convection/inertia ratio $U_p^{[1]}/\omega d_p$.

II. EXPERIMENTAL METHODS

Two different types of experiments were performed in the study. The first type considered the acoustic performance of an entire perforated tube silencer while the second type focused specifically on the behavior of unsteady flow through orifices. The silencer experiments were conducted to provide experimental data for evaluating the accuracy of the time-domain computational approach. The orifice experiments were undertaken to investigate nonlinear perforate behavior and develop empirical expressions for the lumped parameter impedance coefficients.

The silencer measurements were conducted in an extended impedance tube system. In order to perform the orifice experiments in parallel with the silencer studies and use existing experimental equipment, the experimental methods and apparatus for the orifice study were designed for use in the existing impedance tube setup. In the following sections, the impedance tube facility and silencer experiments will be discussed first, followed by descriptions of the test apparatus and methods used in the orifice experiments.

A. Impedance tube

The acoustic performance of the silencers was measured using an extended impedance tube system (see Fig. 2). The overall setup includes the impedance tube itself and a data acquisition and analysis system. The impedance tube uses a 25.4-cm loudspeaker (Madisound 10208) for the acoustic source. The input signal for the loudspeaker is generated by the system signal generator module (B&K 3107) and then amplified by a 100-watt amplifier (Kenwood KR-A5070). A conical section is used for the transition in diameter between the loudspeaker and the inlet duct. The element to be tested is mounted between inlet and exit ducts that have inner diameter of 4.86 cm. Effects caused by reflections downstream of the silencer are avoided by using a nonreflective (anechoic) termination.

The upstream and downstream ducts in the impedance tube each contain a pair of 0.635-cm condenser microphones (B&K 4135) to measure the acoustic pressure in the duct. Signals from the four microphones are acquired and processed by a modular B&K 3550 multichannel analysis system. The system acquires data from the microphone signals and computes the autospectra at each microphone location, and cross spectra between the microphone locations. User-defined functions (user programs) are then employed to manipulate the auto- and cross spectra to the final desired output.

The acoustic performance of the silencers in this study is presented in terms of the transmission loss. For planar wave propagation with inlet and exit ducts of the same diameter, the transmission loss, in decibels, is defined as

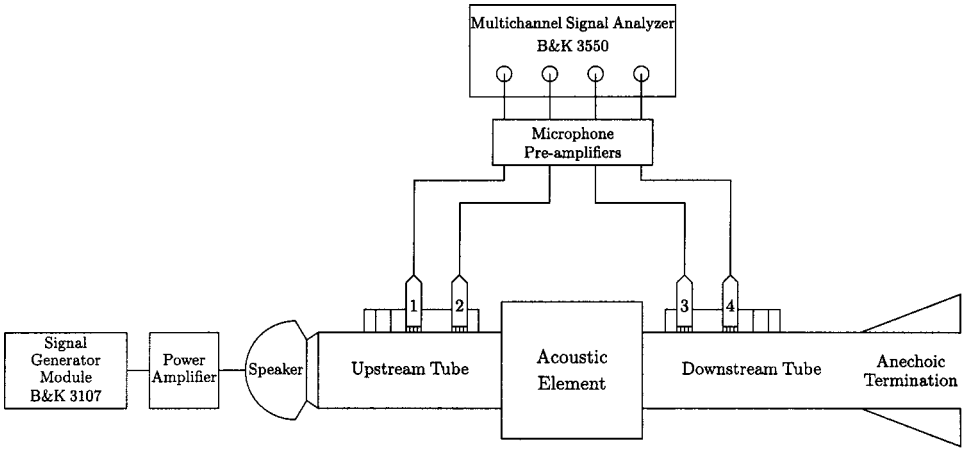


FIG. 2. Impedance tube system.

$$TL = 20 \log_{10} \left| \frac{\mathbf{p}_{\text{inc}}^{[f]}}{\mathbf{p}_{\text{tr}}^{[f]}} \right|, \tag{16}$$

where the subscripts inc and tr denote incident and transmitted components, respectively, and the bracketed superscript denotes an harmonic component at frequency f. The two-microphone method for planar wave separation is used to separate incident and reflected waves in the ducts and compute transmission loss from the acquired spectral data.[25,26] Further details regarding the impedance tube system and procedures used to obtain transmission loss may be found elsewhere.[17,19]

B. Orifice measurements

An acoustic phase-pressure approach, commonly referred to as the two-microphone technique,[4,7,27,28] was adopted for the orifice measurements. In the two-microphone technique, the perforated sample is backed by a cavity of known dimensions and pressure measurements are made at the sample face and within the cavity. Using linear acoustic theory, the pressure at the back of the sample and the flow rate through the perforations are determined from the cavity pressure measurement.

The experimental apparatus used for the orifice measurements is depicted in Fig. 3. An annular cavity surrounds a removable central tube containing a small section of orifices. Although the annular configuration is convenient for use in the impedance tube, the fixed axisymmetric geometry does introduce some limitations as noted by Sullivan.[28] First, the acoustics of an annular cavity is somewhat more complex than the more commonly used sidebranch duct of constant cross section. Second, and more significant, the fixed geometry of the system causes the acoustic response of the cavity at a given frequency to be fixed. Therefore, the cavity can not be ''tuned'' to take maximum advantage of cavity resonances. With the fixed cavity geometry and loudspeaker driver of the current setup, power and distortion limitations were encountered that reduced speaker life and limited the range of measurements.

The test section cavity is fabricated from two aluminum end plates and a short section of acrylic tubing (see Fig. 3). The end plates slide over the central tube and are attached directly to the acrylic tubing. For this study, the circular orifices were fabricated by drilling directly into a central tube having 4.92 cm inner diameter (i.d.) and 5.08 cm outer diameter (o.d.). The annular cavity has inner and outer wall diameters of 5.08 and 10.16 cm, respectively, and an axial length of 3.81 cm. Pressures at the inner surface of the central tube and at the outer cavity wall are measured with flush-mounted piezoresistive pressure transducers (Endevco 8507C-2). The pressure transducers are 0.234 cm in diameter, have a nominal range of 14 kPa, and sensitivity of 21.5 mV/kPa. In the current system, the sensitivity of the trans-

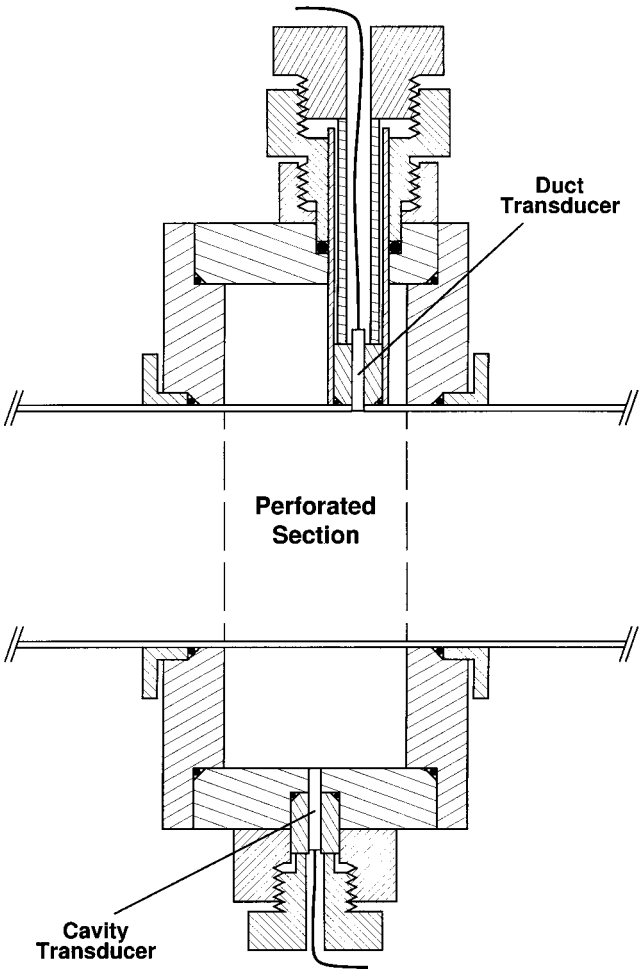


FIG. 3. Schematic of orifice measurement apparatus.

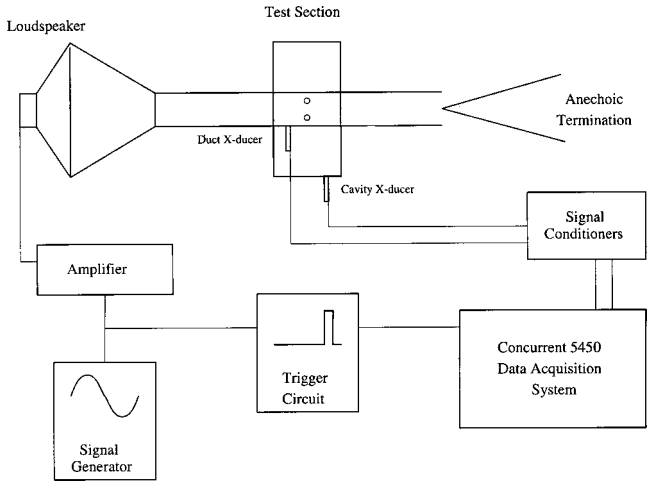


FIG. 4. Data acquisition setup for orifice measurements.

ducers limited the minimum measurable SPL to approximately 100 dB. The maximum rated SPL for the transducers is published as 190 dB, which is well beyond the capability of the driver.

Figure 4 depicts the data acquisition setup for the orifice measurements. The excitation voltage input, and output signal amplification for the pressure transducers, are provided by an Endevco model 106 bridge signal conditioner. The signals from the amplifiers are sampled and stored by a Concurrent 5450 high-speed data acquisition system. In order to reduce the effects of cycle-to-cycle variations, data are taken for a number of cycles and averaged. The phase-locked averaging process is made possible by a trigger circuit that initiates data acquisition at the same phase point for each cycle that is sampled. More detailed descriptions of the perforate experimental apparatus and methods are available in Refs. 19 and 20.

To infer the pressure at the back of the perforate sample and the perforate flow rate from the measured cavity pressure, an harmonic analysis of the cavity is performed. In the analysis of the cavity, two primary assumptions are made: (1) The oscillations in the cavity obey the linearized equations of motion given by the wave equation, and the linearized Euler equation, and (2) the acoustic propagation in the cavity is primarily one-dimensional in the radial direction.

The impedance is determined from the fundamental frequency component as

$$\mathbf{Z}_p = \left(\frac{\mathbf{p}_i^{[1]} - \mathbf{p}_o^{[1]}}{\mathbf{U}_p^{[1]}} \right), \tag{17}$$

where $\mathbf{p}_i^{[1]}$ is available directly from the center tube measurement. The pressure at the outer surface of the perforate ($\mathbf{p}_o^{[1]}$) can be computed from the cavity wall pressure as[8,20]

$$\mathbf{p}_o^{[1]} = \mathbf{p}_2^{[1]} \left[\frac{\mathbf{H}_0^{(1)}(kr_1) - \dfrac{\mathbf{H}_1^{(1)}(kr_2)}{\mathbf{H}_1^{(2)}(kr_2)} \mathbf{H}_0^{(2)}(kr_1)}{\mathbf{H}_0^{(1)}(kr_2) - \dfrac{\mathbf{H}_1^{(1)}(kr_2)}{\mathbf{H}_1^{(2)}(kr_2)} \mathbf{H}_0^{(2)}(kr_2)} \right], \tag{18}$$

where subscripts 1 and 2 denote the inner and outer cavity surfaces, respectively, $\mathbf{H}_a^{(b)}$ is the Hankel function of order a and type b, $k = \omega / c_o$ is the wave number, and r is the radius. Conservation of volumetric flow between the cavity and the orifices allows U_p to be determined from the measured cavity pressure as

$$\mathbf{U}_p^{[1]} = -\mathbf{p}_2^{[1]} \left(\frac{i 2\pi r_1 l_c}{\rho_o c_o A_p} \right) \times \left[\frac{\mathbf{H}_1^{(1)}(kr_1) - \dfrac{\mathbf{H}_1^{(1)}(kr_2)}{\mathbf{H}_1^{(2)}(kr_2)} \mathbf{H}_1^{(2)}(kr_1)}{\mathbf{H}_0^{(1)}(kr_2) - \dfrac{\mathbf{H}_1^{(1)}(kr_2)}{\mathbf{H}_1^{(2)}(kr_2)} \mathbf{H}_0^{(2)}(kr_2)} \right]. \tag{19}$$

The frequency domain impedance parameters $l_{\text{eq}}^{[1]}$ and $R^{[1]}$ are then available from the imaginary and real parts of the impedance, respectively [see Eqs. (8) and (9)].

III. PERFORATE EXPERIMENTAL RESULTS AND MODELS

Impedance parameters were obtained using the methods described in Sec. II for a test section containing eight orifices with diameter (d_p) of 0.249 cm and wall thickness (t_w) of 0.081 cm. The relatively small number of orifices was selected to maintain a low resonance frequency for the apparatus, and minimize the effects of orifice interactions. Experiments were performed for frequencies between 100 and 800 Hz ($0.0046 \leqslant k d_p \leqslant 0.036$). In general, experiments were performed at each frequency for sound-pressure levels across the perforate of 110 to 135 dB in 5-dB increments. However, due to the combination of a fixed cavity geometry and the limitations of the loudspeaker source, the maximum obtainable SPL across the orifices varied somewhat. With increasing frequency, distortion was encountered as the sound-pressure level was increased and, at frequencies above 500 Hz, a maximum of 135 dB across the orifice was not reached. For all frequencies considered, at least 130 dB was reached across the perforate.

A. Linear regime

Figure 5 includes experimental data for the end correction ($\Delta l_{\text{eq}}^{[1]} = l_{\text{eq}}^{[1]} - t_w$) and resistance ($R^{[1]}$) obtained at low to moderate sound-pressure levels. The values for $\Delta l_{\text{eq}}^{[1]}$ are scaled with respect to the Rayleigh end correction ($0.85 d_p$) and are plotted against $U_p^{[1]} / \omega d_p$. The values for $R^{[1]}$ are scaled with respect to $\rho \sqrt{\nu \omega}$ and are plotted against $U_p^{[1]} / \sqrt{\nu \omega}$. The experimental data indicate that for this perforate, the linear regime of behavior occurs for $U_p^{[1]} / \omega d_p$ less than approximately 0.2. Up to this point, $l_{\text{eq}}^{[1]}$ is essentially constant and, since the reactance term is dominant, the impedance is independent of velocity. For the higher amplitude measurements the well-known behavior of decreasing $\Delta l_{\text{eq}}^{[1]}$ and increasing $R^{[1]}$ are observed.

The data for $\Delta l_{\text{eq}}^{[1]} / 0.85 d_p$ do not show any significant dependence on frequency in the linear regime and the end correction is very close to the Rayleigh correction

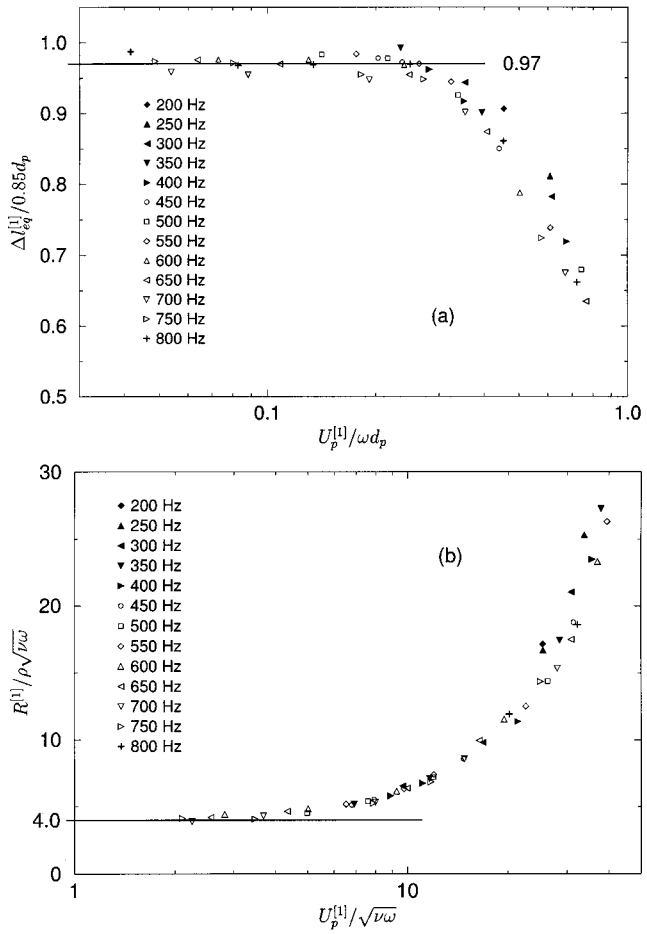


FIG. 5. Scaled $\Delta l_{\mathrm{eq}}^{[1]}$ and $R^{[1]}$ experimental data for the low-amplitude region.

$(\Delta l_{\mathrm{eq}}^{[1]}/0.85d_p \cong 0.97)$. The values of $R^{[1]}/\rho\sqrt{\nu\omega}$ collapse well at low $U_p^{[1]}/\sqrt{\nu\omega}$ and appear to be approaching a limit of approximately 4, although there is not enough data at low $U_p^{[1]}/\sqrt{\nu\omega}$ to be certain of this value. This value is quite close to the result from the often-used empirical expression[24,29]

$$R^{[1]}/\rho\sqrt{\nu\omega} \cong \sqrt{8}(1+t_w/d_p), \tag{20}$$

which, for this orifice geometry, yields $R^{[1]}/\rho\sqrt{\nu\omega} \cong 3.75$.

As theory for oscillating laminar flow indicates, the linear resistance has moderate dependency on frequency (increasing with $\sqrt{\omega}$). However, since $l_{\mathrm{eq}}^{[1]}$ is independent of frequency and the inertia/viscous ratio is large, small inaccuracies in the determination of $R^{[1]}$ are of minor consequence. By assuming an average value for $R^{[1]}$ that neglects this frequency dependence, a number of frequency components can be included in a single simulation to reduce computation time. The coefficients for linear perforate behavior are therefore specified to be invariant with respect to both time and frequency as

$$\frac{\Delta l_{\mathrm{eq}}^{[1]}}{0.85d_p} = 0.97, \tag{21}$$

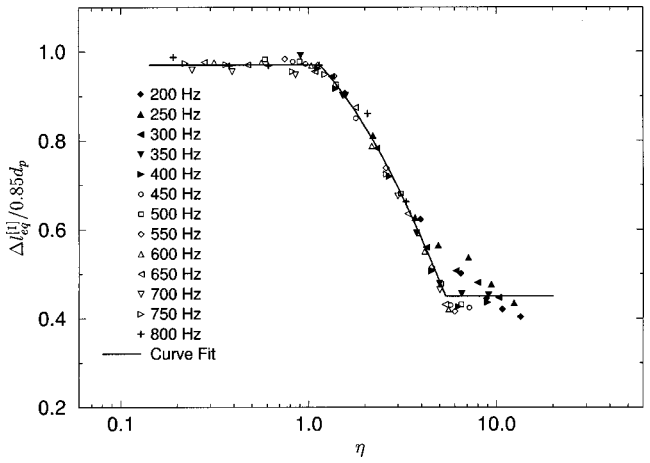


FIG. 6. Scaled $\Delta l_{\mathrm{eq}}^{[1]}$ experimental data for the low- and high-amplitude regions plotted against η.

$$\frac{R^{[1]}}{\rho\sqrt{\nu\omega_{\mathrm{mid}}}} = 4.0, \tag{22}$$

where ω_{mid} is a median value for the frequency range considered in the simulation. For the experimental data obtained, the $l_{\mathrm{eq}}^{[1]}$ values are within 3% of the approximation. The discrepancy between the assumed value of $R^{[1]}$ and measurement will vary with frequency but, once again, is of secondary importance.

B. Nonlinear regime

The nonlinear perforate treatment accounts for spatial variation in the perforate behavior, but assumes that the coefficients are constant in time. This model is formulated on the assumption that Δp and U_p are dominated by a single-frequency component that is known beforehand. Under these conditions, it is reasonable to use empirical relationships for $l_{\mathrm{eq}}^{[1]}$ and $R^{[1]}$ based on the scaling parameters discussed in Sec. I C. For the present work, the empirical expressions for $l_{\mathrm{eq}}^{[1]}$ are developed from experimental data obtained between 200 and 800 Hz. The expressions for $R^{[1]}$ are formulated from experimental data between 100 and 800 Hz. The nonlinear curve fits include the limiting cases of linear and quasisteady orifice behavior, as well as the intermediate region between these limits. The correlation parameters and empirical curve fits are described next for the equivalent length and resistance, respectively.

1. Equivalent length ($l_{eq}^{[1]}$)

Figure 6 includes experimental measurements of the time-invariant end correction ($\Delta l_{\mathrm{eq}}^{[1]}$) for frequencies between 200 and 800 Hz and varied incident sound-pressure levels. Inspection of Fig. 5 indicates that the end correction could be correlated reasonably with respect to $U_p^{[1]}/\omega d_p$ alone. However, the experimental data were found to collapse more satisfactorily with respect to η, a parameter given by

$$\eta = \left(\frac{U_p^{[1]}}{\omega d_p}\right)\left(\frac{d_p}{\delta_{BL}}\right)^{0.4}. \tag{23}$$

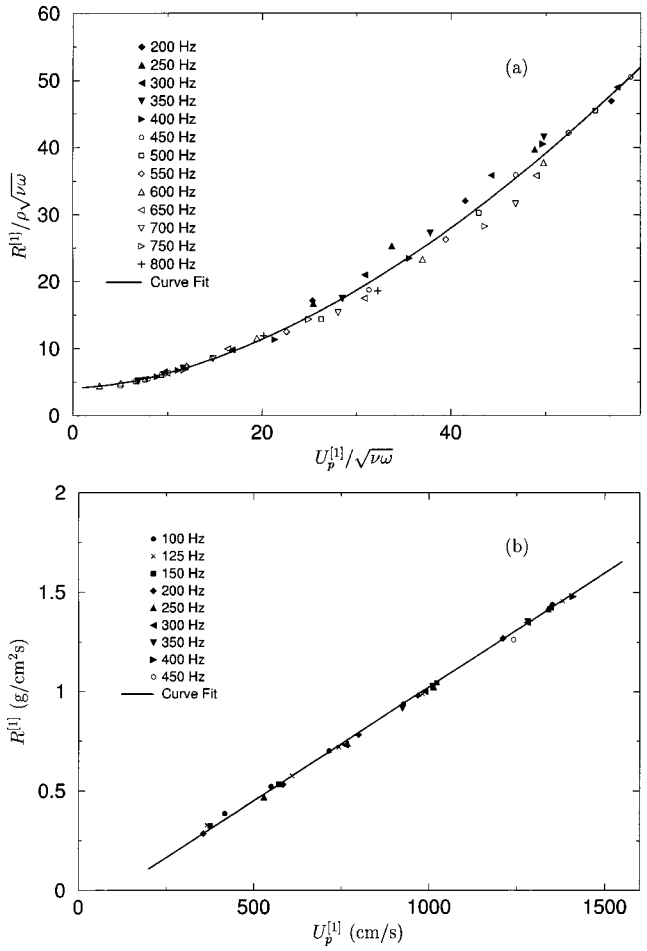


FIG. 7. $R^{[1]}$ experimental data for the (a) linear and transition and (b) highly nonlinear regions.

Figure 6 depicts the experimental data for $\Delta l_{\mathrm{eq}}^{[1]}$ scaled in this manner for the frequency range 200–800 Hz. The data collapse relatively well for the linear and moderately nonlinear ranges. A certain amount of scatter is evident for the data in the nonlinear range where $\Delta l_{\mathrm{eq}}^{[1]}/0.85d_p$ is less than 0.6. In this range, however, small errors in the value of $l_{\mathrm{eq}}^{[1]}$ are of less consequence due to the increased importance of the resistance.

A piecewise continuous curve is used to represent the data shown in Fig. 6. Separate expressions are used to represent the linear, nonlinear transition, and highly nonlinear regimes as follows:

$$\frac{\Delta l_{\mathrm{eq}}^{[1]}}{0.85d_p}=\begin{cases}0.97 & 0\leqslant\eta\leqslant 1.15\\ 1.23e^{-0.18\eta} & 1.15<\eta\leqslant 5.37.\\ 0.46 & 5.37\leqslant\eta\end{cases} \tag{24}$$

The lower limiting value of $\Delta l_{\mathrm{eq}}^{[1]}/0.85d_p=0.46$ is obtained from a potential flow analogy for the flow on the upstream side of the orifice. This value may be slightly higher than the data indicate, which is probably due to separated flow within the orifice itself.

2. Resistance ($R^{[1]}$)

Figure 7 includes measured values of $R^{[1]}$ for frequencies between 100 and 800 Hz. The data for nonlinear orifice resistance are correlated using two curve fit expressions. The first expression represents the linear and transition regimes, while the second expression represents the data in the highly nonlinear regime.

Figure 7(a) includes experimental data for the linear and transition regimes. The scaled parameters are presented as $R^{[1]}/\rho\sqrt{\nu\omega}$ vs $U_p^{[1]}/\sqrt{\nu\omega}$. Consistent with the findings of Panton and Goldman,[24] an acceptable collapse of the resistance data is obtained with these parameters. The data for this regime are approximated by the expression

$$\frac{R^{[1]}}{\rho\sqrt{\nu\omega}}=4.17+0.0421\left(\frac{U_p^{[1]}}{\sqrt{\nu\omega}}\right)^{1.72}. \tag{25}$$

Equation (25) is included in Fig. 7(a). Nearly all of the points are within 10% of this curve, with the maximum deviation for all points being less than 15%.

Figure 7(b) includes experimental data for the highly nonlinear regime, where $R^{[1]}$ is plotted against $U_p^{[1]}$ and the frequency range is 100–450 Hz. The resistance increases linearly with the orifice velocity and is represented by

$$R^{[1]}=-0.121+0.966\rho U_p^{[1]}, \tag{26}$$

where cgs units have been used. Similar to the data of a number of workers, the *y*-intercept falls below zero, suggesting a dependence for the ''average'' discharge coefficient on factors such as *vena contracta* formation or viscous boundary layers. Due to the speaker limitations, the data in the highly nonlinear regime are biased toward the lower frequency experimental observations. To confirm that the curve fit remained reasonable at higher frequencies, the high-amplitude data of Sullivan[8] were compared to Eq. (26). For high-amplitude measurements between 1000 and 1300 Hz, the data of Sullivan were found to be within 5% of Eq. (26).

In the computational model, the distinction between the intermediate and highly nonlinear regimes for the resistance is determined by the intersection of Eqs. (25) and (26). For the frequency range above approximately 100 Hz, these expressions will intersect at two points, the first one occurring for $U_p^{[1]}/\sqrt{\nu\omega}$ below 35 and the second representing the transition point of interest. Therefore, if $U_p^{[1]}/\sqrt{\nu\omega}\leqslant 35$, the resistance in the computational model is determined from Eq. (25). Otherwise, the resistance is determined as the minimum value obtained from Eqs. (25) and (26). At frequencies below 100 Hz, there will be a small discontinuity in $R^{[1]}$ at the transition between the two expressions. However, for these low frequencies and amplitudes, silencer behavior will be predominantly reactive and the effect of this discontinuity is expected to be inconsequential.

IV. SILENCER RESULTS

This section presents the experimental and computational results for perforated tube silencer geometries consisting of both single- and multiple-pass configurations. Following a description of the different silencer geometries considered, silencer performance is studied using the computational approach with the linear and nonlinear perforate models. For the linear simulations, comparisons are made

with experimental data from the extended impedance tube setup. Since the loudspeaker driver of the impedance tube setup was not capable of driving silencers into the nonlinear regime, transmission loss comparisons for a small concentric tube resonator are made with a data set from the published work of Sullivan.[13,28]

In order to calculate the scaling parameters and use the nonlinear impedance curve fits in the computational model, a value for $U_p^{[1]}$ must be obtained. For single-frequency simulations with small wave steepening effects, $U_p^{[1]}$ is available as the peak orifice velocity during an oscillatory cycle. Monitoring the peak orifice velocity at each computational cell thus allows the scaling parameters $U_p^{[1]}/\omega d_p$ and $U_p^{[1]}/\sqrt{\nu\omega}$ to be obtained. The additional scaling parameters are available from the input frequency, fluid properties, and orifice geometry. The time-invariant model is applied by first initiating the model with the equivalent length and resistance set at linear values. As the computations proceed, the peak orifice velocities at each cell are monitored. When the orifice flow for a given computational cell changes direction, the value of $U_p^{[1]}$ for that cell is updated as the peak velocity obtained during the previous half cycle. The values for $l_{\text{eq}}^{[1]}$ and $R^{[1]}$ are updated at each flow changeover until a steady state is reached.

A. Geometries

The silencer geometries considered in the study consist of three single-pass and two multiple-pass configurations. The perforated tubes and orifices in the silencers are consistent with that of the orifice test section with tube o.d.=5.08 cm, $t_w=0.081$ cm, and $d_p=0.249$ cm. Also, to minimize the effects of orifice interactions, the maximum porosity (σ) considered is 5%. The single-pass silencers include both ''short'' and ''long'' concentric tube geometries where, consistent with the work of Sullivan and Crocker,[8] a concentric tube silencer is designated as short if the volume-controlled (Helmholtz) resonance occurs before the first axial mode frequency ($f=c_0/2l_e$), and is otherwise designated as long. The multiple-pass geometries are variations of a prototype three-pass silencer with dimensions representative of a production automotive muffler.

Figure 8 includes schematics of the basic geometries of the single-pass and multi-pass perforated tube silencers. For convenience, all single-pass silencers will be referred to as resonators. Two short and one long resonator geometries were fabricated having the dimensions shown in Table I. All silencers were perforated evenly over the central portion of the cavity and the porosities are computed with respect to the outer diameter of the central tube. Resonator 1 has dimensions and porosity consistent with the silencer used in the nonlinear experiments of Sullivan.[28] This resonator was fabricated to obtain data to supplement the comparisons to Sullivan's nonlinear experiments. Resonators 2 and 3 were fabricated specifically for comparing model predictions with the impedance tube measurements.

The overall structure of the prototype multi-pass muffler consists of an external tube, four baffle plates, and the perforated tubes. Each of the baffle plates and tubes is removable, which allows different silencer configurations to be assembled from the same components. Two three-pass geometries are considered in the study, which will be referred to as mufflers. Dimensions for each geometry are included in Table I. In muffler 1, all tubes spanning the central chamber are perforated evenly and have the same porosity. Muffler 2 is identical to muffler 1 except that the pass tube connecting the two end chambers is replaced by a nonperforated tube.

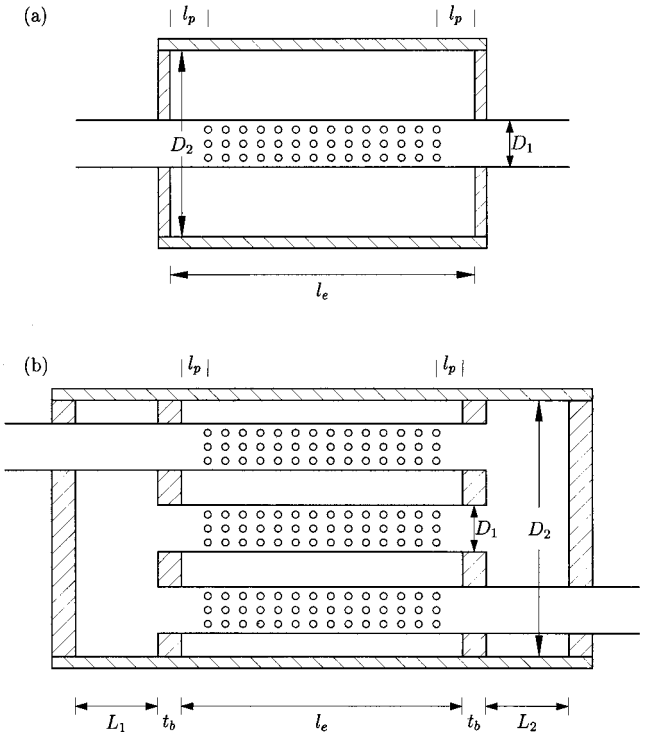


FIG. 8. Schematic drawings of (a) single-pass resonator, and (b) prototype muffler geometries.

For the muffler simulations, the end chambers are treated as lumped volumes. At the expansion from the central tubes to the end chambers, an end correction is necessary in the model to account for the local multidimensional effects.[30] For simplicity, the end correction is taken from the work of Ingard[29] for concentric Helmholtz resonators as

$$\Delta l_{\text{end chamber}}=0.425d_{\text{pipe}}\left(1-1.25\frac{d_{\text{pipe}}}{d_{\text{chamber}}}\right). \tag{27}$$

B. Computational and experimental results

To develop confidence in the nonlinear perforate model, the computational approach is first applied to resonator 1 so the results can be compared to the high-amplitude measure-

TABLE I. Dimensions of the fabricated silencers (lengths in cm).

Silencer	D_1	D_2	l_e	l_p	L_1	L_2	t_b	t_w	σ
Resonator 1	5.08	10.16	6.67	0.36	N/A	N/A	N/A	0.081	0.037
Resonator 2	5.08	7.27	6.67	0.34	N/A	N/A	N/A	0.081	0.037
Resonator 3	5.08	10.16	25.72	0.36	N/A	N/A	N/A	0.081	0.020
Muffler 1	5.08	16.51	25.48	0.55	4.60	4.29	1.27	0.081	0.048
Muffler 2	5.08	16.51	25.48	0.55	4.60	4.29	1.27	0.081	0.048[a]

[a]Pass tube not perforated.

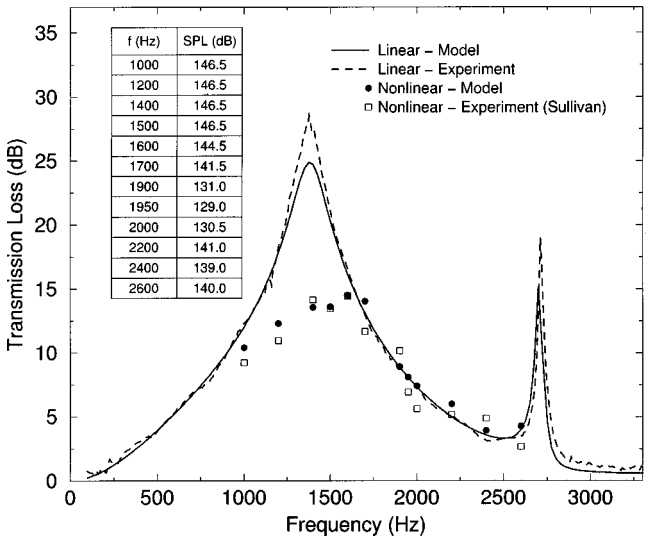


FIG. 9. Predicted and measured transmission loss for resonator 1 at low and high sound-pressure levels (σ=0.037).

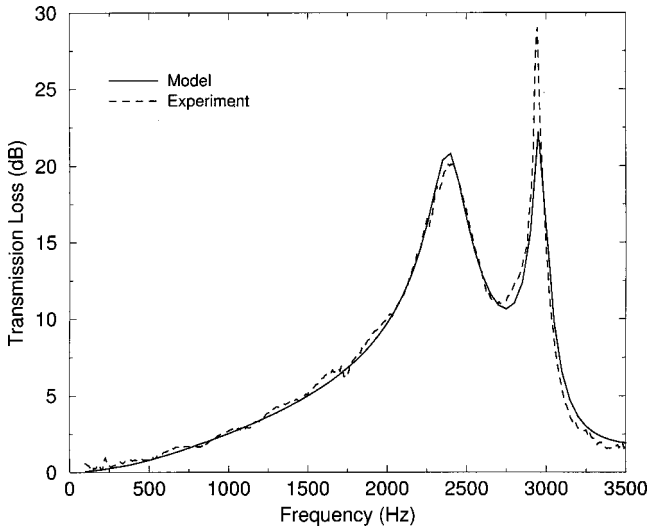


FIG. 10. Predicted and measured transmission loss for resonator 2 at low sound-pressure levels (σ=0.037).

ments of Sullivan.[13] For the remaining silencers, computations using the linear perforate model are first compared to experimental measurements with the extended impedance tube. The effects of high-amplitude oscillations are then investigated computationally using the nonlinear impedance perforate model and pure tone incident sound-pressure levels (SPL_i) between 130 and 160 dB.

For incident sound-pressure levels up to 160 dB, questions may arise regarding the effects of nonlinear wave steepening on the transmission loss. An estimate of the importance of wave steepening may be obtained from the Fubini–Ghiron solution for nonlinear wave propagation.[31] This solution applies for one-dimensional inviscid propagation of pressure disturbances much smaller than ambient in a constant area duct. The Fubini–Ghiron solution indicates that even without any wave attenuation by a silencer, the amount of reduction in the fundamental pressure component will correlate to a change in transmission loss of less than 1 dB for the geometries and conditions considered. In addition, the dissipative effects of heat transfer and viscosity for acoustic propagation in the main ducts are expected to be small (see, for example, Ref. 32), particularly in comparison to the dissipation at the perforations.

1. Resonator 1

Predictions and measurements for the transmission loss of resonator 1 at both low (linear perforate behavior) and high (nonlinear perforate behavior) sound-pressure levels are included in Fig. 9. The low-amplitude measurements were obtained as part of the present study while the high-amplitude measurements are taken from the published work of Sullivan.[13] Due to limitations in the driver (Ling 94B) used by Sullivan, the incident sound-pressure level varies between the measurements at different frequencies. The incident sound-pressure level for each measurement, rounded to the nearest 0.5 dB, is included in the inset table of Fig. 9.

The linear behavior of the silencer exhibits the typical behavior of a short concentric tube resonator. The first portion of the transmission loss includes a rather broad peak due to a volume-controlled (Helmholtz) resonance. At higher frequencies, a transition in behavior occurs that results in length-controlled resonances characterized by narrow spikes in the transmission loss curve. For the high-amplitude cases, reductions in transmission loss of 10 dB or more near resonance clearly demonstrate the adverse effect that nonlinear perforate behavior can have on the attenuation characteristics of a silencer. The increase in orifice impedance at high sound-pressure levels inhibits the cavity/duct communication, and therefore reduces the amount of attenuation.

The linear computations and experiments in Fig. 9 show good agreement for the entire frequency range considered. A moderate underprediction of the transmission loss at the volume-controlled resonance occurs, along with a slight underprediction of the length-controlled resonance frequency near 2700 Hz. A comparison between Sullivan's data and the computational predictions shows that the model captures the general behavior due to high-amplitude disturbances. Discrepancies between the model and high-amplitude experimental data are less than 2 dB for all points except the measurement at 1700 Hz, where the difference is approximately 2.3 dB. The average discrepancy for all 12 measurements is 1.1 dB. For the present case, this is considered good agreement since the measured data have experimental uncertainty of 1–2 dB.[28] An indication of the experimental uncertainty is provided by the measured data between 1900 and 2000 Hz. These experiments were performed well away from resonance with incident sound-pressure levels near 130 dB, suggesting essentially linear silencer behavior. Model predictions for these cases are consistent with the linear results, while the experimental data show deviations of $\pm$1–2 dB.

2. Resonator 2

Comparisons of predicted and measured transmission loss of resonator 2 for low sound-pressure levels and linear perforate behavior are shown in Fig. 10. The behavior for this short resonator is quite similar to resonator 1. For resonator 2, however, the volume- and length-controlled reso-

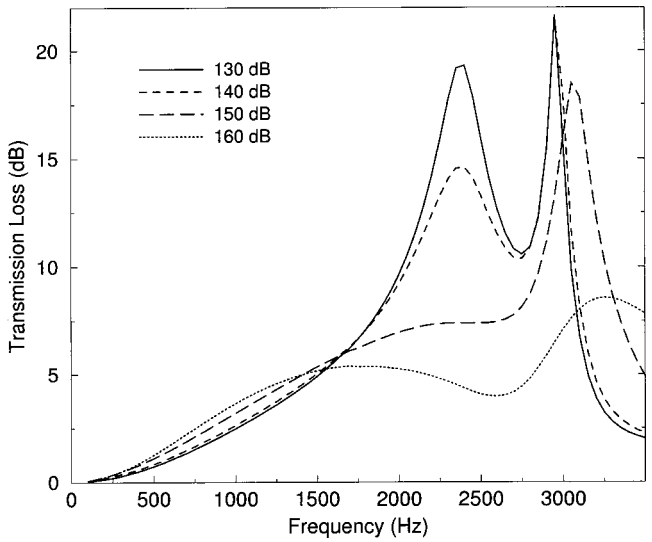


FIG. 11. Predicted transmission loss for resonator 2 with varied incident sound-pressure level ($\sigma=0.037$).

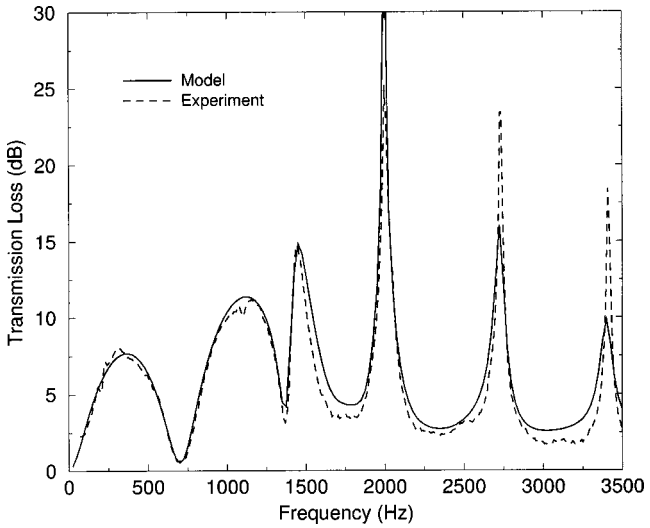


FIG. 12. Predicted and measured transmission loss for resonator 3 at low sound-pressure levels ($\sigma=0.020$).

nances are closer to each other, and their interactions tend to increase the transmission loss at frequencies in between. Agreement between the linear computations and experiments is good. Other than perhaps a very slight underestimation of the second resonance frequency, both of the peaks are predicted well by the computational model. The underprediction of the transmission loss at the length-controlled resonance is largely due to the frequency resolution of the simulation (25 Hz), which is too coarse to fully resolve a narrow spike. The higher frequencies may also be slightly affected by numerical dissipation.

Transmission loss predictions for resonator 2 with incident sound-pressure levels of 130–160 dB and the nonlinear impedance model are shown in Fig. 11. The results in Fig. 11 further demonstrate the dramatic effect that nonlinear perforate behavior can have on Helmholtz-type perforated tube resonators. At an incident sound-pressure level of 130 dB, the silencer behavior is essentially linear, which may be seen by a comparison to Fig. 10. As the incident sound-pressure level is increased, the effects of nonlinearity are first evident near the volume-controlled resonance, where the orifice velocities are highest for a particular sound-pressure level. At $\mathrm{SPL}_i=150$ dB, the transmission loss peak for the volume-controlled resonance is almost entirely lost, and changes in the length-controlled resonance are more evident. Noticeable, but relatively insignificant increases in transmission loss at lower frequencies are due to increased energy dissipation at the orifices. Finally, at $\mathrm{SPL}_i=160$ dB, both the volume- and length-controlled resonances are diminished substantially. The transmission loss is broadband in nature and remains below 6 dB for much of the frequency range.

3. Resonator 3

Comparisons of predicted and measured transmission loss of resonator 3 for low sound-pressure levels and linear perforate behavior are shown in Fig. 12. For this long resonator, the transmission loss curve at lower frequencies includes two broadband domes with the second dome being larger in magnitude. These domes correspond to the (somewhat modified) behavior of a simple expansion chamber which, for the geometry of resonator 3 with no perforated tube, would give 7.0-dB domes that repeat at 670-Hz intervals. Immediately after the second dome a resonance occurs that causes a transmission loss spike. This spike is followed by three similar sharp resonances away from which the transmission loss drops rapidly to only a few dB. Agreement between the model and experiment for resonator 3 is good for both the broadband and resonance behaviors. Again, due to the frequency resolution of the simulations, and possibly some numerical dissipation, differences are observed in the peak transmission loss for the length-controlled resonances.

Transmission loss predictions for resonator 3 with incident sound-pressure levels of 130–160 dB and the nonlinear impedance model are shown in Fig. 13. The transmission loss changes substantially with varying SPL_i but, in contrast to resonators 1 and 2, the changes can have rather large positive effects. As expected, the 130-dB case is essentially

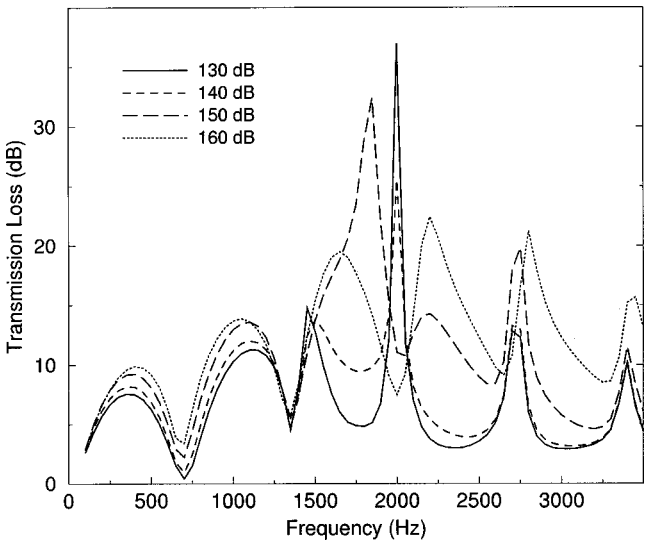


FIG. 13. Predicted transmission loss for resonator 3 with varied incident sound-pressure level ($\sigma=0.020$).

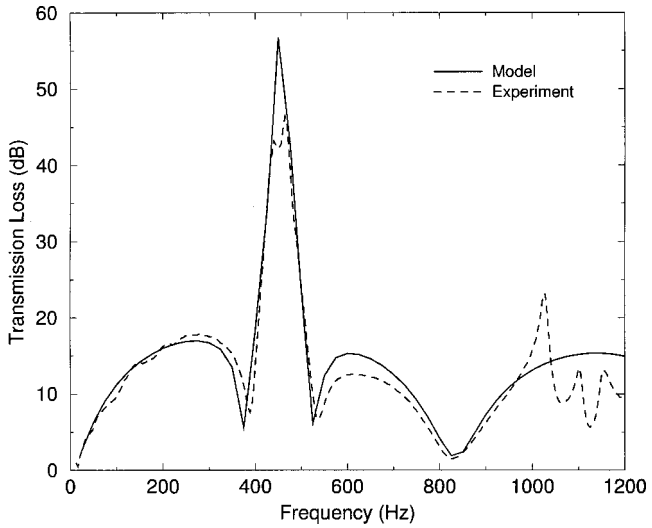


FIG. 14. Predicted and measured transmission loss for muffler 1 at low sound-pressure levels (σ=0.048).

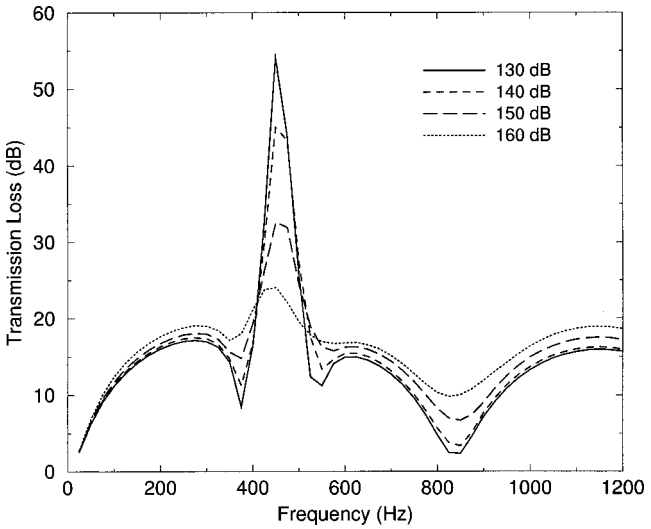


FIG. 15. Predicted transmission loss for muffler 1 with varied incident sound-pressure level (σ=0.048).

linear and very close to the results in Fig. 12. Below approximately 250 Hz, the behavior is relatively insensitive to SPL_i. Above this frequency, a number of changes occur at higher sound-pressure levels. Increasing SPL_i generally tends to increase the transmission loss in the troughs between the length-controlled resonances. At $SPL_i = 150$ dB, the frequency range between 1500 and 2000 Hz becomes a region of high attenuation and the trough between 2000 and 2750 Hz is largely eliminated. Finally, at $SPL_i = 160$ dB, the transmission loss minima occur near $nc_0/2l_e$ which, at lower sound-pressure levels, were length-controlled resonance locations. Above 2000 Hz, the behavior seems to be somewhat of a cross between the broadband and resonance behaviors. The dome shape is replaced by a rapid rise to a peak after the minima location, and then the transmission loss falls more gradually toward the next minima.

4. Muffler 1

Predictions of the transmission loss for muffler 1 with low sound-pressure levels and linear perforate behavior are compared to experimental data in Fig. 14. The linear regime transmission loss of muffler 1 is similar to an expansion chamber with a superimposed resonance near 450 Hz. The model accurately predicts the location of the resonance and then overestimates the transmission loss slightly for the part of the dome between 550 and 750 Hz. The location of the pass band near 830 Hz and the beginning of the following dome are well determined up to approximately 1000 Hz. Above this frequency, the model predicts a smooth domelike behavior whereas the experimental data deviate rapidly, and show multiple peaks and troughs. The discrepancies indicate that multidimensional effects are significant at these frequencies and the coupled one-dimensional approach is no longer valid (see, for example, Ji and Selamet[33]).

Figure 15 includes the computed transmission loss results for muffler 1 with varied incident sound-pressure levels and the nonlinear impedance model. As might be expected for the larger geometry and lower frequencies, the effects of nonlinear orifice behavior are somewhat less dramatic than for the concentric tube resonators. The most evident changes in transmission loss appear for the resonance. The magnitude of the transmission loss at resonance is diminished as the sound-pressure level increases, but the resonance frequency seems only slightly affected. The work of Munjal[34] indicates that mean flow in a multipass muffler can result in rather similar changes in the transmission loss behavior near resonance for small-amplitude disturbances. This is not surprising since the grazing flow will increase the perforate resistance and decrease the perforate reactance, effects similar to those of high-amplitude orifice velocities, although occurring by different mechanisms. Away from resonance, broadband transmission loss behavior improves by 2–3 dB at the higher sound-pressure levels, with more significant increases occurring near the passband location.

5. Muffler 2

Comparisons of predicted and measured transmission loss for muffler 2 with low sound-pressure levels and linear perforate behavior are shown in Fig. 16. The computational model captures the overall behavior of muffler 2 well but overpredicts the primary resonance frequency by approximately 20 Hz. A comparison of the transmission loss for muffler 2 with that of muffler 1 (Fig. 14) shows that each of the perforated tubes plays a significant role in the behavior of a multipass silencer. For muffler 2, the primary resonance occurs at approximately 280 Hz, much lower than the resonance frequency for muffler 1. It appears that multidimensional effects become significant at a slightly lower frequency for muffler 2, with predictions deviating from experiment above approximately 900 Hz.

Figure 17 includes the computed transmission loss results for muffler 2 with varied incident sound-pressure levels and the nonlinear impedance model. Similar to muffler 1, the magnitude of the transmission loss at resonance is diminished as the sound-pressure level increases. However, the location of the resonance frequency (which was only slightly affected for muffler 1), exhibits a substantial decrease. Away

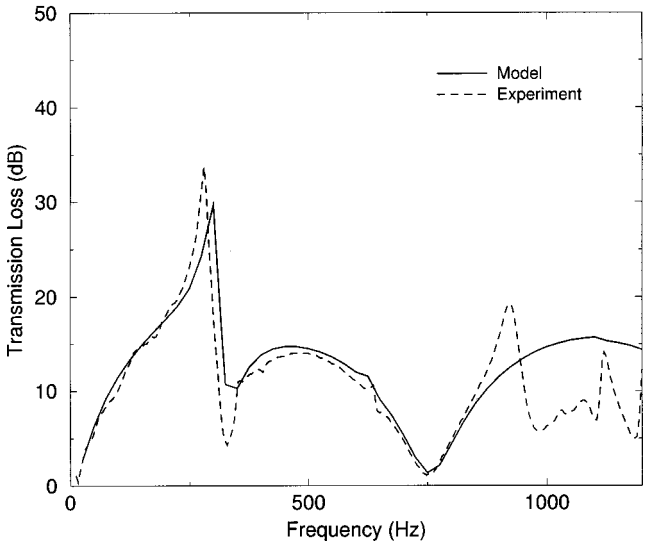


FIG. 16. Predicted and measured transmission loss for muffler 2 at low sound-pressure levels (σ=0.048).

from resonance, the effects of increased dissipation are again evident as a moderate increase in transmission loss at all but the lowest frequencies.

V. CONCLUDING REMARKS

The time-domain computational model has been used to predict the acoustic behavior of both single- and multiple-pass perforated tube silencers. Empirical expressions were developed from experimental data to relate the nonlinear behavior of $l_{\mathrm{eq}}^{[1]}$ and $R^{[1]}$ to relevant nondimensional parameters. Predictions have been performed for a range of incident sound-pressure levels between the low-amplitude limit and 160 dB. For automotive silencers, the effects of mean flow and multifrequency excitation, which have been excluded from the present study, require consideration. However, the basic findings of the work, as well as the computational approach itself, are expected to be useful in the design and analysis of these elements.

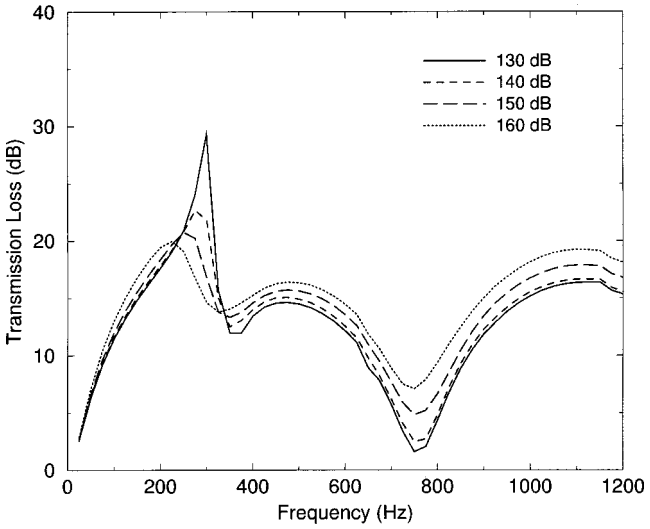


FIG. 17. Predicted transmission loss for muffler 2 with varied incident sound-pressure level (σ=0.048).

In the regime of linear perforate behavior, transmission loss predictions compare well with experimental data obtained with the extended impedance tube setup. For concentric tube resonators, the model accurately predicted the volume- and length-controlled resonances for both short and long geometries. The model also captured the transmission loss behavior of the multipass silencers with generally good agreement between the experiments and predictions.

A comparison to the experimental results of Sullivan[13] for a short concentric tube resonator (resonator 1) indicates that the nonlinear time-invariant perforate model predicts the effects of orifice nonlinearity reasonably. For the concentric tube resonators, high-amplitude disturbances tend to eliminate both volume- and length-controlled resonances and promote broadband behavior. This dramatically reduces the effectiveness of short geometries designed for volume-controlled resonance behavior, while possibly improving the broadband behavior of long resonators. The multipass geometries are less sensitive to high-amplitude disturbances, although resonances that occur in the low-amplitude limit may be affected substantially. Resonance frequencies and transmission loss magnitudes near resonance may be expected to decrease while increased dissipation is expected to moderately improve the broadband behavior.

[1] U. Ingard and S. Labate, ''Acoustic circulation effects and the nonlincar impedance of orifices,'' J. Acoust. Soc. Am. **22**(2), 211–218 (1950).
[2] G. B. Thurston and C. E. Martin, Jr., ''Periodic fluid flow through circular orifices,'' J. Acoust. Soc. Am. **25**(1), 26–31 (1953).
[3] T. Notomi and M. Namba, ''Numerical calculation of periodic viscous flow through a circular hole,'' J. Sound Vib. **157**(1), 309–332 (1992).
[4] B. Phillips, ''Effects of high-wave amplitude and mean flow on a Helmholtz resonator,'' NASA TM X-1582, 1968.
[5] G. D. Garrison, A. C. Schnell, C. D. Baldwin, and P. R. Russell, ''Suppression of combustion oscillations with mechanical damping devices,'' Pratt and Whitney Report PWA FR-3299, 1969.
[6] D. Ronneberger, ''The acoustical impedance of holes in the wall of flow ducts,'' J. Sound Vib. **24**, 133–150 (1972).
[7] T. H. Melling, ''The acoustic impedance of perforates at medium and high sound pressure levels,'' J. Sound Vib. **29**(1), 1–65 (1973).
[8] J. W. Sullivan and M. J. Crocker, ''Analysis of concentric-tube resonators having unpartitioned cavities,'' J. Acoust. Soc. Am. **64**(1), 207–215 (1978).
[9] A. Selamet, S. H. Yonak, J. M. Novak, and M. Khan, ''The effect of vehicle exhaust system components on flow losses and noise in firing spark-ignition engines,'' SAE 951260, 1995.
[10] K. Jayaraman and K. Yam, ''Decoupling approach to modeling perforated tube muffler components,'' J. Acoust. Soc. Am. **69**(2), 390–396 (1981).
[11] M. L. Munjal, K. N. Rao, and A. D. Sahasrabudhe, ''Aeroacoustic analysis of perforated muffler components,'' J. Sound Vib. **114**(2), 173–188 (1987).
[12] K. S. Peat, ''The acoustical impedance at discontinuities of ducts in the presence of a mean flow,'' J. Sound Vib. **127**(1), 123–132 (1988).
[13] J. W. Sullivan, ''A method for modeling perforated tube muffler components. I. Theory. II. Applications,'' J. Acoust. Soc. Am. **66**(3), 772–788 (1979).
[14] I. J. Chang and A. Cummings, ''A time domain solution for the attenuation, at high amplitudes, of perforated tube silencers and comparison with experiment,'' J. Sound Vib. **122**(2), 243–259 (1988).
[15] M. Chapman, J. M. Novak, and R. A. Stein, ''Numerical modeling of inlet and exhaust flows in multi-cylinder internal combustion engines,'' in *Flows in Internal Combustion Engines*, edited by Teoman Uzkan (ASME WAM, Phoenix, AZ, 1982).
[16] A. Selamet, N. S. Dickey, and J. M. Novak, ''A time-domain computational simulation of acoustic silencers,'' J. Sound Vib. **117**(3), 323–331 (1995).
[17] A. Selamet, N. S. Dickey, and J. M. Novak, ''The Herschel-Quincke tube:

A theoretical, computational, and experimental investigation,'' J. Acoust. Soc. Am. **96**(5), 3177–3185 (1994).
[18] A. Selamet, N. S. Dickey, and J. M. Novak, ''A theoretical, computational and experimental investigation of Helmholtz resonators with fixed volume: Lumped versus distributed volume,'' J. Sound Vib. **187**(2), 358–367 (1995).
[19] Nolan S. Dickey, ''A time domain approach for acoustic analysis of perforated tube silencers,'' Ph.D. thesis, The University of Michigan, Ann Arbor, MI, 1999.
[20] N. S. Dickey and A. Selamet, ''Acoustic nonlinearity of a circular orifice: An experimental study of the instantaneous pressure/flow relationship,'' Noise Control Eng. J. **46**(3), 97–107 (1998).
[21] B. T. Zinn, ''A theoretical study of non-linear damping by Helmholtz resonators,'' J. Sound Vib. **13**(3), 347–356 (1970).
[22] P. K. Tang and W. A. Sirignano, ''Theory of a generalized Helmholtz resonator,'' J. Sound Vib. **26**(2), 247–262 (1973).
[23] A. Cummings and W. Eversman, ''High amplitude acoustic transmission through duct terminations: Theory,'' J. Sound Vib. **91**(4), 503–518 (1983).
[24] R. L. Panton and A. L. Goldman, ''Correlation of nonlinear orifice impedance,'' J. Acoust. Soc. Am. **60**(6), 1390–1396 (1976).
[25] A. F. Seybert and D. F. Ross, ''Experimental determination of acoustic properties using a two-microphone random-excitation technique,'' J. Acoust. Soc. Am. **61**(5), 1362–1370 (1977).
[26] J. Y. Chung and D. A. Blaser, ''Transfer function method of measuring in-duct acoustic properties. i: Theory,'' J. Acoust. Soc. Am. **68**(3), 907–913 (1980), ''Transfer function method of measuring in-duct acoustic properties. ii: Experiment,'' **68**(3), 914–921 (1980).
[27] P. D. Dean, ''An in situ method of wall acoustic impedance measurement in flow ducts,'' J. Sound Vib. **34**(1), 97–130 (1974).
[28] J. W. Sullivan, ''Theory and methods for modelling acoustically-long, unpartitioned cavity resonators for engine exhaust systems,'' Ph.D. thesis, Purdue University, West Lafayette, IN, 1974.
[29] U. Ingard, ''On the theory and design of acoustic resonators,'' J. Acoust. Soc. Am. **25**(6), 1037–1061 (1953).
[30] N. S. Dickey, A. Selamet, and J. M. Novak, ''Multi-pass perforated tube silencers: A computational approach,'' J. Sound Vib. **211**(3), 435–448 (1998).
[31] E. Fubini-Ghiron, ''Anomalies in acoustic wave propagation of large amplitude,'' Alta Freq. **4**, 530–581 (1935).
[32] A. D. Pierce, *Acoustics* (Acoustical Society of America, New York, 1989).
[33] Z. L. Ji and A. Selamet, ''Boundary element analysis of a three-pass perforated duct muffler, in *Proceedings of Inter-Noise 99*, edited by J. Cuschieri, S. Glegg, and Y. Young, 1999, pp. 557–562.
[34] M. L. Munjal, ''Analysis of a flush-tube three-pass perforated element muffler by means of transfer matrices,'' Int. J. Acoust. Vib. **2**(2), 63–68 (1997).

Interference from degraded auditory stimuli: Linear effects of changing-state in the irrelevant sequence

Dylan M. Jones,[a] David Alford, William J. Macken, Simon P. Banbury, and Sébastien Tremblay
School of Psychology, Cardiff University, Cardiff, United Kingdom

(Received 24 May 1999; revised 5 October 1999; accepted 9 June 2000)

Cognitive performance, particularly on a number of tasks involving short-term memory for order, is impaired by the mere presence of irrelevant background sound. The current study examines the features of the irrelevant sound that determine its disruptive potency. Previous research suggests that the amount of variability in an irrelevant stream is related to the degree of disruption of memory. The present experiments used a parametric approach to manipulate degree of change more precisely. Increasing levels of degradation, effected either by low-pass filtering (speech) or by digital manipulation (speech and nonspeech), monotonically decreased the degree of interference. The findings support the following propositions: (i) the degree of physical change within an auditory stream is the primary determinant of the degree of disruption; and, (ii) the effects of irrelevant speech and irrelevant nonspeech sounds are functionally similar. [S0001-4966(00)02809-5]

PACS numbers: 43.50.Qp [DWG]

I. INTRODUCTION

Exposure to background sound substantially impairs memory for order (e.g., Colle and Welsh, 1976). The most notable feature of this effect is that it occurs reliably even though individuals are fully aware that the background sound is irrelevant and are instructed to ignore it. In a typical irrelevant sound experiment, a series of digits or letters is presented sequentially on a computer screen for recall at the end of the sequence presentation or after a short (typically 10 sec) delay. Sound, either speech or nonspeech, presented during such tasks appreciably disrupts recall of the sequence. The robustness of disruption is evidenced by the frequency with which the effect has been replicated and by the high proportion of individuals sensitive to the effect (see Ellermeier and Zimmer, 1997). Not only can the effect reach up to a 50% decrement but it is also reliable; typically disruption endures over trials (e.g., Tremblay and Jones, 1998) and is just as marked when tested in experimental sessions separated by up to a week (e.g., Ellermeier and Zimmer, 1997; Hellbrück *et al.*, 1996). However, a substantial amount of research indicates that some sounds are more disruptive than others (e.g., LeCompte *et al.*, 1997), and also that some tasks are more prone to disruption than others (e.g., Beaman and Jones, 1997). The functional character of the task is not the subject of the present study; instead, interest centers on the claim that the degree of change in the irrelevant sequence is the primary source of disruption.

Neither the semantic nor the physical similarity of the irrelevant sound to the to-be-remembered material seems to be a major determinant of the effect. It seems fairly certain that the semantic similarity of the irrelevant material to the items being rehearsed in memory is not an important factor. Numerous attempts to find effects of meaning have failed (e.g., Buchner *et al.*, 1996; Jones *et al.*, 1990; LeCompte and Schaibe, 1997) and those that have succeeded only found small and marginally significant effects (e.g., Neely and LeCompte, 1999). Also, there is mounting evidence that physical similarities between the stimuli in the irrelevant stream and the rehearsed material are not a major determinant of disruption. Given that the material rehearsed in memory is usually words or syllables, one might expect that disruption would be brought about only by verbal material such as words or syllables. But it has been shown that speech is neither a necessary (e.g., Salamé and Baddeley, 1989) nor a sufficient (e.g., Jones *et al.*, 1992) condition for substantial disruption to occur. The key determinant of disruption seems to be the degree of change within the irrelevant sound sequence. The prime example of this is that presenting a sequence of different tokens such as letters or syllables results in substantially more disruption of serial recall than when a single token is repeatedly played (e.g., Jones *et al.*, 1992; Jones and Macken, 1995a; LeCompte, 1995). This effect of variation applies for nonverbal and less discrete stimuli also. Tones varying either in frequency or timbre are more disruptive than a repeated tone (e.g., Jones and Macken, 1993) and instrumental music with many sharp changes in energy tends to be more disruptive than music with slow changes (Klatte *et al.*, 1995). The notion of changing-state can be defined as the amount of acoustic variation in frequency, timbre, tempo but not in sound level (see Tremblay and Jones, 1999).

In the experiments that follow, we attempted to show how the degree of change in the auditory stream is related to the disruption of serial recall. One way of testing the concept of changing-state is by selecting naturally occurring stimuli and arranging them on a continuum which is supposed to map onto variability. So, if speech tokens are chosen to rhyme with one another, then the degree of disruption is markedly less than if the tokens are phonologically dissimi-

[a]Author to whom correspondence should be addressed; electronic mail: jonesdm@cardiff.ac.uk

lar (Jones and Macken, 1995a). However, this approach has a number of drawbacks: First, it is far from clear what criteria should be employed in the selection of stimuli to represent the various levels of variability, and, second, it would be difficult to assign a value to each stimulus that could be arranged along a continuum. Overcoming this last drawback would be particularly desirable given that parametrically derived relationships between variables are more informative in analytical terms than a simple contrast between two experimental conditions.

An alternative method is to construct sequences embodying different degrees of change by using a fixed set of stimuli and reducing by degrees the extent to which they change. The general approach adopted in the current study was to systematically degrade the spectral character of changing sequences. By taking a stimulus sequence known to produce an irrelevant sound effect, and systematically degrading it so that the amount of variation within it is gradually reduced, it should be possible to establish a function relating the degree of change to the disruption of memory for order. A similar approach was used by Ellermeier and Hellbrück (1998). They used changes in the signal-to-noise ratio of the irrelevant sequence as the device for degrading the stimulus. That is, they manipulated the degree of change within a sequence by adding more or less noise to a speech sequence. As might be expected on the basis of the changing-state concept, when the signal-to-noise ratio became smaller, so the degree of disruption diminished; moreover, this relationship was monotonic. Abstractly, this is explained by the fact that sequences of noise-masked stimuli have fewer prominent acoustic features; therefore, the degree of variation is reduced gradually as the signal-to-noise ratio diminishes.

However, the technique of using a noise mask is not without its drawbacks (as Ellermeier and Hellbrück, 1998, acknowledge) insofar as the masking effect might be plausibly one primarily of reducing the perceivable number of tokens (distinctive acoustic events) within a stream. This would happen if some of the tokens tended to be more susceptible to masking than others so that, as the signal-to-noise ratio is decreased, they would be reduced to a level below audibility within the noise, with gradually fewer tokens becoming perceptible. Thus, the beneficial effect of adding the noise could be one of reducing the number of tokens, and not one of the relative degrees of change. There is strong evidence to suggest that the degree of disruption is related to the number of tokens per time unit in the irrelevant sequence (referred to as the token-dose effect; see Bridges and Jones, 1996). For example, if the rate at which a changing irrelevant sound sequence is presented is increased, the degree of disruption also increases. However, there is no such effect for a repeated irrelevant sequence.

II. EXPERIMENT 1

In experiment 1, low-pass filtering was used to degrade a stimulus sequence to different degrees (so that some sequences of sound become less distinct). It was predicted that as the degree of filtering increased so the disruption would diminish. The filter is set in such a way as to allow frequencies around the fundamental frequency of the speech sequence to pass unaffected. The degree of degradation of the stimulus by the filter was set by manipulating the rate of roll-off of the filter which served to attenuate progressively the frequencies higher than the fundamental. Steeper rates of roll-off (such as 24 dB per octave) would attenuate many more of these frequencies than shallower degrees of roll-off (say, 6 dB per octave). Phenomenally, the effect of the roll-off becoming steeper would be that the differences between elements in the sound sequence would become less distinct. Importantly, with the settings used in experiment 1, the sound was never completely obliterated even at the steepest rates of roll-off; hence, the token dose remained constant over all levels of roll-off thereby circumventing the difficulty to which the masking technique of Ellermeier and Hellbrück (1998) is prey. The prediction was that, as the roll-off became increasingly steep, the degree of disruption would diminish; that is, the relation between the degree of roll-off and disruption would be inverse and monotonic.

III. METHOD

A. Participants

Twenty-five Cardiff University students each reporting normal hearing and normal or corrected-to-normal vision, received a small honorarium for participating in the experiment.

B. Apparatus and stimuli

1. Visual stimuli for the memory task

Lists of items to be recalled were presented serially on the screen of an Apple Macintosh Performa 6200 microcomputer. Lists were constructed from the random arrangement of the nine letters f, k, l, m, q, r, s, t, and y, with the constraints that a letter could not appear in the same serial position in two consecutive lists and that recognizable words, acronyms, or letter strings were excluded.

2. Auditory stimuli

The words—turn, kilt, band, jaw, fruit, rod, porch, pier, nest, cow, pin, hat, boat, top—were recorded digitally in a male voice, spoken at an even pitch. The person producing the words was asked to produce them at a pitch matching that of a reference pure tone of 103 Hz (A♭2) which was played periodically to him as a reminder throughout the recording session. During the preparation of the materials, any word not matching this pitch was re-recorded. The words were edited using digital signal processing techniques to last 500 msec with a 200-msec inter word interval. The sequence was then band pass filtered using a Kemo Dual Variable Filter Type VBF/14. For each of the auditory conditions, the center frequency of the filter was set at 85 Hz and the lower bound of the filter (the high-pass element) was fixed at a roll-off of 24 dB/octave. Only the rate of attenuation of the frequencies above the center frequency (the low-pass component) was manipulated across auditory conditions within the experiment. The roll-off values for the low-pass component were set at either 0, 6, 12, 18, or 24 dB/octave. The

value of the center frequency of the filter—at 85 Hz, lower than the fundamental frequency of the speech—was arrived at following pilot trials to meet the criterion that the lower degrees of roll-off (6 dB/octave vs 0 roll-off) produced a perceptible difference in attenuation, at least when judged in focal attention. The range of values of roll-off used in the study was constrained by the values available on the Kemo filter.

The stimuli were recorded on digital audio tape, and transferred thereafter to Sound Designer digital editing software where the amplitude was digitally edited to constant root-mean-square (rms) value for all stimuli. Generally, this meant that the stimuli were broadly similar in loudness. There is a substantial body of evidence showing that both the overall level of intensity (e.g., Ellermeier and Hellbrück, 1998) and the degree of change in intensity (Salamé and Baddeley, 1989; Tremblay and Jones, 1999) do not influence the degree of disruption of serial recall. All samples were recorded to 16-bit resolution at a sampling rate of 44.1 kHz, stored digitally, and presented during experimental trials within a SuperCard environment.

C. Design

A repeated-measures design was used, all participants undertaking the recall task under all five auditory conditions. There were 75 trials in all, 15 for each level of roll-off. Presentation of the conditions was quasirandomized from trial to trial, such that every condition was presented before any one was repeated.

D. Procedure

Participants were tested individually, seated in a soundproof laboratory approximately 0.5 m from the computer screen. At the outset, standard instructions were read by participants. These informed them of the nature of the recall task, asked them to ignore any sounds they might hear, and reassured them that the content of the auditory material would not be tested. In each trial, the nine letters were displayed in random order as described above. Each letter was displayed for 0.5 sec with an interstimulus interval also of 0.5 sec. Every list was preceded by a warning tone. After presentation of the last item in the list, the word ''wait'' was displayed on the screen for 10 sec during which the subject was expected to rehearse the list. After this retention interval, once the word ''wait'' disappeared, the subject was required to recall the list in strict serial order. Sound was played via headphones (Sony CD250 Digital Reference) throughout both presentation and retention of the lists at a level of approximately 60 dB(A). Responses were written on a blank grid comprising rows of nine boxes corresponding in order to the to-be-remembered sequence. The experimental trials were preceded by a short practice session, and the experiment lasted approximately 45 min.

IV. RESULTS

The results were scored to a strict serial recall criterion. Each item in the recalled sequence was scored as correct only if it corresponded exactly to the item at that position in the to-be-remembered sequence. The main feature of the results is depicted in Fig. 1, which shows that as the degree of filtering increased the degree of disruption diminished; moreover, this relationship is monotonic. The error percentages were subjected to a two-way repeated measures analysis of variance, which revealed significant main effects of serial position (nine positions), $F(8,192)=30.81$, $MSE=16.77$, $p<0.001$, of auditory condition (five levels of roll-off), $F(4,96)=4.91$, $MSE=7.84$, $p<0.01$, and also of their interaction, $F(32,768)=1.86$, $MSE=2.50$, $p<0.05$. No functional significance is attached to the interaction between serial position and auditory condition since it may simply reflect a ceiling effect in recency (analysis of serial position data revealed the same pattern in experiment 2 but are not reported).

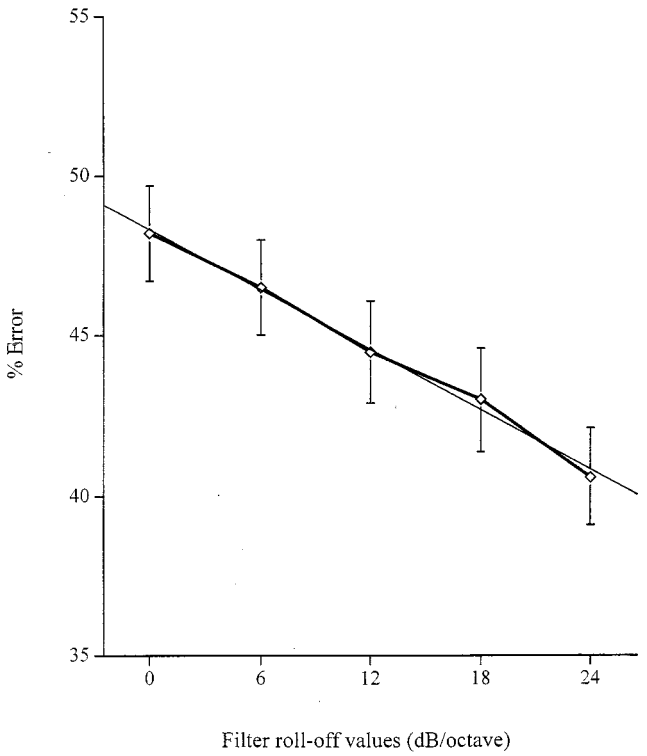


FIG. 1. Experiment 1: Percent serial recall errors in relation to the degree of low-pass filtering of the irrelevant sound. Errors are pooled over serial positions. A linear fit is superimposed to the curve (error bars represent standard error).

Further scrutiny of the significant effect of auditory conditions shows that, using a test of orthogonal polynomials, the trend has a strongly linear component, $F(4,96)=19.34$, $p<0.0001$, and, as expected, a nonsignificant quadratic component, $F(4,96)=0.19$, $p>0.05$. It could be argued that the linear function is the result of averaging many nonlinear functions. Given the relatively great range of roll-offs employed, the average of nonlinear functions should have led to more variability in the middle range, however, visual inspection of the error bars in Fig. 1 (representing standard error) clearly shows that the variability is similar across the range of roll-off points.

V. DISCUSSION

The monotonic relation between the degree of degradation in the stimulus and the degree of disruption points to the

fact that the effect of changing-state is continuous; there is no evidence of a cutoff point above which the disruption becomes more pronounced. That is, there is no evidence here of a signal level which is necessary for the effect to occur, as one might suppose if the effect is dependent upon the categorical perception of the words or their constituent phonemes, for example. This result supports the suggestion that the effect of irrelevant sound is not dependent on the identity of the sounds be that identity semantic, lexical, or phonological.

Although the filtering technique is one that provides clear results for the purposes of supporting the concept of changing-state, its main shortcoming is that it is difficult to apply when attempting to make a comparison across several different classes of stimuli varying greatly in complexity. For example, in making a comparison of the degradation in a speech sequence with that in a sequence of notes from a musical instrument, it would be necessary to adopt vastly different degrees of roll-off in each case (given that the distribution of the energy across the audio spectrum is far less wide in the case of a typical musical instrument). Moreover, the fundamental frequency of each irrelevant stimulus would have to be the same. Given these differences, it might be difficult to be sure that very different ranges of value of roll-off have equivalent effects with each stimulus. Ideally, it would be useful to use a technique that would employ a common metric for the degree of degradation with different classes of stimuli.

We attempted to use just such a technique in experiment 2; instead of low-pass filtering, we employed a method involving a digital signal processing technique in which an intact stimulus (say, a spoken word) is sampled digitally and then the polarity of each of the sample points constituting the stimulus is reversed with a certain probability. By systematically varying the likelihood of reversal, a range of stimuli may be generated that bear different degrees of resemblance to the original token, at one extreme being composed entirely of amplitude-modulated noise, for which the likelihood of polarity reversal is .5, through intermediate stages in which there are approximations to the word (accompanied by noise), to the other extreme in which none of the samples is changed and the stimulus is in its original form. Of course, this interference function can in theory be applied to any auditory stimulus with a common metric, namely the probability of a change in polarity. This allows a contrast of the effect of speech and nonspeech stimuli to be made. This is what we attempted in experiment 2, by comparing speech tokens with tokens of a cello note played at different frequencies.

In phenomenal terms the steps in the degradation of the stimulus will give rise to progressive lack of clarity of the stimulus, accompanied by a loss of lexical and phonological identity to the point where only very little of the original stimulus can be discerned. Eventually, when the stimulus is composed entirely of amplitude-modulated noise, a noise burst will be heard. In the case of the cello note, the amplitude of each token is the same, and hence the noisiest version of the transformation will be the same for each token. For the words, there will be small variations in the pattern of noisiness, reflecting minor variations in the amplitude envelopes of the words that served originally as the basis for the transformations.

The purpose of experiment 2 was to explore the shape of the interference function relating degradation of the stimulus sequence and serial recall for different classes of stimuli. It was predicted that the function would be monotonic and general; that is, the degree of disruption would increase as the degree of change increased, and this would be true for both speech and nonspeech (instrumental notes).

VI. EXPERIMENT 2

In experiment 2, interference functions were established for nonspeech (cello notes) and speech (syllable sequences) stimuli. Again, the concept of changing-state makes the prediction that both will have monotonic interference functions, but that, because there are more changing elements in the speech than in the cello note, the slope of the function will be steeper for speech than for cello notes. Six levels of noisiness were compared.

VII. METHOD

A. Participants

Forty-seven participants volunteered to take part in the experiment for a small honorarium. All reported normal hearing and vision.

B. Apparatus and stimuli

1. Visual stimuli

The short-term memory task was the same as that used in experiment 1.

2. Auditory stimuli

Two types of stimuli were employed: cello notes and spoken single-syllable words. Seven words[1]—pier, nest, cow, pin, hat, boat, and top—were recorded digitally in a male voice. Digital samples of a plucked cello note were taken from a compact disk containing musical samples (Digidesign SampleCell). Twelve samples of the plucked cello note covering the octave C4 to B4 served as the basis for the transformations. Words and cello notes were edited to last approximately 500 msec with a 100-msec interstimulus interval. From each of the speech and cello note samples, five additional versions were generated by digital signal processing to form a set of transformed samples along a continuum of noisiness. This was achieved by first representing the sample digitally (with a sampling rate of 22 kHz) and then reversing the polarity of each sample point with different probabilities (Schroeder, 1968; Repp and Frost, 1988). In essence, this changed the proportion of the sample that consisted of noise. As the probability of changing sample points increased, so did the proportion of noise in relation to the signal increase, but the length of the sample and its amplitude envelope remained the same. Six levels of degradation of the original sample were prepared: 0%, 20%, 40%, 60%, 80%, and 100% (corresponding to probabilities of reversing polarity of 0, 0.1, 0.2, 0.3, 0.4, and 0.5). These values en-

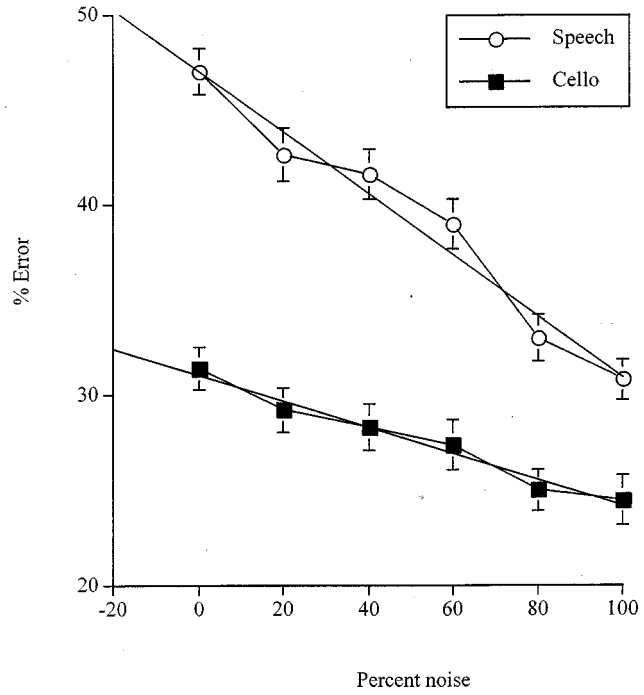


FIG. 2. Experiment 2: Percent recall errors for speech sounds and cello notes as a function of the degradation. Linear fits are superimposed (error bars represent standard error).

compass the full range of degradation; at 0%, the original signal was intact; at 100%, the sample was completely reduced to amplitude-modulated noise.

Using this transformation technique, 6 versions of each of the 12 samples of cello notes and speech were obtained (one original and five copies at different levels of degradation). Sequences of sound at each level of degradation were then assembled to form the experimental conditions. Thus, for example, one condition contained sequences made up of untransformed cello notes, another of notes at another level of degradation, and so forth. The order of stimuli in each sequence was random, with the constraint that no two successive stimuli were the same. These sequences were stored digitally, and retrieved via software according to a predetermined sequence during the memory task.

C. Design and procedure

The design used here was similar to that employed in experiment 1, save that there were two repeated-measures factors, namely, class of sound (speech vs cello) and degree of degradation (six levels). Thus participants were exposed to 12 different conditions. There were 12 trials per condition for a total of 144 trials. Because of the large number of trials there was a 10-min rest break half way through the experimental session. The general procedure was the same as for experiment 1. The experiment took about 100 min.

VIII. RESULTS

As in experiment 1, participants' responses were scored according to the strict serial order criterion. The main features of the results are illustrated in Fig. 2. Errors decreased as the level of noise increased; the slope of this improvement was less steep for the cello notes than for the speech stimuli. The two factors—degradation (six levels in the range 0% to 100% noise) and source (cello speech)—were entered into a repeated-measures analysis of variance (ANOVA). The disruption produced by speech was greater than that produced by cello notes, $F(1,46)=5.63$, $MSE=23.09$, $p<0.02$. The main effect of percent noise was highly significant, the level of disruption decreasing markedly as the proportion of noise increased, $F(5,230)=13.56$, $MSE=0.92$, $p<0.0001$. The two factors also interacted significantly, $F(5,230)=2.37$, $MSE=0.92$, $p<0.05$. The results are broadly in line with predictions. Speech showed more disruption than nonspeech, and the degree of disruption rose more rapidly as the percentage of noise decreased. In separate within-participants ANOVAs, each function produced a significant fit to linear polynomials, but not to higher-order polynomials [for cello, linear contrast: $F(1,46)=13.72$, $p<0.001$; quadratic contrast: $F<1$; for speech, linear contrast $F(1,46)=56.25$, $p<0.001$; quadratic contrast: $F<1$].

IX. DISCUSSION

The results may be summed up succinctly: As stimulus degradation was reduced, the effect of interference from the auditory stimulus increased, and this effect was significantly linear. Just as in experiment 1, the form of these results is interesting, in that there is no discontinuity in the effect of degradation, as one might expect if there were categorical processes at work, such as the emergence of meaning or the detection of lexical, phonological, or other sublexical units. There is, in other words, no cutoff point that a degraded stimulus must exceed in order to become classified as a word before the disruption becomes marked. This monotonic function can be explained by the concept of changing-state. Simply, as more of the signal becomes available, the amount of variation will be greater and so the degree of disruption will also increase.

X. GENERAL DISCUSSION

Gradual degradation of a coherent but changing stimulus sequence leads to a smooth monotonic function in relation to the degree of disruption. This confirms the result of Ellermeier and Hellbrück (1998) with techniques that do not suffer the potential drawback of confounding token dose with degree of mismatch. The result also suggests that there is no single point above which the effect of disruption becomes evident, such as might be the case if some threshold of lexical or sublexical identity had to be breached. Thus far the experimental series has generated outcomes in line with the proposition that, as the successive items in a series of irrelevant sounds show less variation, the degree of disruption of serial recall diminishes.

With regard to the linear relationship between degradation and disruption, there are indications that both speech and nonspeech stimuli are functionally equivalent in this regard. In experiment 2, simple cello notes seemed to yield similar trends to those produced by speech. This finding constitutes support for the claim that speech and nonspeech are functionally similar in producing disruption of memory for order. The fact that fully degraded speech is still more dis-

ruptive than fully degraded nonspeech, may, however, be taken as evidence to the contrary. One possible explanation of this difference in disruptive potency is that the fully degraded speech sounds still contain some spectral changes from token to token, given the different amplitude envelopes associated with the different speech tokens. While previous studies have shown that amplitude-modulated broadband noise is not disruptive of recall (e.g., Salamé and Baddeley, 1989), the stimuli used in those studies may not have been perceptually segmentable into discrete tokens, a condition necessary for disruption to occur with any signal (Jones and Macken, 1993). However, the 100-ms silent interval between tokens used in experiment 2 is likely to have been sufficient for such segmentation to occur. In any case, although the levels of performance were different with cello notes and with speech, they seemed to be related to the degree of degradation of the irrelevant sequence in the same lawful way.

A similar argument about the functional equivalence of speech and nonspeech sounds can be derived from other factors such as the way in which token set size and disruption are related for tones and speech (Tremblay and Jones, 1998). That is, both speech and tones show the same relation between disruption and token set size: There is an increase in error as the number of distinguishable sounds increases from one to two, but there is very little increase in error as set size increases thereafter. Differences in terms of the relative degrees of disruption for speech and nonspeech are sometimes significant (e.g., LeCompte *et al.*, 1997) and in some other studies nonsignificant (e.g., Jones and Macken, 1993). The interpretation of such differences depends in part on how one regards the comparison between simple speech sounds and simple nonspeech sounds. Of course, at one level, this comparison is of doubtful value, given that in acoustic terms speech sounds that are nominally simple are typically more complex than nonspeech sounds.

While the studies reported here clearly demonstrate a linear relationship between the amount of change within an auditory sequence and its power to disrupt recall, the precise mechanism of this effect is perhaps less clearly established. In particular, an as-yet unanswered question is whether change is important as a global property of a sound sequence, or operates on a local, token-to-token mismatch basis. Results from Jones *et al.* (1992) suggest that global properties are not the critical determinants of disruption, at least insofar as predictable sequences of tokens (e.g., A, B, C, D, A, B, C, D, etc.) are no less disruptive than randomly changing, unpredictable sequences (e.g., A, C, D, B, C, A, B, D, etc.). Further, there is evidence pointing to the fact that change is derived from information on the basis of two immediately successive stimuli: The largest increase in disruption occurs when the set of distinguishable tokens goes from one repeated token to two distinct tokens, with little or no further increase in disruption from sequences containing larger numbers of tokens (see above; Tremblay and Jones, 1998). However, the effects of manipulating the rate of transition between different tokens in the irrelevant sound are not in line with this mismatch hypothesis. A low rate of transition, at which the token was changed every three stimuli (e.g., AAABBBCCC), should produce less disruption than a sequence of irrelevant sound in which the transition is between immediately successive tokens (e.g., ABC). However, these two conditions do not produce significantly different degrees of disruption (see experiment 4 of Tremblay and Jones, 1998). Whether the effect of changing-state is calculated over adjacent tokens remains to be further investigated.

One intuitively appealing and parsimonious account of the irrelevant sound effect is that repeated tokens (or unvarying continuous sound) produce less disruption because by virtue of their repeated presentation they do not produce an attentional response. Based on the concept of the orienting response (see Sokolov, 1963), it was suggested that irrelevant sound recruits attention away from the task at hand but repetitive stimuli cause much less disruption because the attentional response to the sound attenuates as a result of repeated presentation (see Cowan, 1995). However, a number of lines of empirical evidence suggest that habituation plays little or no role in the irrelevant sound effect. First, a key prediction of the habituation framework is that the degree of disruption should increase as the number of different tokens in the irrelevant sound sequence increases. With few tokens in a sequence, each token is repeated more often than if there are many, and therefore habituation should be brought about more rapidly. However, as mentioned earlier, it has been shown that disruption increases sharply when the set of distinct tokens increases from one to two, but beyond two tokens the degree of disruption does not increase significantly (Tremblay and Jones, 1998). Second, a further strong prediction is that the effect of irrelevant sound should diminish as trials are repeated, particularly so if they contain repeated tokens, but empirical evidence shows no hint of diminution in disruption over successive trials or over experimental sessions (Ellermeier and Zimmer, 1997; Hellbrück *et al.*, 1996; see also Jones *et al.*, 1997). A further difficulty is a logical one. The degree of disruption is not the product solely of the nature of the sound; it is less pronounced in tasks not requiring serial order. If the irrelevant sound effect was due wholly to the recruitment of attention from the task, it seems problematic that the nature of the task determines in part the degree of disruption (e.g., Beaman and Jones, 1997, 1998; Salamé and Baddeley, 1990).

A more tenable account of the effect of changing-state is that, generally, the effect of irrelevant sound on serial recall is due to concurrent processes of ordering (or seriation).[2] One source of seriation relates to the maintenance of order during rehearsal of the lists in the memory task (see, e.g., Jones *et al.*, 1996, for a discussion). The other is automatic and reflects the seriation of sound. Generally, it has been assumed that the greater the quantity of changing-state information within the sound sequence, the stronger the cues to seriation (and so the greater the disruption). However, there is some evidence to suggest that the degree of disruption is not predicted solely by a consideration of the degree of change, but it seems the effects of changing-state can be modulated by the rules of auditory organization generally known as ''streaming'' (Bregman, 1990). For example, if the effect of presenting a fixed sequence of three syllables monaurally (so that a single stream containing a changing sequence is formed) is contrasted with the effect of same se-

quence presented with each syllable assigned to a different location in auditory space (now forming three streams of unchanging syllables), disruption is much less marked in the latter case (Jones and Macken, 1995b). Thus it seems that the effect of degree of change can be understood only in relation to the organization of sounds into streams and to the ensuing consequences of change within such streams. The results of the current series reinforce the generalization that the degree of change within a single coherent stream is related to the magnitude of the irrelevant sound effect.

ACKNOWLEDGMENTS

This work received support from the UK's Economic and Social Research Council (ESRC) in the form of a project grant. Dylan Jones is also a Visiting Fellow at the Air Systems' Sector of the Defense Evaluation and Research Agency, Farnborough, Hampshire, UK. David Alford received support from an ESRC CASE studentship. Particular thanks are due to Clive Frankish for his help in undertaking the signal transformation used in experiment 2. Thanks are also due to Karen Anderson Howes and Robert Houghton for a critical reading of an earlier version of the paper.

[1]We used a set of 7 irrelevant words in experiment 2 and a set of 14 words in experiment 1. There is a difference in the set size because the two experiments were run in different laboratories, and the available sound files were employed. Also, 12 cello notes were used in experiment 2 as there are 12 semitones in one octave. As demonstrated in Tremblay and Jones (1998), the degree of disruption by irrelevant sound is independent of the irrelevant token set size.

[2]There are two other prominent accounts of the irrelevant speech effect, the phonological loop hypothesis (Salamé and Baddeley, 1982) and the feature model (Neath, 2000), but they do not offer an explanation for the effect of irrelevant nonspeech sounds.

Beaman, C. P., and Jones, D. M. (**1998**). ''Irrelevant sound disrupts order information in free as in serial recall,'' Q. J. Exp. Psych. **51A**, 615–636.

Beaman, C. P., and Jones, D. M. (**1997**). ''The role of serial order in the irrelevant speech effect: Tests of the changing state hypothesis,'' J. Exp. Psychol. **23**, 459–471.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA).

Bridges, A. M., and Jones, D. M. (**1996**). ''Word-dose in the disruption of serial recall by irrelevant speech: Phonological confusions or changing state?,'' Q. J. Exp. Psychol. **49A**, 919–939.

Buchner, A., Irmen, L., and Erdfelder, E. (**1996**). ''On the irrelevance of semantic information for the 'irrelevant speech' effect,'' Q. J. Exp. Psychol. **49A**, 765–779.

Colle, H. A., and Welsh, A. (**1976**). ''Acoustic masking in primary memory,'' J. Verbal Learn. Verbal Behav. **15**, 17–31.

Cowan, N. (**1995**). *Attention and Memory: An Integrated Framework* (Oxford University Press, Oxford).

Ellermeier, W., and Hellbrück, J. (**1998**). ''Is level irrelevant in 'irrelevant speech'? Effects of loudness, signal-to-noise ratio, and binaural masking,'' J. Exp. Psychol. **24**, 1406–1414.

Ellermeier, W., and Zimmer, K. (**1997**). ''Individual differences in susceptibility to the 'irrelevant speech' effect,'' J. Acoust. Soc. Am. **102**, 2191–2199.

Hellbrück, J., Kuwano, S., and Namba, S. (**1996**). ''Irrelevant background speech and human performance: Is there long-term habituation?,'' J. Acoust. Soc. Jpn. **17**, 239–247.

Jones, D. M., Beaman, C. P., and Macken, W. J. (**1996**). ''The object-oriented episodic record model,'' in *Models of Short-Term Memory*, edited by S. Gathercole (Lawrence Erlbaum Associates, London), pp. 209–238.

Jones, D. M., and Macken, W. J. (**1993**). ''Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory,'' J. Exp. Psychol. **19**, 369–381.

Jones, D. M., and Macken, W. J. (**1995a**). ''Phonological similarity in the irrelevant speech effect: Within- or between-stream similarity?,'' J. Exp. Psychol. **21**, 103–115.

Jones, D. M., and Macken, W. J. (**1995b**). ''Organizational factors in the effect of irrelevant speech: The role of spatial location and timing,'' Mem. Cognit **23**, 192–200.

Jones, D. M., Madden, C., and Miles, C. (**1992**). ''Privileged access by irrelevant speech to short-term memory: The role of changing state,'' Q. J. Exp. Psychol. **44A**, 645–669.

Jones, D. M., Miles, C., and Page, J. (**1990**). ''Disruption of proof-reading by irrelevant speech: Effects of attention, arousal or memory?,'' Appl. Cogn. Psychol. **4**, 89–108.

Jones, D. M., Macken, W. J., and Mosdell, N. (**1997**). ''The role of habituation in the disruption of recall performance by irrelevant sound,'' Br. J. Psychol. **88**, 549–564.

Klatte, M., Kilcher, H., and Hellbrück, J. (**1995**). ''Wirkungen der zeitlichen struktur von hintergrundschall auf das arbeitsgedachtnis und ihre theoretischen und praktischen implikationen'' [''The effects of temporal structure of background noise on working memory: Theories and applications''], Z. Exp. Psychol. **42**, 517–544.

LeCompte, D. C. (**1995**). ''An irrelevant speech effect with repeated and continuous background speech,'' Psych. Bull. Rev. **2**, 391–397.

LeCompte, D. C., Neely, C. B., and Wilson, J. R. (**1997**). ''Irrelevant speech and irrelevant tones: The relative importance of speech to the irrelevant speech effect,'' J. Exp. Psychol. **23**, 472–483.

LeCompte, D. C., and Shaibe, D. M. (**1997**). ''On the irrelevance of phonological similarity to the irrelevant speech effect,'' Q. J. Exp. Psychol. **50A**, 100–118.

Neath, I. (in press). ''Modeling the effects of irrelevant speech on memory,'' Psychon. Bull. Rev.

Neely, C. B., and LeCompte, D. C. (**1999**). ''The importance of semantic similarity to the irrelevant speech effect,'' Memory Cogn. **27**, 37–44.

Repp, B. H., and Frost, R. (**1988**). ''Detection of words and non-words in two kinds of noise,'' J. Acoust. Soc. Am. **84**, 1929–1932.

Salamé, P., and Baddeley, A. D. (**1982**). ''Disruption of short-term memory by unattended speech: Implications for the structure of working memory,'' J. Verbal Learn. Verbal Behav. **21**, 150–164.

Salamé, P., and Baddeley, A. D. (**1989**). ''Effects of background music on phonological short-term memory,'' Q. J. Exp. Psychol. **41A**, 107–122.

Salamé, P., and Baddeley, A. D. (**1990**). ''The effects of irrelevant speech on immediate free recall,'' Bull. Psych. Soc. **28**, 540–542.

Schroeder, M. R. (**1968**). ''Reference signal for signal quality studies,'' J. Acoust. Soc. Am. **43**, 1735–1736.

Sokolov, E. N. (**1963**). *Perception and the Conditioned Reflex* (Pergamon, London).

Tremblay, S., and Jones, D. M. (**1998**). ''Role of habituation in the irrelevant sound effect: Evidence from the effects of token set size and rate of transition,'' J. Exp. Psychol. **24**, 659–671.

Tremblay, S., and Jones, D. M. (**1999**). ''Change of intensity fails to produce an irrelevant sound effect: Implications for the representation of unattended sound,'' J. Exp. Psychol. **25**, 1005–1015.

Acoustic optic hybrid (AOH) sensor

Anthony D. Matthews[a] and Lisa L. Arrieta
Naval Surface Warfare Center, Dahlgren Division, Coastal Systems Station, Panama City, Florida 32407-7001

(Received 17 March 1997; revised 7 April 1999; accepted 17 March 2000)

The ability of laser vibrometers to receive and process acoustic echoes from the water surface above a submerged target is established and evaluated. Sonar echoes from a submerged target are collected from the water surface by a laser vibrometer. Feasibility of this approach to sensing underwater sound is demonstrated. If the acoustic excitation at an otherwise undisturbed water surface is 195 to 168 dB *re*: 1 μPa, signal-to-noise ratio (SNR), at the vibrometer output, is shown to range from about 46 to 6 dB. Capillary waves and gravity waves at the water surface are expected and shown to have some destructive effect on the process of echo retrieval. A series of experiments to quantify the surface wave effects is described. The wave experiment results are reported. A successful attempt to acquire echoes from a submerged target over a grid of points for further processing into a three-dimensional image is made and described. The data acquisition and beamforming techniques constitute a three-dimensional, acoustic optic, synthetic aperture sonar (SAS). Beamformed images are included. For an aircraft towing acoustic sensors through the water with a mechanical link, this technique holds the promise of increased safety and improved fuel efficiency.
[S0001-4966(00)05906-3]

PACS numbers: 43.60.Gk, 43.60.Rw, 43.30.Yj [JCB]

INTRODUCTION

Measurement of suspended particle velocity by observation of Doppler shifts in laser light scattered from those particles was demonstrated by Yeh and Cummins[1] in 1964. The technique became known as laser Doppler velocimetry (LDV) and depends on the premise that particles will be suspended in, and move with, the fluid being examined. In the mid 1970s, Taylor[2,3] measured particle velocities in air that supported standing and traveling acoustic waves. In 1990, Vignola, Berthelot, and Jarzynski[4] demonstrated LDV measurement of acoustic particle displacements of standing waves generated in a water-filled tube. This study differs from previous work in that it explores retrieval of acoustic energy from the fluid surface rather than particles in the fluid, quantifies the interference imposed by surface waves, and attempts to beamform acoustic echoes from a submerged target by employing synthetic aperture sonar (SAS) techniques. Motivation for this study stems from a desire to conduct sonar operations from an aircraft without the onerous and dangerous need for towing transmitters and receivers in the water. It should be noted, however, that this method has broad application for searches in murky or opaque liquids. Key contributions of this work include experimental evaluation of an affordable LDV in terms of SNR, given a calm surface, quantification and analysis of the effects of surface waves on the technique, and synthesis of a three-dimensional aperture from LDV data. The first provides an estimate of how good this approach could be if the surface wave effects could be resolved, as well as establishing feasibility. The second highlights the magnitude of the surface wave problem and offers a suggestion as to the mechanism of the problem. The third illustrates a path for integration of the technique into an affordable system architecture.

[a]Electronic mail: matthewsad@ncsc.navy.mil

I. DYNAMIC RANGE ESTIMATES

Consider an acoustic projector transmitting a continuous frequency in a test pool with a calibrated hydrophone at a distance from the source equal to the source depth. An LDV device, referred to as a vibrometer, suspended above the water surface over the acoustic source, may be expected to measure the acoustic energy if its operating envelope contains the dynamic range of signals. From Urick[5] (assuming the wave to be planar over the measurement spot), the particle of water just beneath the surface has a velocity related to the pressure, sound velocity in water, and fluid density by

$$p = \rho c \nu,$$

where p is the pressure in Pa, ρ is the fluid density in kg/m^3, c is the sound velocity in the fluid in m/s, and ν is the particle velocity in m/s. Since the particle velocity doubles at the water surface, displacement and velocity can be related as follows:

$$\nu = \omega X = 2(p/\rho c)(R_0/D),$$

where X is particle displacement at the water surface, ω is the radian frequency, p is the pressure measured at the reference position, R_0 is the range from source to pressure reference measurement, and D is the vertical distance from source to surface. Note that this equation relates only the amplitudes of particle velocity, particle displacement, and pressure, and not their respective phases. This simplistic model is based on spherical waves, which is an acceptable assumption for proof of concept.

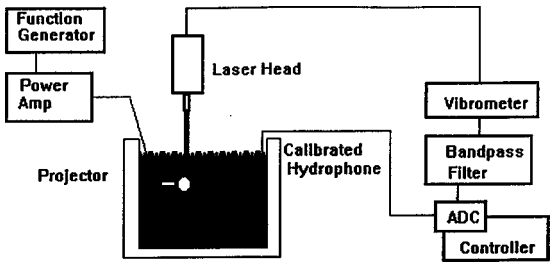


FIG. 1. AOH feasibility experiment.

Consider the case where the measurement is made at the surface directly above a 7-kHz source. If R_0 is made equal to D, the last term in the equation conveniently becomes 1. Thus, if the sound-pressure level just beneath the water surface is 195 dB *re*: 1 μPa, the pressure just beneath the surface is 5623 Pa, the surface velocity is 7.5 mm/s, and the particle displacement at the surface is 1.7×10^{-7} m in amplitude. At a sound-pressure level of 165 dB *re*: 1 μPa, the pressure just beneath the water surface is 178 Pa, the surface velocity is 2.34×10^{-4} m/s, and the particle displacement is 5.46×10^{-9} m.

A laser vibrometer, whose operating envelope, at 7 kHz, runs from peak vibration amplitudes of approximately 2×10^{-7} to 1.0×10^{-10} m,[6] was selected for the experiment. This vibrometer was a TSI model 1930 controller together with a model 1940 optic probe. Since the particle displacements expected are within the limits imposed by the instrument operating envelope, the experimental data can be expected to remain linear (linear here indicates that signal clipping or amplifier saturation will not occur) and, therefore, beamformable in this power region. The laser wavelength was 780 nm, and consequently the penetration depth in the water was substantial. However, the principle of the probe operation is based on the reflection from a surface at a calibrated range of 19.2 in. As the specular surface reflection was much larger than any other, the penetration was considered negligible.

II. EXPERIMENTS AND RESULTS

Four experiments were conducted. A feasibility experiment was followed by a capillary wave sensitivity experiment, a gravity wave sensitivity experiment, and a target imaging experiment.

A. Feasibility experiment

The basic experiment setup, illustrated in Fig. 1, was designed to demonstrate feasibility. After establishment of feasibility, the experiment was extended to find the SNR sensitivity to excitation level.

The feasibility experiment collected digitized records of one-half-second duration from the vibrometer output, sampled at 33.33 kHz. Ten time records were taken with the ITC 4-in. spherical projector transmitting at 7 kHz. Ten more records were made with the projector silent.

Fourier transforms of these 20 records were used to estimate the vibrometer SNR in the 7-kHz frequency bin.

The calibrated hydrophone used was an ITC model 6069. The hydrophone receive response is listed by the manufacturer as -206 dB *re*: 1 V/1 μPa. Because the reference hydrophone was placed at a distance from the projector equal to the projector depth, the surface excitation (SE), in dB *re*: 1 μPa, may be expected to obey the following relation:

$$\mathrm{SE_{dB}} = 206 + 20\log_{10}(V_{\mathrm{pk}}/\sqrt{2}),$$

where V_{pk} is the measured voltage peak at the hydrophone and SE is arbitrarily defined as the acoustic level delivered by the transducer to a point directly above the transducer and just beneath the water surface.

Observing an oscilloscope that monitored the reference hydrophone provided the surface excitation entry for any corresponding data set taken with the vibrometer.

TABLE I. Excitation level vs AOH sensor SNR.

Excitation level (dB *re*: 1 μPa)	Vibrometer output SNR (dB)
195.5	47.1
189.9	39.1
179.6	34.7
172.4	22.9
168.6	6.7

1. Feasibility experiment results

Acoustic optic hybrid (AOH) feasibility experiment observations are shown in Table I. These results are associated with a calm surface. No surface waves were intentionally induced during these trials.

Figure 2 graphs the spectral power density in the vibrometer output with the 7-kHz source turned on in trial numbers 11 through 20. The acoustic source is silent in trials 1 through 10. The nominal frequency of 7 kHz was actually centered closer to 6.92 kHz.

B. Capillary wave experiment

By adding a variable-speed fan and a wave-height measurement device to the feasibility experiment, the vibrometer output SNR could be measured as a function of surface wave height. Capillary waves are defined as those waves whose restoring force is surface tension. In water, this type of wave is arbitrarily assigned peak wave heights of 3.2 mm or less

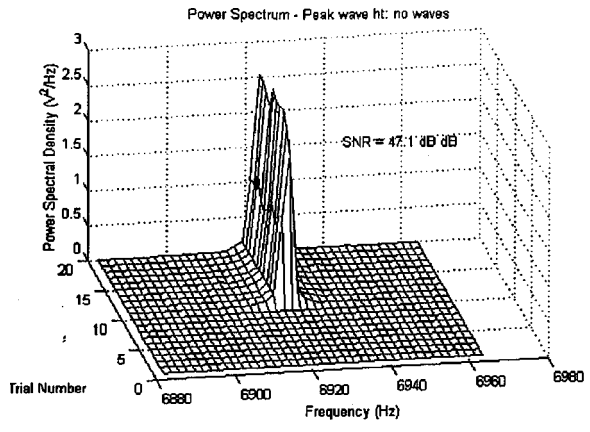


FIG. 2. Vibrometer output power spectral density with acoustic source off, then on.

TABLE II. AOH sensor SNR as a function of capillary wave height.

Positive peak wave height (mm)	Vibrometer SNR (dB)
0	47.1
1	33.7
2	28.5
3.2	28.5

for these experiments. The visual wave meter was assembled using a magnifying glass and a ruler suspended vertically across the water–air interface. The surface excitation was held constant at 196 dB *re*: 1 μPa.

1. Capillary wave experiment results

Capillary wave experimentation results appear in Table II. The SNR deterioration is surprising for waves of such modest height.

Data gathered and processed in similar fashion to that described in the AOH feasibility experiment, using an excitation level of 196 dB *re*: 1 μPa, provided the values shown in Table II.

C. Gravity wave experiment

Gravity waves are defined as those surface waves whose restoring force is gravity. Replacement of the fan with a manual paddle allowed the generation of peak wave heights from zero to above 50 mm. Paddle motion producing waves with positive peak heights of 12.7, 25.4, and 50.2 mm was employed. Waves of these heights are considered to be gravity waves. The gravity wave experiment was, in other respects, identical to the feasibility and capillary wave experiments.

1. Gravity wave experiment results

Table III tabulates the results of the gravity wave sensitivity trials. These data were gathered and processed in the same way as those of the capillary wave experiment series.

The wave experiments were embarked upon with the expectation that surface waves would introduce a noise source that would add linearly with the desired acoustic signal to decrease the vibrometer SNR.

Careful examination of the time-series data suggests that the noise is more complicated than anticipated. Three examples of such time-series data appear in Fig. 3. The acoustic signal is obvious when it is present. The surface waves are almost always apparent. The acoustic signal, however, disappears from view for periods that lengthen as the surface wave height increases.

Conventional linear digital filtering will not restore a declining SNR due to a signal record that contains signal holidays of duration proportional to wave height.

TABLE III. AOH sensor SNR as a function of gravity wave height.

Positive peak wave height (mm)	Vibrometer SNR (dB)
12.7	9.4
25.4	0.6
50.8	−0.1

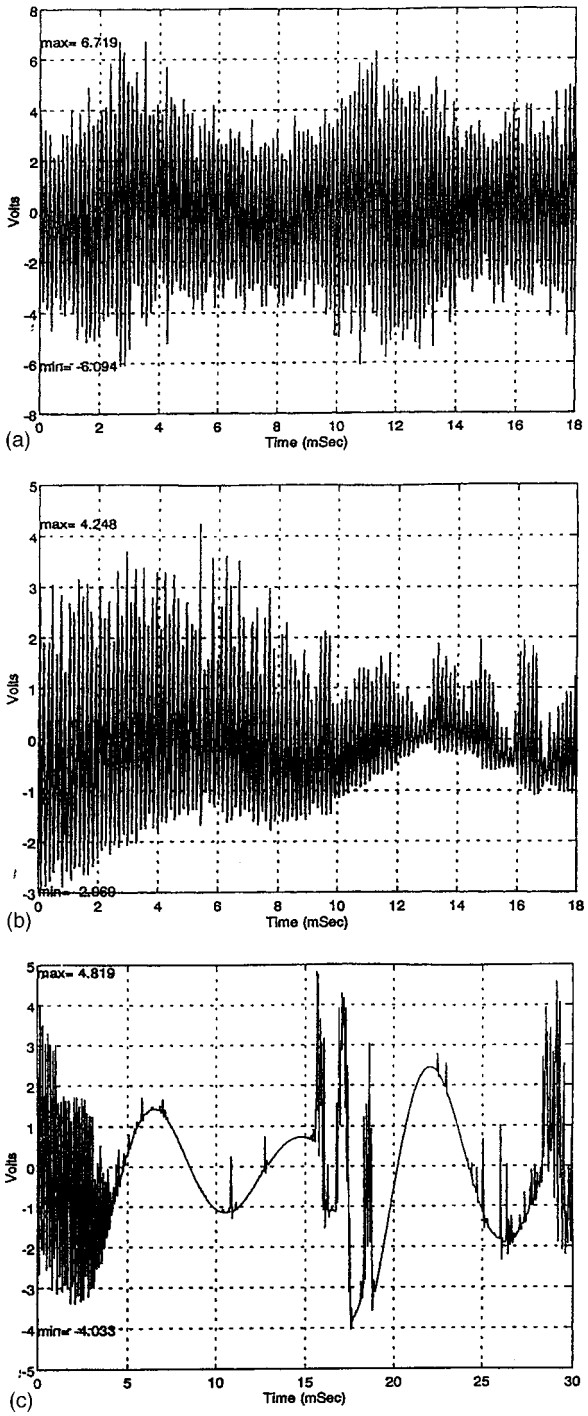


FIG. 3. Vibrometer time series (a) no surface waves, (b) with 2-mm surface waves, and (c) with 3-mm surface waves present.

A speculation of interest concerns the vibrometer device and its measurement approach. The laser beam is rather narrow. It travels 19.2 in. to the surface, is reflected, and returns to a small aperture. The alignment is sufficiently critical that a separate visible light source is directed at the surface to be interrogated, so that the range can be sighted in with two bullseyes that are to be coincident. The interrogated surface is expected to be a diffuse scatterer. In the water surface case, the laser return is specular and likely is lossy if the laser head is not perpendicular to the surface. The speculation is that the surface waves deflect the specular (narrow

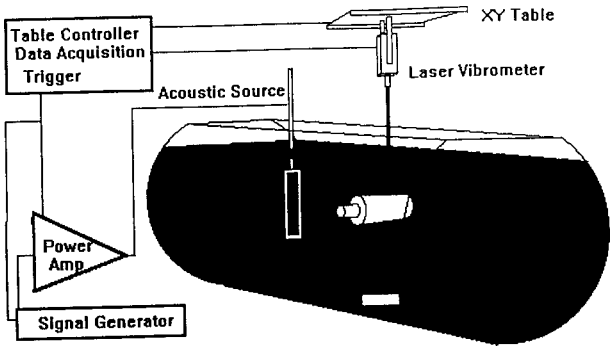


FIG. 4. AOH imaging experiment diagram.

beam) reflection out of the laser head receiver aperture. Another possibility is that the acoustic wave itself does not behave as expected in certain parts of the surface wave. The wave peak and the skirts adjacent to it may be examples of wave parts that contain acoustic signals, the behavior of which is poorly understood due to the dynamic and complicated boundaries.

D. AOH imaging experiment

An experiment was devised to learn if there were any hidden obstacles to the formation of high-quality beamformed images of near-field targets using surface vibrometry receive elements. A cylindrical test pool, approximately 3 ft in diameter by 4 ft in length was used for this experiment. A directive acoustic source was suspended at one end of the pool so that its main lobe ensonified a submerged target. The target was a 10-in. by 3-in., air-filled, anchored, plastic soda container. Target and source placements were made to keep multipath energy out of the receive time window as long as possible. The data acquisition was automated by a program in the experiment control computer. A synthetic aperture sonar approach to the data acquisition relieved the need for many vibrometers (one vibrometer was available). An $x-y$ positioning table was interfaced to the computer so that it could route the laser head around to 32 by 16 receive spots on the water surface. A transmission was initiated at each location followed by acquisition of an echo record from the vibrometer. The analog-to-digital converter was triggered by the same pulse that triggered the transmitter, making the system coherent. The x direction is arbitrarily defined as across the pool. The y direction is defined as along the axis of the pool cylinder (increasing range from the transmit element). The z direction is defined as depth. While the transmission was a linear frequency-modulated (LFM) pulse, swept from 63 to 73 kHz and of 0.11 ms duration, the sample rate was 250 kHz. Oversampling was considered to have a coherence maximizing influence on the system. The receive locations were separated by 1.150 cm (0.435 in.), which is half-wavelength spacing at 68 kHz, in both the x and the y directions. The resulting aperture was 34.25 cm in x by 16.57 cm in y (13.49 in. by 6.53 in.). The AOH imaging experiment arrangement is illustrated in Fig. 4.

Results of the AOH target imaging experiment must be evaluated in a subjective manner. The original data form a three-dimensional block of samples. Consequently, the beamformer output is a three-dimensional block of samples. Imaging experiment beamformed data are shown in Fig. 5 with some panels removed for better viewing of the peak panel. Figure 5 data were beamformed on a matched-filter processing strategy. This beamformer output can be subdivided into a sequence of two-dimensional sample planes or images. The beamformer generated 32 images in the y sequence. Each image was 64 points (in x) wide by 512 (time) samples deep. The 11th image contained the data block peak value. An inverse-filtering beamformer was convolved with the original raw data block. A portion of the resulting output peak panel, bearing the 25.4-cm by 7.62-cm container signature, is reproduced as a magnitude raised to the 3.5 power plot in Fig. 6.

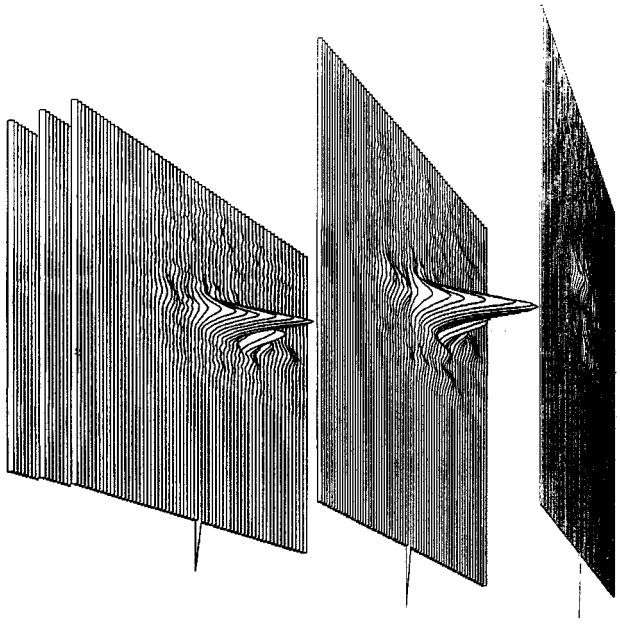


FIG. 5. Matched-filter beamformer output.

III. CONCLUSIONS

AOH feasibility is established. Acoustic signals of practical power levels can be retrieved from a calm water surface with laser vibrometry.

Capillary and gravity waves of very modest height sharply reduce the SNR of this measurement technique. A

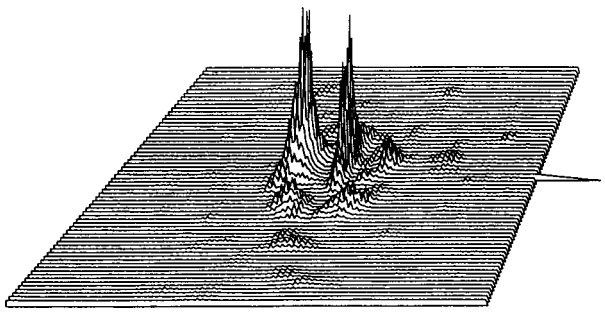


FIG. 6. Peak panel of inverse-filtering beamformer shown as magnitude to the 3.5 power. The x direction, which is cross pool, proceeds from back to front. The z, or depth direction, proceeds from the surface at left to deeper depths at the right.

method for mitigation of this effect must be found before a practical system can be fielded for ocean use.

The surface wave problem is not merely the superposition of unwanted surface wave signatures and desired acoustic signal. A temporal disappearance of the acoustic signal is associated with the surface waves. The acoustic signal absence is of a duration that is directly related to the surface wave height.

High-quality acoustic target images can be acquired from a wave-free surface. Three-dimensional synthetic aperture techniques can be combined with the AOH methods to produce highly resolved target images. The resolution arises from the proximity of the aperture to the target. The receive range is equal to the target depth. In shallow or very shallow water, this range becomes short and productive of high resolution.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to Mr. James Kennedy of the Red River Army Depot for his generous efforts in obtaining the equipment necessary for these experiments. Likewise, thanks are extended to Dr. John Jellinek of the Tank Automotive Command in Warren, MI, who authorized the equipment loan. This work was sponsored by Internal Research funding from Coastal Systems Station Dahlgren Division, Naval Surface Warfare Center, Panama City, FL.

[1] H. Yeh and H. Z. Cummins, ''Localized Fluid Flow Measurements with a He–Ne Laser Spectrometer,'' Appl. Phys. Lett. **4**, 176–178 (1964).
[2] K. J. Taylor, ''Absolute Measurement of Acoustic Particle Velocity,'' J. Acoust. Soc. Am. **59**, 691–694 (1976).
[3] K. J. Taylor, ''Absolute Calibration of Microphones by a Laser Doppler Technique,'' J. Acoust. Soc. Am. **70**, 939–945 (1981).
[4] J. Vignola, Y. H. Berthelot, and J. Jarzynski, ''Laser Detection of Sound,'' J. Acoust. Soc. Am. **90**, 1275–1286 (1991).
[5] R. J. Urick, *Principles of Underwater Sound* (McGraw-Hill, New York, 1975).
[6] TSI manufacturer's specification for Laser Vibrometer, models 1930 and 1940 (1988).

Acoustic impedances at the oval window, and sound pressure transformation of the middle ear in Norwegian cattle

Magne Kringlebotn
Department of Physics, Norwegian University of Science and Technology, N 7491 Trondheim, Norway

(Received 7 February 2000; revised 3 May 2000; accepted 18 May 2000)

In 15 cadaver ears from Norwegian cattle, sound pressure transfer functions have been measured (1) for sound input to the tympanic membrane, (2) for sound input to the oval window with the footplate in place, but with the ossicular chain removed, and (3) for sound input to the oval window with also the footplate removed. The output pressure was measured in an enclosure cemented to the round window. The data allow calculation of equivalent sound pressures at the input positions, as well as the acoustic input impedances at the oval window with intact footplate, Z_{sc}, and with the footplate removed, Z_c. The difference $Z_s = Z_{sc} - Z_c$ is the acoustic impedance contribution of the footplate and annular ligament. Z_{sc} is mainly determined by the stiffness of the annular ligament at low frequencies, and by the cochlear input impedance Z_c at higher frequencies. Z_c is predominately resistive, a minor reactive part at low frequencies is attributed to the stiffness of the round window membrane. Z_s and Z_c are equal in magnitude at about 0.4 kHz. Rather close RLC fits have been obtained for all the three impedances, Z_{sc}, Z_s, and Z_c. The fitted values for the resistive parts of Z_{sc} and Z_c are 62.9 and 58.2 acoustic GΩ, respectively. The relatively small difference, 4.7 GΩ, is attributed to the resistance of the annular ligament. The fitted resistance of Z_s is somewhat larger, 8.6 GΩ, but is anyway of minor importance relative to the dynamic stiffness of the annular ligament. This stiffness depends on the static pressure difference across the footplate. Each of the averaged Z_{sc} corresponds to minimum stiffness. The fitted acoustic compliance is 6.89×10^{-15} m^3/Pa. The acoustic inertance plays a minor role. It is attributed to the mass of the footplate and the co-vibrating liquid in the inner ear, and has a fitted value of 4.7×10^5 Pa s^2/m^3. A sound pressure at the eardrum is equivalent to a larger pressure at the footplate, about 16 dB larger at frequencies below 100 Hz, increasing to about 30 dB at 10 kHz. In the vestibulum at the inner side of the footplate, the sound pressure at 20 Hz is about 20 dB below the equivalent pressure at the outer side. The two pressures approach toward higher frequencies, and above 1 kHz they are nearly equal. © *2000 Acoustical Society of America.* [S0001-4966(00)00709-8]

PACS numbers: 43.64.Bt, 43.64.Kc, 43.58.Bh [BLM]

INTRODUCTION

Knowledge of the acoustic input impedance at the oval window is required when establishing or modifying models for the ear, and may also be useful for physicians when analyzing surgical measures in the middle ear. A thorough explanation of the significance of this impedance has been given by Merchant *et al.* (1996). In the present study, the impedance at the oval window will be measured with the footplate in place, as well as with the footplate removed. The difference between the two impedances is the impedance contribution of the annular ligament and the footplate.

It is well known that the middle ear acts as a sound pressure transformer: the sound pressure at the eardrum is transformed to a higher equivalent sound pressure at the stapes. Sound pressures are equivalent if they cause identical volume displacements in the oval window. Contributing factors to the sound pressure increase from eardrum to stapes are the lever ratio of the rotating ossicles, and the larger area of the tympanic membrane as compared to the footplate. Equivalent sound pressures will be measured at the eardrum, the stapes, and at the entrance of the oval window after the stapes has been removed. Knowledge of the sound pressure transformation in the middle ear is important for the same reasons as mentioned above for the acoustic impedance at the oval window.

METHOD

Applying cement to the ear bone

To insure attachment of the cement to the bone, soft tissue is removed from the contact area, which is then carefully dried with paper towel. To improve attachment, some Loctite 401 (Manufacturer: Loctite Corporation, Dublin, Ireland) is applied. Plastic padding elastic (Manufacturer: Loctite Sweden AB, Gøteborg, Sweden) is used as a cement. It hardens in about 10 min, or even faster if a larger than normal amount of hardener is used. To avoid leakage from the perilymphatic duct and the inner ear meatus, the surrounding tissue and bone are carefully dried and Loctite 401 is applied before the openings are coated with cement.

Measuring the sound pressure transfer functions in the ear

The measurements are made on 15 cadaver ears from NFR, which is a Norwegian cattle breed. Animal sex, age,

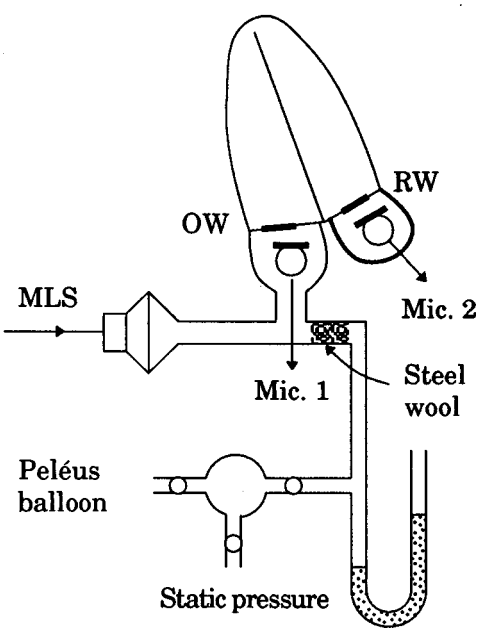


FIG. 1. Experimental setup for measuring the acoustic input impedance at the oval window, and sound pressures ratios in the ear.

and weight are unspecified. The ears have been kept moist at a temperature just above 0 °C, and are measured at room temperature 1–11 days after slaughtering.

The output sound pressure is measured in a chamber formed by closing the bony entrance to the round window with cement and a microphone adapter made of aluminum. Access to the chamber is through a 1 mm diameter hole in the adapter. The microphone input tube fits exactly into this hole, so that the sound pressure in the air filled chamber is exposed to the microphone.

The experimental setup when the input sound pressure is applied at the tympanic membrane is described elsewhere (Kringlebotn, 2000a). An identical setup, except that the sound input is now to the oval window, is shown schematically in Fig. 1. A four-armed glass tube is cemented to the oval window. Microphone 1 is directed toward the oval window through one arm of the glass tube and measures the input sound pressure about 2–4 mm in front of the oval window. This tube arm has an inner diameter of 4 mm. The microphone cable fits airtight into a second arm not shown in Fig. 1. A third arm is connected to a loudspeaker, and the fourth to a manometer. The entrance to this arm is filled with

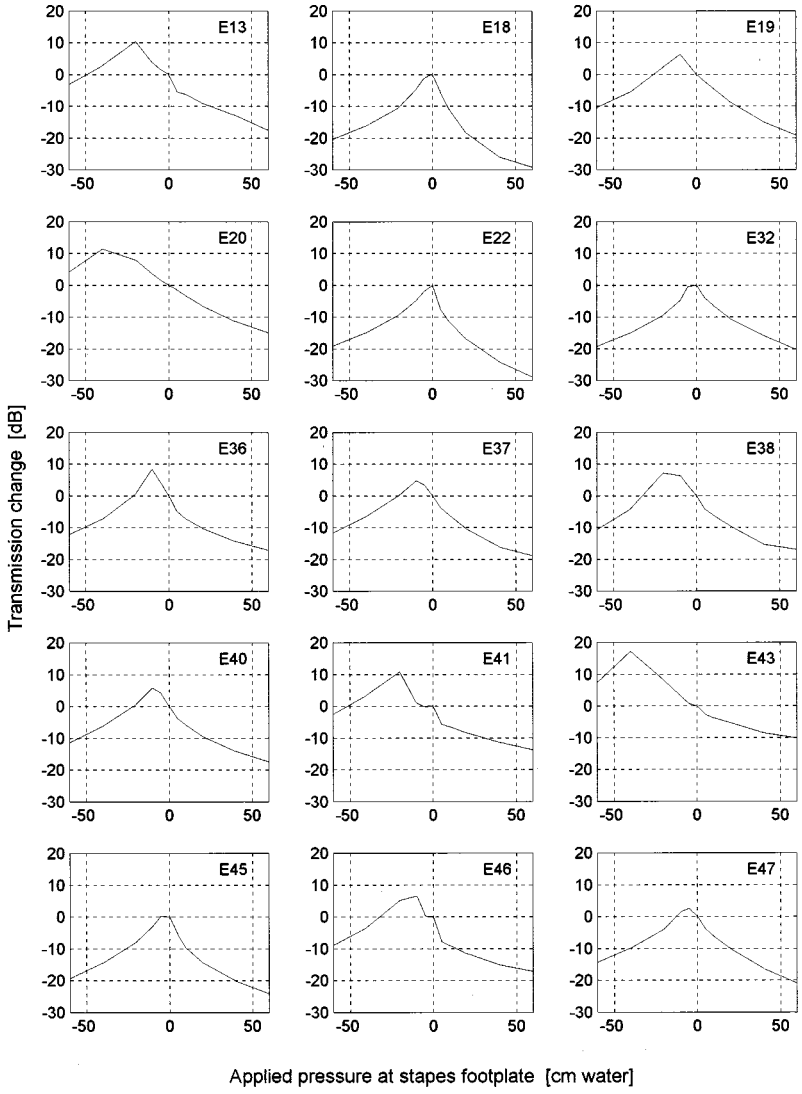


FIG. 2. Change in sound transmission through the oval window for each of 15 cattle ears, when various static pressures are applied at the footplate. The ossicular chain is disconnected. Measuring frequency: 96 Hz.

steel wool in order to reduce the effect of standing waves. The microphones are small electret microphones with built in amplifiers (Knowles Type 3046).

A computer equipped with a suitable soundcard (Fiji from Turtle Beach Systems, Inc.), and software (WinMLS, Morset Sound Development, Trondheim, Norway), for generating and analyzing MLS signals is used to derive the inner ear transfer function, from microphone 1 to microphone 2. The MLS signal from the computer is fed to a loudspeaker, via an equalizer and an amplifier, in order to optimize the signal to noise ratio, by adjusting the equalizer so that an approximately flat spectrum is obtained from microphone 1. The microphone signals are fed via measuring amplifiers to the two line inputs of the sound card. The static pressure in the inner ear may differ from the outside atmospheric pressure. It is reasonable to assume that the footplate moves most easily when the static pressure difference across it is zero. The acoustic impedance at the oval window will be reported for this condition.

The static pressure difference across the footplate is varied by varying the static pressure in the enclosure cemented to the oval window, by means of a Peléus balloon (a small rubber balloon having three inlet/outlet sphere valves). The opening of the tube connection to the Peléus balloon and the manometer is filled with steel wool, in order to reduce the effect of standing waves.

The volume of the enclosure around the round window membrane is of the order of 30 mm^3. When approximated by a sphere, the diameter will be of the order 4 mm and equal to a quarter of a wavelength at about 20 kHz. In air at 0 °C the density is $\rho_0 = 1.293\,kg/m^3$ and the velocity of sound $c_0 = 331.3\,m/s$. The acoustic compliance for the enclosed air volume V at standard atmospheric pressure and arbitrary temperature is:

$$C = \frac{V}{\rho_0 c_0^2}, \tag{1}$$

corresponding to an acoustic impedance, as seen by the round window membrane:

$$Z_e = \frac{1}{j\omega C}, \tag{2}$$

where ω is the angular frequency. The volume V of the round window enclosure is found at the end of the experimental session by filling it with water from a small syringe, and weighing the corresponding mass loss of the syringe.

The sound volume velocities in the cochlear windows are equal (Kringlebotn, 1995; Voss *et al.*, 1996). For a volume velocity U, the sound pressure in the round window enclosure is:

$$p_e = Z_e U. \tag{3}$$

If only the footplate is left in the oval window, the input acoustic impedance is denoted by Z_{sc}, and when also the footplate is removed by Z_c. For the same volume velocity U, the corresponding sound pressures at the oval window will be:

$$p_s = Z_{sc} U, \tag{4}$$

$$p_v = Z_c U. \tag{5}$$

The sound pressure at the eardrum is denoted by p_d, and the following transfer functions will be measured:

$$H_1 = \frac{p_e}{p_d}, \tag{6}$$

$$H_2 = \frac{p_e}{p_s}, \tag{7}$$

$$H_3 = \frac{p_e}{p_v}. \tag{8}$$

The measurements of H_1 have been reported earlier (Kringlebotn, 2000a). They were made with sound input to the tympanic membrane, and with an intact ossicular chain and an open middle ear.

H_2 is measured for the same ears, with sound input to the oval window and with an intact footplate. The ossicular chain and the outer ear are first removed. The incudostapedial joint is disconnected by means of a sharp, pointed tool. The ear canal and the outer part of the middle ear with the tympanic membrane, malleus, and incus are removed with a single cut of a band saw. A direct view to the oval window is then obtained. The stapes crura are easily broken and removed by means of a pair of tweezers, without any risk of injuring the annular ligament. The annular ligament is in fact very robust. In average, it may withstand a static pressure of up to 29.4 atm at the footplate (Kringlebotn, 2000b).

H_3 is measured after the footplate has been removed. The glass tube to the oval window is disconnected by means of a scalpel, by removing a sufficient amount of cement. The footplate and the annular ligament may then be removed. This is done by carefully pressing the sharp, pointed tool through the footplate, which is broken and removed bit by bit. In this operation a small amount of perilymph escapes and is replaced by isotonic NaCl solution, without introducing air bubbles in the inner ear. After adjusting the liquid level in the oval window, the glass tube connection to the oval window is re-established by applying new cement to compensate for what was removed.

The differences in frequency characteristics for the two measuring channels, including the microphones, were regularly measured and corrected for by exposing the microphones to the same sound pressure in a small coupler with a condenser microphone as a sound source.

From Eqs. (1) to (8), the acoustic input impedances may now be determined:

$$Z_{sc} = \frac{Z_e}{H_2}, \tag{9}$$

$$Z_c = \frac{Z_e}{H_3}, \tag{10}$$

$$Z_s = Z_{sc} - Z_c. \tag{11}$$

The sound pressure transformation from the eardrum to the footplate will be given by:

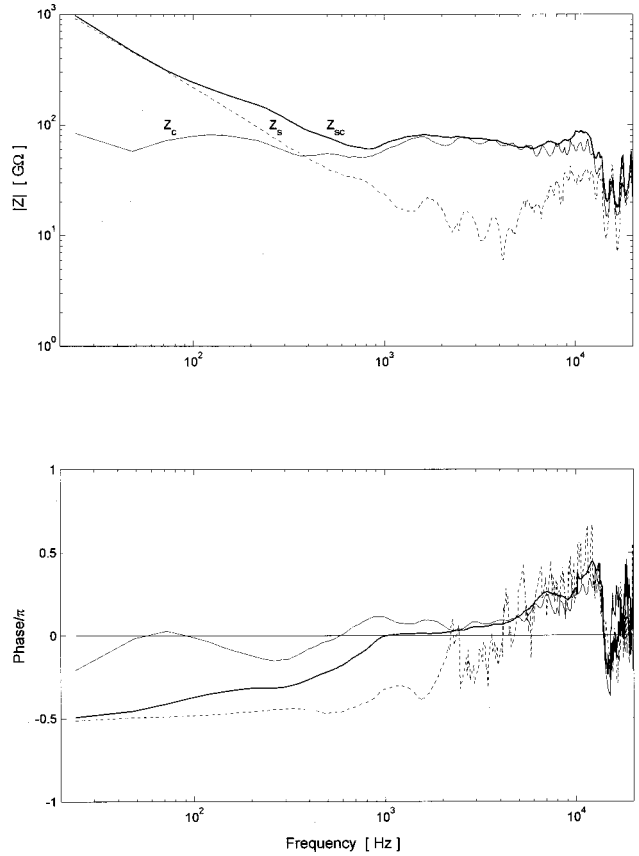


FIG. 3. Acoustic impedances at the oval window. Mean for 15 cattle ears. Z_{sc} : input impedance with footplate in place. Z_c : input impedance with footplate and annular ligament removed. Z_s : impedance contribution of footplate and annular ligament.

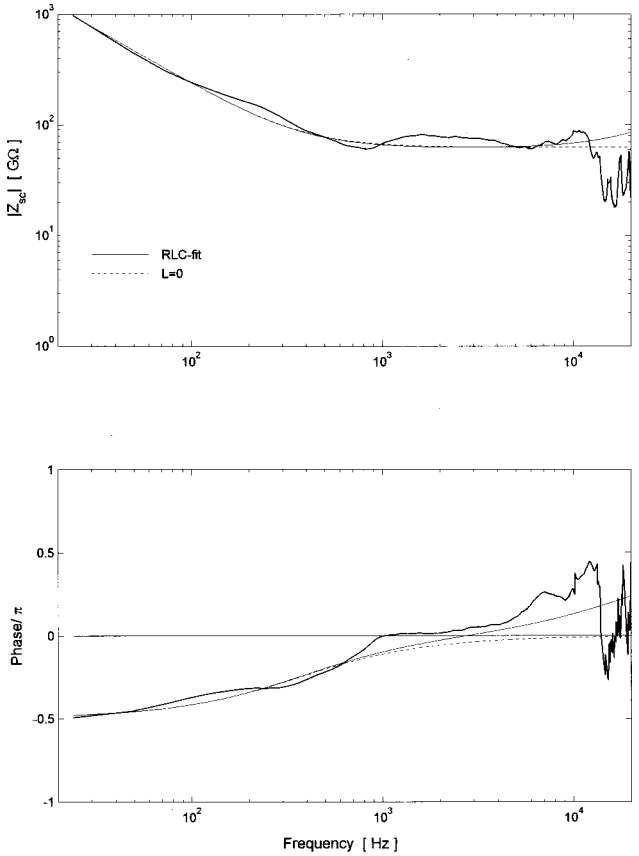


FIG. 4. RLC fit to the oval window impedance with intact footplate, Z_{sc}. $R=6.29\times10^{10}$ Pa s/m^3, $L=4.68\times10^5$ Pa s^2/m^3, $C=6.89\times10^{-15}$ m^3/Pa.

$$\frac{p_s}{p_d}=\frac{H_1}{H_2}, \tag{12}$$

and from the eardrum to the vestibulum at the inner side of the footplate by:

$$\frac{p_v}{p_d}=\frac{H_1}{H_3}, \tag{13}$$

and finally, from the outer to the inner side of the footplate by:

$$\frac{p_v}{p_s}=\frac{H_2}{H_3}. \tag{14}$$

RESULTS

For 15 ears and at a frequency 96 Hz, the change in the sound pressure transfer function $H_2=p_e/p_s$ with varying static pressure at the footplate is shown in Fig. 2. At this frequency, the sound transmission is determined by the dynamic stiffness of the annular ligament. The transfer functions H_2 that correspond to minimum stiffness, i.e., maximum sound transmission through the oval window, have been used to calculate acoustic impedances and sound pressure relationships.

The acoustic impedances at the oval window are calculated from Eqs. (9) to (11). Averaged results are shown in Fig. 3, and RLC fits to the mean impedances are in Figs. 4–6. Mean impedances±standard deviations are compared to data from other investigations in Figs. 7–11. The impedance magnitudes are given in acoustic GΩ, $1\ \mathrm{G\Omega}=10^9\ \mathrm{N\,s/m^5}$.

The ratios between equivalent sound pressures in the ear are calculated from Eqs. (12) to (14). Averaged results are shown in Fig. 12. Ratios±standard deviations are shown in Figs. 13–16, and have been compared to data from other investigations.

All mean data were obtained by geometric averaging of the magnitudes and arithmetic averaging of the phases.

For one ear (E45), the impedance dependence on static pressure at the footplate is shown in Fig. 17.

DISCUSSION

Measurement of the transfer function $H_1=p_e/p_d$ is reported elsewhere (Kringlebotn, 2000a). Fifteen of the ears were the same as those used in the present study. With an intact ossicular chain, it was shown that a static pressure across the footplate had only a minor influence on the transfer function H_1. Static pressures in the range 0–60 cm water caused mean sound transmission losses less than 5 dB below 1 kHz, and negligible losses at higher frequencies.

A static pressure across the footplate is accompanied with a static displacement of the stapes and an increase in the dynamic stiffness of the annular ligament. It is to be expected that the stiffness increase will be smaller when the displacement of the stapes is obstructed by the ossicular

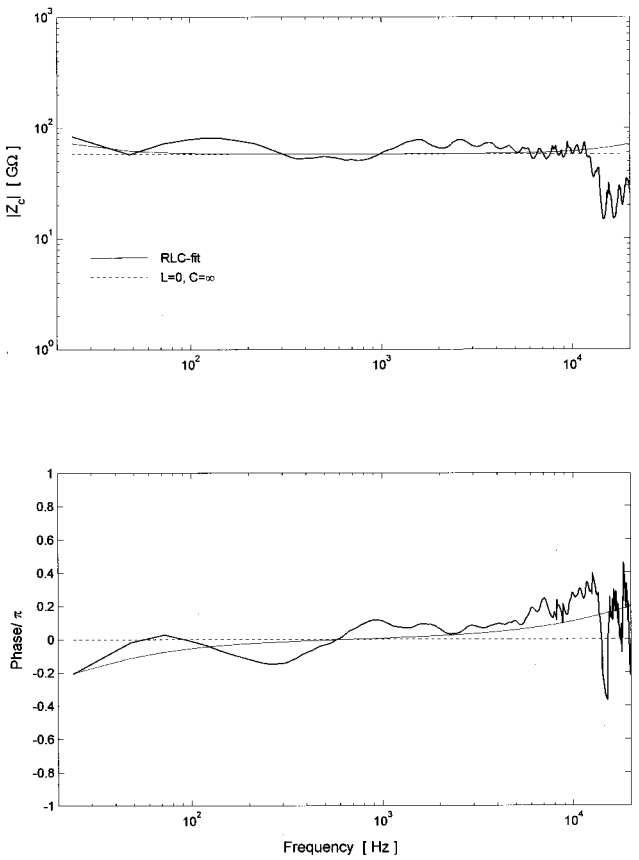


FIG. 5. RLC fit to the oval window impedance, when the footplate and annular ligament have been removed, Z_c. $R=5.82\times10^{10}$ Pa s/m^3, $L=3.28\times10^5$ Pa s^2/m^3, $C=1.52\times10^{-13}$ m^3/Pa.

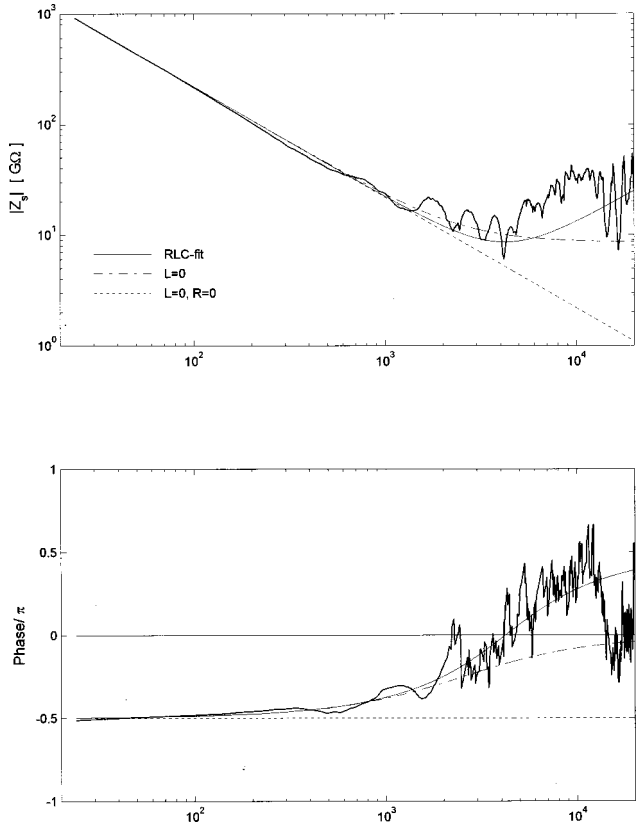


FIG. 6. RLC fit to the footplate and annular ligament contribution to the oval window impedance, Z_s. $R=8.61\times10^9$ Pa s/m^3, $L=1.96\times10^5$ Pa s^2/m^3, $C=7.27\times10^{-15}$ m^3/Pa.

chain, than when the chain is removed so that the stapes moves more freely. This expectation is confirmed by the experimental findings. When the ossicular chain is removed, Fig. 2 shows that the transfer function $H_2=p_e/p_s$ depends strongly on the static pressure at the footplate. Maximum in transmission occurs at zero or negative pressures. A possible explanation is that liquid slowly escapes from the inner ear by penetrating the bone. The footplate and the round window membrane will then move inward to account for the reduction in liquid volume. When a sufficient underpressure is applied at the oval window, the footplate will move back to its optimal position, corresponding to zero pressure across the footplate. Simultaneously, the round window membrane moves further inward, but the stiffness changes here are believed to be of little significance. This explanation seems to be inconsistent with the data of Lynch *et al.* (1982), for which a maximum in transmission occurs at a positive pressure at the oval window. However, the placement of a transducer in the inner ear may have pressed the footplate and the round window membrane outward in this case. The observed result is then to be expected.

From Fig. 4 it is seen that at low frequencies, the acoustic impedance at the oval window with intact footplate, Z_{sc}, is stiffness controlled. The stapes moves most easily for minimum stiffness, i.e., for maximum H_2 in Fig. 2. Z_{sc} is reported for this condition. The applied static pressure at the stapes is then assumed to equal the inner ear pressure. When also the footplate is removed, such a pressure effect will of course be absent, as the inner ear pressure then will equal the atmospheric pressure.

Mean results for the acoustic impedances Z_{sc}, Z_c and $Z_s=Z_{sc}-Z_c$ are shown in Fig. 3. In Figs. 4–6, the mean impedances are fitted rather nicely by simple series RLC circuits.

The impedance contribution of the annular ligament and the footplate, Z_s, is obtained simply by subtraction. At lower frequencies, Z_s gives the dominating contribution to Z_{sc}, and is stiffness controlled. This is also seen from the phase, which approach $-\pi/2$ at the lowest frequencies.

Z_{sc} and Z_c are the results of completely independent measurements. At frequencies above 1 kHz, the input impedance at the oval window with intact footplate, Z_{sc}, very nearly equals the input impedance of the cochlea, Z_c. Z_s contributes negligibly to Z_{sc} in this frequency range. A possible exception is the high frequency range above about 13 kHz. But the results here are considered to be uncertain, due to the more or less unavoidable distances between the microphone acoustic inputs and the optimal measuring positions near the oval and round windows. Figures 4 and 5 show that the results above 13 kHz are inconsistent with the RLC representation of Z_{sc} and Z_c in this frequency range. The Z_c data have a fine structure that is not reflected in the Z_{sc} data; see Fig. 3. The reason for this disagreement is unknown.

Figure 4 demonstrates that in the fitted impedance for Z_{sc}, the inertance L is of minor importance, but contributes to a somewhat better fit than a pure RC circuit. Figure 5

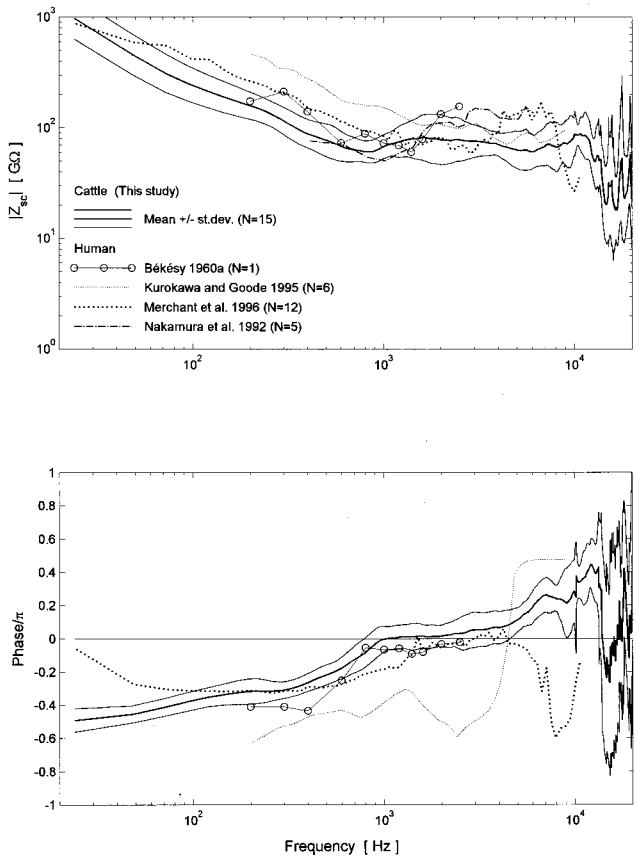


FIG. 7. Oval window impedance with stapes footplate in place, Z_{sc}. Comparison with data for human ears.

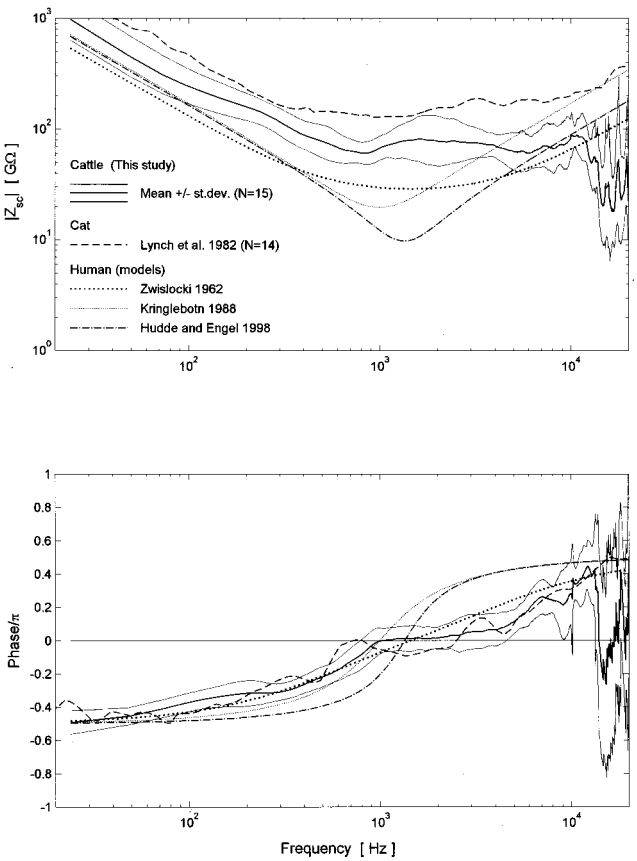


FIG. 8. Oval window impedance with stapes footplate in place, Z_{sc}. Comparison with human ear models and data for cat ears.

demonstrates that the cochlear input impedance, Z_c, is predominately resistive. Inclusion of C and L is seen to slightly improve the curve fit, especially for the phase. If the reactive part at low frequencies is attributed to the stiffness of the round window membrane, the fitted value for this stiffness (inverse of compliance) is $K_c = 6.6 \times 10^{12}\,\mathrm{Pa/m^3}$, in fair agreement with the value $7.5 \times 10^{12}\,\mathrm{Pa/m^3}$, obtained from the difference in the fitted stiffnesses in Z_{sc} and $Z_s : K_{sc} = 1.451 \times 10^{14}\,\mathrm{Pa/m^3}$ and $K_s = 1.376 \times 10^{14}\,\mathrm{Pa/m^3}$ ($K_c = K_{sc} - K_s$). It is seen that K_c contributes only about 5% to K_{sc}.

Excitation of the hair cells is determined by the movement of the cochlear partition. For a given sound pressure input to the cochlea, this movement is only, to a small extent, influenced by the stiffness of the round window at the lowest frequencies, and by the liquid mass load at the footplate at the highest frequencies.

Figure 6 clearly demonstrates that the impedance contribution due to the annular ligament (and the footplate), Z_s, is stiffness controlled up to at least 1 kHz. The excellent fit toward low frequencies leaves no doubt that the impedance contribution of the annular ligament is really stiffness controlled at these frequencies, and that the annular resistance plays only a minor role. As Z_c is shown to be predominately resistive, it is then reasonable to believe that the low frequency phase asymptote for Z_{sc} should be $-\pi/2$, not only for the cattle ear, but also for the human ear. This asymptote is approached when the magnitude of Z_c becomes small as compared to the magnitude of Z_s.

With footplate area $A = 2.3\,\mathrm{mm^2}$ (Kringlebotn, 2000a), $L = 4.68 \times 10^5\,\mathrm{Pa\,s^2/m^3}$ in Z_{sc} corresponds to a vibrating mass $m = 2.48\,\mathrm{mg}$. The footplate radiates sound into the perilymph, and the footplate dimensions are small compared to the wavelength in the whole audio range. For a circular piston with the same area in an infinite baffle the radiation impedance may then be represented by adding a radiation mass to the piston mass (Kinsler and Frey, 1962): $m_r = \rho_0 a^3 8/3 = 1.73\,\mathrm{mg}$. Here $a = (A/\pi)^{1/2} = 0.856\,\mathrm{mm}$ is the piston radius and $\rho_0 = 1034\,\mathrm{kg/m^3}$ (Békésy, 1960a) is the density of the perilymph. This estimate for the radiation mass happens to be in excellent correspondence with the fitted $L = 3.28 \times 10^5\,\mathrm{Pa\,s^2/m^3}$ in the cochlear impedance $Z_c : m_r = LA^2 = 1.74\,\mathrm{mg}$. As is to be expected, the footplate mass obtained from $m_f = m - m_r = 0.75\,\mathrm{mg}$ is in fair correspondence with the fitted $L = 1.96 \times 10^5\,\mathrm{Pa\,s^2/m^3}$ in $Z_s : m_f = LA^2 = 1\,\mathrm{mg}$. According to a single measurement the stapes mass is about 3 mg, so a footplate mass of about 1 mg seems quite reasonable.

In Figs. 7 and 8, the input acoustic impedance at the oval window with intact footplate, Z_{sc}, is compared to results from other experiments with human ears and cat ears, and to predictions from models of the human ear. The similarities in the experimental impedance data are more striking than the differences. Some of the results have been thoroughly discussed by Merchant *et al.* (1996). The magnitude data for cattle ears seem to be fairly representative also for human ears, except that the Kurokawa and Goode magnitude data

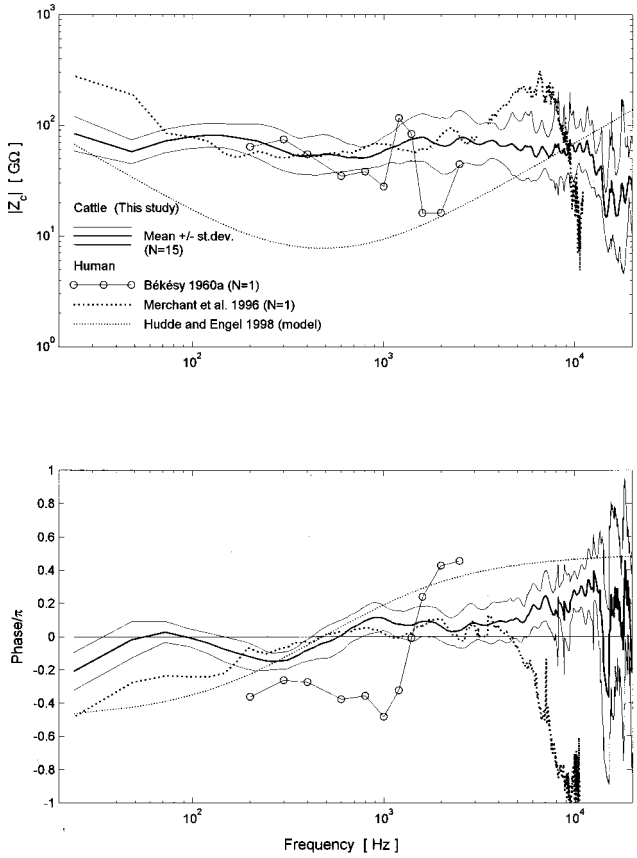


FIG. 9. Oval window impedance with the stapes footplate and annular ligament removed, Z_c. Comparison with data for human ears.

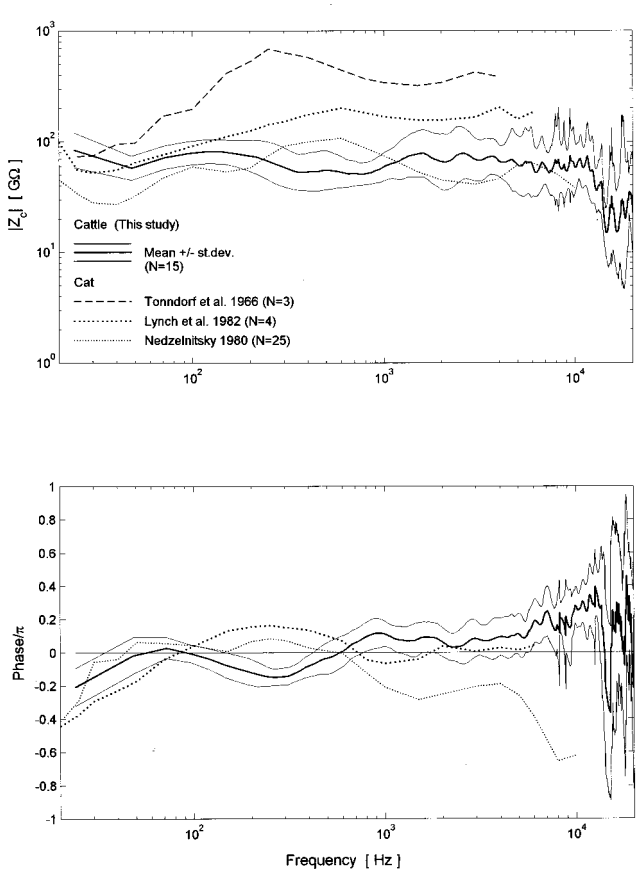


FIG. 10. Oval window impedance with the stapes footplate and annular ligament removed, Z_c. Comparison with data for cat ears.

are 6–10 dB larger at frequencies below 1 kHz. The fitted acoustic resistance of Merchant *et al.* (1996), 72 GΩ for human ears, is in approximate agreement with the fitted value for cattle ears, 63 GΩ. Relative to the results for cattle ears, it is seen that the impedance magnitudes for cat ears (Lynch *et al.*, 1982) are generally larger but with nearly the same frequency dependence.

The phase curve for cattle ears agrees well with the curve for cat ears and is believed to be rather representative also for human ears. This belief is based on the following assumptions. The phase data of Kurokawa and Goode (1995) are dubious as they markedly deviate from the rest of the phase data. The phase data of Merchant *et al.* (1996) should approach $-\pi/2$ at the lowest frequencies. At the highest frequencies the phase is expected to be positive due to the mass of the stapes and the co-vibrating fluid. In the cattle ears only the footplate was left in the oval window. The inertance is therefore somewhat smaller than for intact stapes.

For one ear (E45), Fig. 17 shows how Z_{sc} depends on the static pressure at the footplate. The frequency range in which Z_{sc} is stiffness controlled, increases with increasing static pressure across the footplate. At higher frequencies, Z_{sc} is little affected. Wether the results for Z_{sc} in previous investigations have been influenced by inner ear pressures is not known. If they have, it is to be expected that the mean impedance magnitudes for zero pressure across the footplate will be lower than reported values in the low frequency range where the impedance is stiffness controlled. In the present study, this reduction is about 6 dB, as may be deduced from Fig. 2. If the fitted acoustic stiffness for 12 human ears (Merchant *et al.*, 1996) is reduced by the same amount, i.e., from 3.26×10^{14} to 1.63×10^{14} Pa/m^3, the result will be in approximate agreement with the fitted value for cattle ears, 1.45×10^{14} Pa/m^3. For one human ear and with the inner ear drained (Merchant *et al.*, 1996), the fitted stiffness was 2.19×10^{14} Pa/m^3.

The human ear models obviously give to small magnitudes for Z_{sc}, except at the highest frequencies. But the phase curve according to Zwislocki's model is in fair agreement with the phase curve for cattle ears as well as for cat ears.

Data for the input acoustic impedance of the cochlea, Z_c, are scant. Figure 9 shows the results for one human ear according to Békésy (1960c). The magnitude and phase curves are rather irregular, but the magnitudes are of the same order as for cattle ears. The human ear data of Merchant *et al.* (1996) were obtained as the difference $Z_{sc}-Z_s$, where Z_s is for one ear and Z_{sc} is the average for 12 ears. In the frequency range from 100 to 3000 Hz these data agree well with the results for cattle ears. But again, at higher frequencies the phase is expected to be positive. The input impedance of the cochlea has also been discussed by Zwislocki (1975). He argues that it should be resistive and have twice the value estimated from Békésy's data, i.e., 35 GΩ instead of 17.6 GΩ. However, both values are probably too small. The estimated value was a mean value for the frequen-

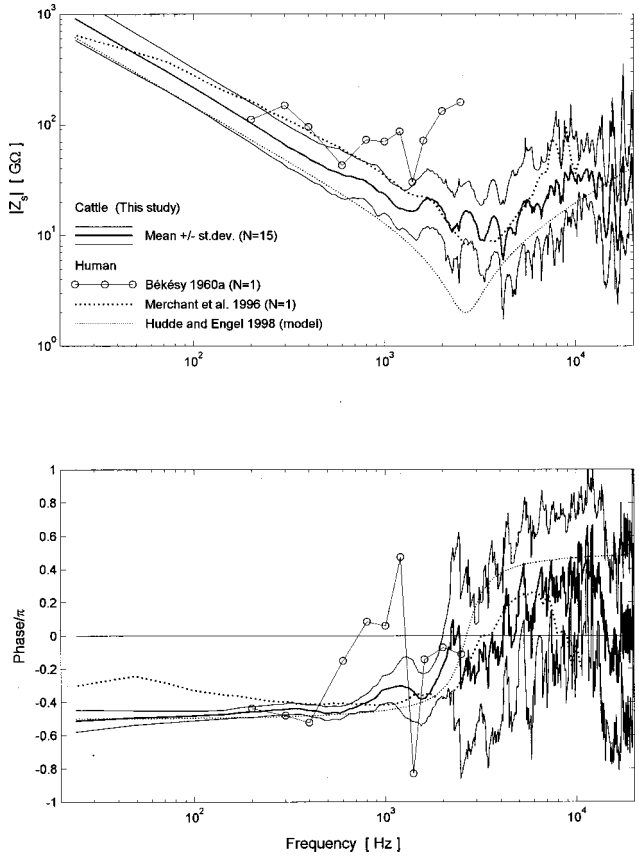


FIG. 11. Contribution to the oval window impedance, Z_s, from the footplate and annular ligament. Comparison with data for human ears.

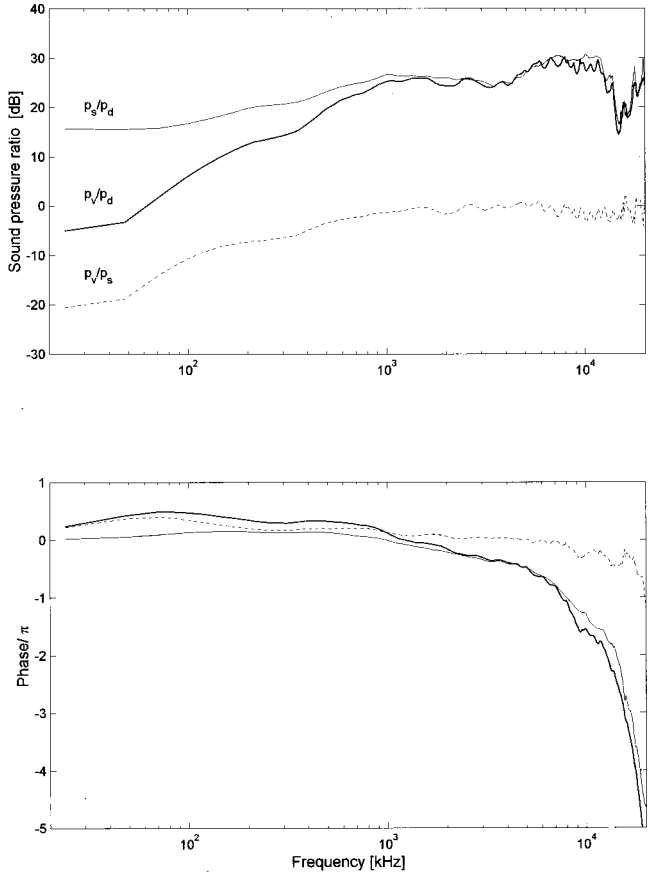


FIG. 12. Ratios of equivalent sound pressures in the ear. Mean for 15 ears. p_s/p_d: sound pressure at the footplate relative to at the eardrum. p_v/p_d: sound pressure in the vestibulum, at the inner side of the footplate, relative to at the eardrum. p_v/p_s: sound pressure at the inner side of the footplate relative to at the outer side.

cies 200, 300, and 400 Hz, but must erroneously have been obtained from a magnitude curve labeled ''Fluid of cochlea alone'' (Békésy, 1960c). The correct curve labeled ''Fluid of cochlea+membrane of the round window'' is presented in Fig. 9. It is seen that the mean value for 200, 300, and 400 Hz happens to be very close to the mean value for cattle ears as well as for the human ear according to Merchant *et al.* (1996). For cat ears, Fig. 10 demonstrates quite large disagreement between data from different investigations.

It is seen from Fig. 11 that the impedance contribution from the footplate and the annular ligament, Z_s, agrees quite well with data for one human ear according to Merchant *et al.* (1996), and also with values predicted by the human ear model of Hudde and Engel (1998). The magnitude curve for cattle ears seems to be a reasonable compromise between the other two. The human ear data of Békésy (1960c) are obtained as the differences $Z_{sc}-Z_c$. They are quite irregular, and some of the data deviate markedly from the frequency characteristics for cattle ears.

Figure 12 shows relationships between equivalent sound pressures in the ear. A sound pressure at the eardrum is equivalent to a larger pressure at the footplate. The pressure ratio is about 16 dB at frequencies below 100 Hz, increasing to about 30 dB at 10 kHz. In the vestibulum at the inner side of the footplate, the sound pressure at 20 Hz is about 20 dB below the equivalent pressure at the outer side, but increases by about 6 dB/octave up to 1 kHz. At higher frequencies, the two pressures are nearly equal. This is consistent with the statement above, that the impedances Z_{sc} and Z_c are nearly equal above 1 kHz.

Figure 13 shows the ratio between the equivalent sound pressures at the footplate and the eardrum, p_s/p_d. Below 1 kHz, the two pressures are nearly in phase. In the frequency range 100–2000 Hz the magnitude data agrees well with data obtained for human ears. It should be noted that the Békésy data were obtained for immobilized stapes. At higher frequencies the ratio for cattle ears remains large, but falls off rather steeply for human ears. This difference is related to the difference in transfer functions from sound pressure at the eardrum to volume displacement in the cochlear windows. The slope in transfer functions above 1 kHz is found to be about −15 dB/octave for human ears (Kringlebotn and Gundersen, 1985) and −5 dB/octave for cattle ears (Kringlebotn, 2000a). The reason for this large discrepancy should be further investigated.

At frequencies above 1 kHz, the equivalent sound pressures at the two sides of the footplate are expected to be nearly equal. In this frequency range, the pressure ratios in Figs. 13–15 should be comparable, and it is seen that the discrepancies in ratio magnitudes are larger between cattle ears and human ears than between cattle ears and other animal ears. A striking difference in the magnitude data for cat ears in Fig. 15, is the notch at about 3 kHz in the Décory

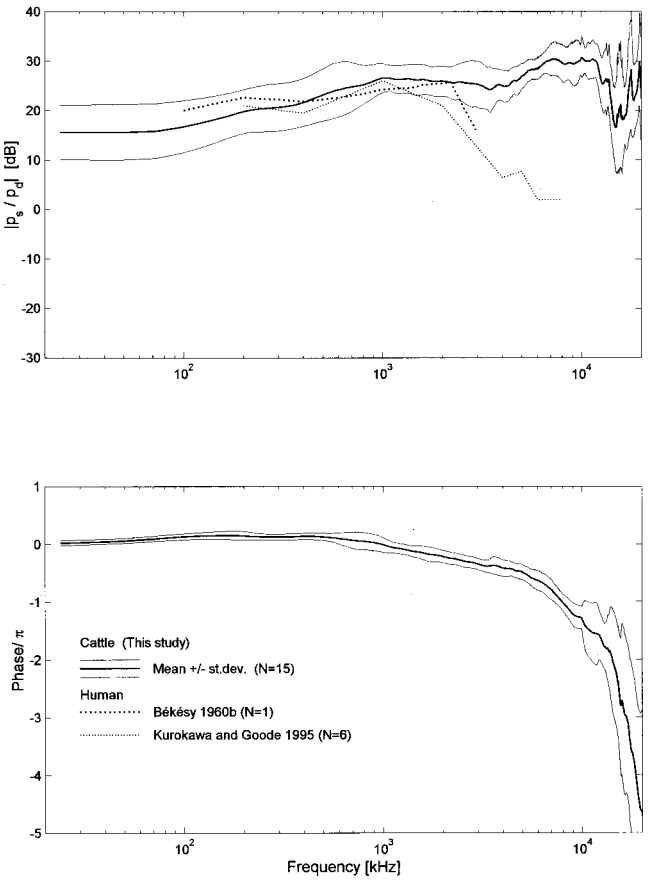


FIG. 13. Ratios of equivalent sound pressures in the ear. p_s/p_d: sound pressure at the footplate relative to at the eardrum. Comparison with human ears.

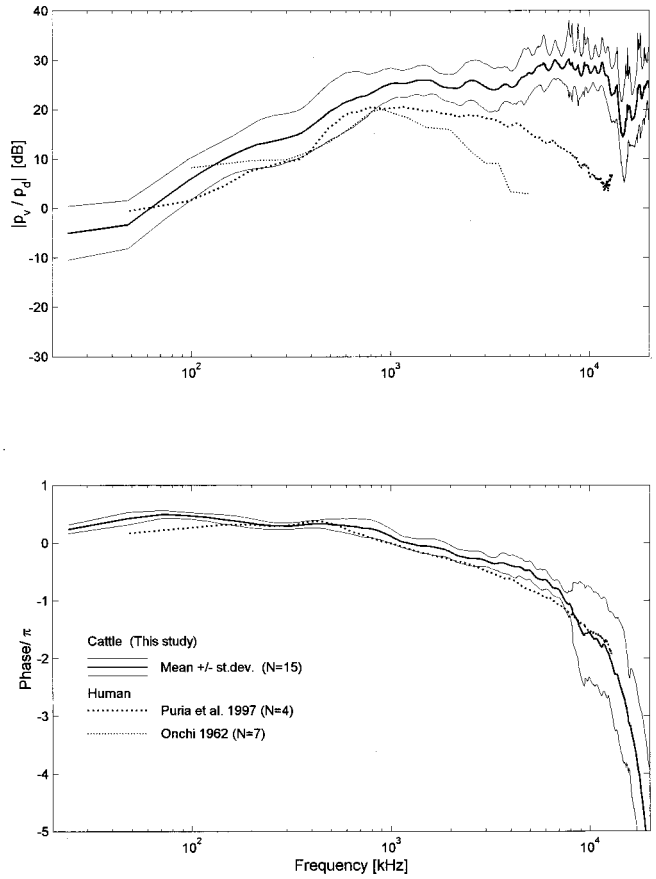


FIG. 14. Ratios of equivalent sound pressures in the ear. p_v/p_d: sound pressure at the inner side of the footplate relative to at the eardrum. Comparison with human ears.

et al. (1990) data. This difference is probably explained by different experimental conditions. In cat, the middle ear is divided in two cavities by a bony septum. The two cavities is acoustically coupled by a hole in the septum, and constitutes a resonator system. The data of Décory *et al.* (1990) were obtained with open bulla and intact septum, and the notch was ascribed to the remaining part of the middle-ear resonator system. The Nedzelnitsky (1980) data were obtained with the bulla widely opened and the septum removed. As in the present study, the influence of the middle-ear cavities is thereby excluded.

Figures 14 and 15 indicate that the vestibulum sound pressure phase for cattle ears is fairly representative also for human ears and the other animal ears.

In Fig. 16, the ratio between equivalent sound pressures at the inner and outer side of the footplate for cattle ears are compared to data for cat ears. The pressure ratio decreases toward low frequencies, due to the stiffness of the annular ligament.

An indication of the cochlear dimensions in ears from Norwegian cattle may be of some interest. For one ear, the cochlea was cut parallel to and close to the modiolus axis, and through the round window. The number of cochlear turns could then be counted. The result was $2\frac{3}{4}$ turn. By a subjective curve fit to the (r,θ) data in Table I, an approximate projection of the cochlear central line along the basilar membrane was drawn in a plane normal to the modiolus axis and on the scale of 10:1. Each mm was marked along the projected line. The lengths of the subsequent half turns were 12.2, 6.0, 5.9, 4.1, 3.3 mm, and 1.5 mm for the last quarter turn, giving a total length 33 mm. The rise in the central line along the axis is only 3.85 mm. The length of the central line, and thus the basilar membrane, could therefore be set equal to the length of the projected line, i.e., 33 mm. The measurements were made by means of a microscope and an *xy* micrometer table. The obtained values for number of turns and length of the basilar membrane, $2\frac{3}{4}$ and 33 mm, respectively, are comparable to the values for the human ear, $2\frac{5}{8}$ and 34 mm (Yost and Nielsen, 1977), but somewhat smaller than the values given by Békésy (1960d) for the cow ear, 3 $\frac{1}{2}$ and over 38 mm.

CONCLUSIONS

With intact stapes, the acoustic input impedance at the oval window, Z_{sc}, is the sum of the cochlear input impedance, Z_c, and the impedance contribution due to the stapes and the annular ligament, Z_s. Z_{sc} is determined mainly by the stiffness of the annular ligament at low frequencies, and by the cochlear input impedance at higher frequencies. For cattle ears, the two impedances Z_s and Z_c are equal in magnitude at about 0.4 kHz. The cochlear input impedance is predominately resistive. A small inertance component due to the stapes and co-vibrating fluid is apparent at high frequen-

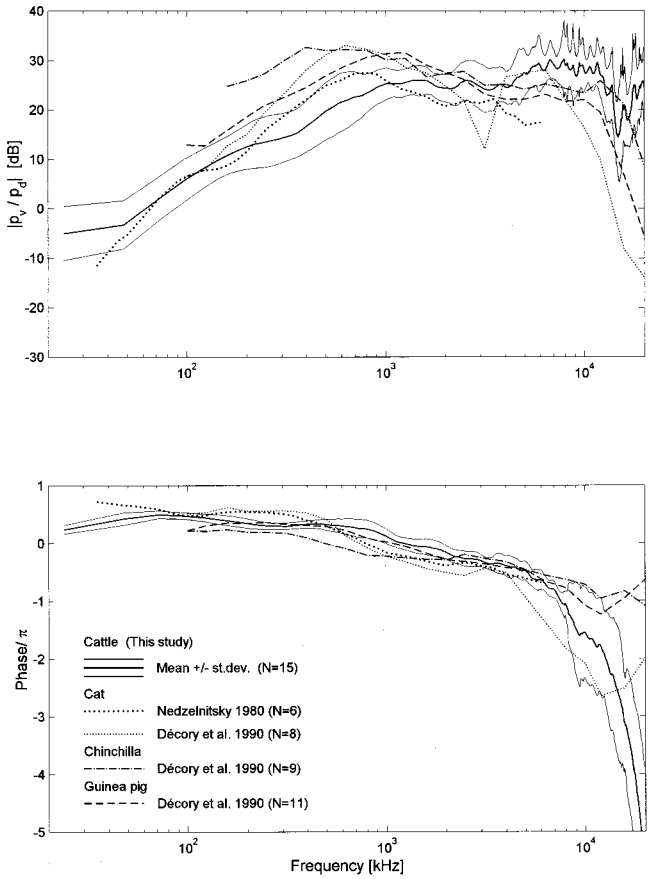


FIG. 15. Ratios of equivalent sound pressures in the ear. p_v/p_d: sound pressure in the vestibulum at the inner side of the footplate relative to at the eardrum. Comparison with animal ears.

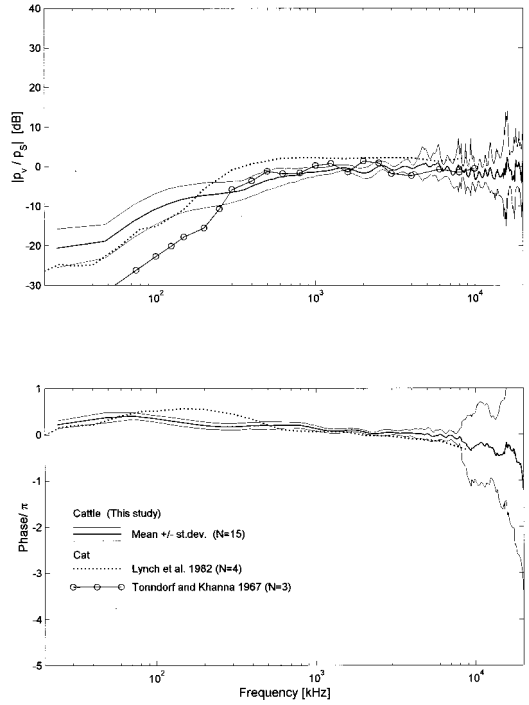


FIG. 16. Ratios of equivalent sound pressures in the ear. p_v/p_s: sound pressure at the inner side of the footplate relative to at the outer side. Comparison with cat ears.

cies. At low frequencies, a small capacitive component is attributed to the stiffness of the round window membrane.

In the referenced models for the human ear, the impedance representations for Z_{sc} seem rather unsatisfactory. For cattle ears, all impedances Z_{sc}, Z_s and Z_c are well represented by simple series RLC circuits:

$$Z_{sc}=\left[62.9+j\left(\omega\,4.68\cdot10^{-4}-\frac{1}{\omega\,6.89\cdot10^{-6}}\right)\right]\mathrm{G}\Omega, \tag{15}$$

$$Z_c=\left[58.2+j\left(\omega\,3.28\cdot10^{-4}-\frac{1}{\omega\,1.52\cdot10^{-4}}\right)\right]\mathrm{G}\Omega, \tag{16}$$

$$Z_s=\left[8.6+j\left(\omega\,1.96\cdot10^{-4}-\frac{1}{\omega\,7.27\cdot10^{-6}}\right)\right]\mathrm{G}\Omega. \tag{17}$$

Here ω is the magnitude of the angular frequency, and 1 acoustic $\mathrm{G}\Omega=10^9$ Pa s/m^3.

The cochlea from Norwegian cattle and the human cochlea are shown to be of comparable size, and the measured impedances seem to be representative also for human ears. To quantify the inter-species impedance differences with some certainty, further investigations are needed. The same measuring method should then be used.

In the whole audio frequency range, the equivalent sound pressure at the stapes is considerably larger than the sound pressure at the eardrum, about 16 dB at low frequencies and up to about 30 dB at higher frequencies. The equivalent sound pressures at the inner and outer side of the footplate are nearly equal at frequencies above 1 kHz. This is consistent with the finding that the input acoustic impedance

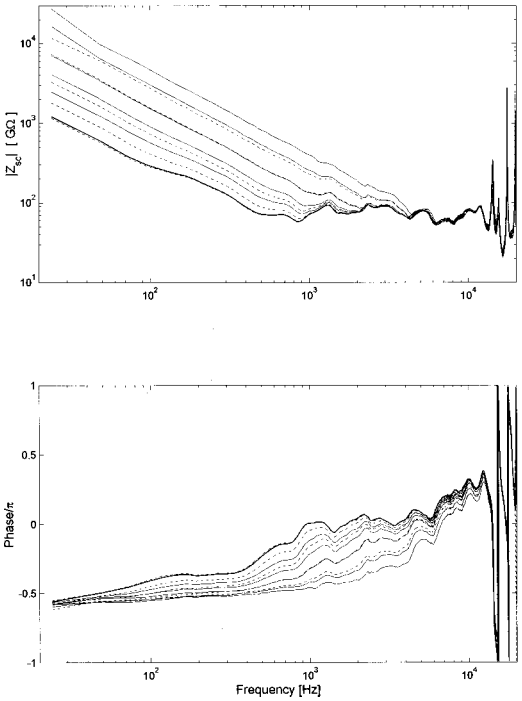


FIG. 17. Variation in oval window impedance with static pressure at the footplate for ear E45. Thick reference curve corresponds to zero pressure, thin continuos curves to overpressures 5, 10, 20, 40, and 60 cm H_2O, and thin dashed curves to underpressures -5, -10, -20, -40, and -60 cm H_2O.

TABLE I. Cochlea dimensions in one ear from Norwegian cattle. The values $a=4$ mm and $b=3.5$ mm for $\theta=0$ are rough estimates.

Cylindrical coordinates of cochlea centerline points			Cross-section axes and area when approximated by an ellipse ($A=\pi ab/4$)		
r (mm)	θ	z (mm)	a (mm)	b (mm)	A (mm^2)
7.07	0	0	4	3.5	11
2.14	π	1.54	1.59	1.35	1.69
1.98	2π	2.82	1.30	1.24	1.27
1.66	3π	2.81	1.21	1.12	1.06
1.00	4π	3.74	1.49	1.17	1.37
1.15	5π	3.85	1.35	0.99	1.05

at the oval window, with and without the footplate in place, are nearly equal in this frequency range. The impedance contribution of the annular ligament is then negligible.

In the vestibulum, the sound pressure gain decreases toward low frequencies. This decrease is caused by the stiffness of the annular ligament. At the low frequency end of the audio range, the sound pressure in the vestibulum is even smaller than at the eardrum.

ACKNOWLEDGMENTS

The author is indebted to Trondheim Slaktehus for easy access to temporal bones from Norwegian cattle, and to J. J. Rosowski for supplying curve data from the articles of Merchant *et al.* (1996) and Puria *et al.* (1997).

Békésy, G. von (**1960a**). *Experiments in Hearing* (McGraw-Hill, New York), p. 407.

Békésy, G. von (**1960b**). *Experiments in Hearing* (McGraw-Hill, New York), p. 100.

Békésy, G. von (**1960c**). *Experiments in Hearing* (McGraw-Hill, New York), p. 436.

Békésy, G. von (**1960d**). *Experiments in Hearing* (McGraw-Hill, New York), p. 507.

Décory, L., Franke, R. B., and Dancer, A. L. (**1990**). ''Measurement of the middle ear transfer function in cat, chinchilla and guinea pig,'' in *Lecture Notes in Biomathematics, Vol. 87: The Mechanics and Biophysics of Hearing, Proceedings of a Conference Held at the University of Wisconsin, Madison, WI, June 25–29, 1990*, edited by P. Dallos *et al.* (Springer-Verlag, Berlin), pp. 270–277.

Hudde, H., and Engel, A. (**1998**). ''Measuring and modeling basic properties of the human middle ear and ear canal. Part III: Eardrum impedances, transfer functions and model calculations,'' Acustica **84**, 1091–1108.

Kinsler, L. E., and Frey, A. R. (**1962**). *Fundamentals of Acoustics* (Wiley, New York), p. 181.

Kringlebotn, M., and Gundersen, T. (**1985**). ''Frequency characteristics of the middle ear,'' J. Acoust. Soc. Am. **77**, 159–164.

Kringlebotn, M. (**1995**). ''The equality of volume displacements in the inner ear windows,'' J. Acoust. Soc. Am. **98**, 192–196.

Kringlebotn, M. (**2000a**). ''Frequency characteristics of sound transmission in middle ears from Norwegian cattle, and the effect of static pressure differences across the tympanic membrane and the footplate,'' J. Acoust. Soc. Am. **107**, 1442–1450.

Kringlebotn, M. (**2000b**). ''Rupture pressures of membranes in the ear,'' Ann. Otol. Rhinol. Laryngol. (in press).

Kurokawa, H., and Goode, R. (**1995**). ''Sound pressure gain produced by the human middle ear,'' Otolaryngol.—Head Neck Surg. **113**, 349–355.

Lynch III, T. J., Nedzelnidsky, V., and Peake, W. T. (**1982**). ''Input impedance of the cochlea in cat,'' J. Acoust. Soc. Am. **72**, 108–130.

Merchant, S. N., Ravicz, M. E., and Rosowski, J. J. (**1996**). ''Acoustic input impedance of the stapes and cochlea in human temporal bones,'' Hear. Res. **97**, 30–45.

Nedzelnitsky, V. (**1980**). ''Sound pressures in the basal turn of the cat cochlea.'' J. Acoust. Soc. Am. **68**, 1676–1689.

Puria, S., Peake, W. T., and Rosowski, J. J. (**1997**). ''Sound-pressure measurements in the cochlear vestibule of human-cadaver ears,'' J. Acoust. Soc. Am. **101**, 2754–2770.

Voss, S. E., Rosowski, J. J., and Peake, W. T. (**1996**). ''Is the pressure difference between the oval and round windows the effective acoustic stimulus for the cochlea?'' J. Acoust. Soc. Am. **100**, 1602–1616.

Yost, W. A., and Nielsen, D. W. (**1977**). *Fundamentals of Hearing* (Holt, Rinehart, and Winston, New York), p. 64.

Zwislocki (**1975**). ''The role of the external and middle ear in sound transmission,'' in *The Nervous System, Vol. 3: Human Communication and Its Disorders*, edited by D. B. Tower (Raven, New York), pp. 45–53.

An automated procedure for identifying spontaneous otoacoustic emissions

Edward G. Pasanen and Dennis McFadden[a)]
Department of Psychology and Institute for Neuroscience, Mezes Hall 330, University of Texas, Austin, Texas 78712

(Received 30 December 1999; revised 3 May 2000; accepted 18 May 2000)

An algorithm is described for objectively identifying and measuring spontaneous otoacoustic emissions (SOAEs) using the spectrum that results from transformation of the acoustic waveform measured in the outer ear canal. Prior to spectral analysis, the rms level is calculated for successive short segments of the waveform and only the weakest 25% of the segments are retained for the spectral analysis [the quietest 150 when using 16k-point fast Fourier transforms (FFTs)]. The resulting *initial* spectrum is scanned for peaks (potential SOAEs) which are then deleted from the spectrum. New values are estimated for the deleted values using linear extrapolations from frequency ranges on either side of the deleted values. The end result is a *smoothed* spectrum devoid of all local peaks. The initial spectrum is then compared peak-by-peak with the smoothed spectrum, and those peaks having differences that exceed an objectively determined decision criterion are identified as likely SOAEs. The effects of varying some of the important parameter values of the algorithm are described, and the sensitivity of the procedure is evaluated by measuring the detection rate for a Lorentzian peak of known amplitude added to a spectrum otherwise devoid of SOAEs. [S0001-4966(00)00609-3]

PACS numbers: 43.64.Jb, 43.64.Yp, 43.58.Ta [BLM]

INTRODUCTION

Spontaneous otoacoustic emissions (SOAEs) are narrow-band sounds that are generated in the cochlea and then propagate back through the middle ear into the outer ear canal where they are detectable using a sensitive microphone system. SOAEs are continuously present, relatively stable, and can be found in the majority of ears with normal hearing (see Probst *et al.*, 1991, for a general review). While these sounds can be as strong as 30 dB SPL, the majority are much weaker, and the weakest ones detected are only about 1 dB above the noise floor of the measuring system. As instrumentation and analysis methodology have developed over the years since SOAEs were first detected by Kemp (1979), the reported prevalence and number of SOAEs in human ears has steadily increased (see Talmadge *et al.*, 1993; Zhang and Penner, 1998). It appears that SOAEs are likely to be found in greater numbers as the state of the art in instrumentation and analysis procedures continues to advance.

Historically, SOAEs have typically been identified by an observer simply examining a spectral representation of the waveform collected from the ear canal and noting those peaks that appeared to exceed the background noise floor. However, there are a number of situations where an objective, automated system for detecting SOAEs is desirable. If the number or size of the SOAEs in two or more populations or treatment groups are being compared, the scoring procedure should be free of bias. Identification of SOAEs ''by eye'' involves a certain amount of practice, and when the task falls upon the same experimenters who generated the hypotheses under test, readers and reviewers can become concerned about possible biases. Furthermore, if there are many ears to be analyzed, automation can save labor. In addition to objectivity and efficiency, an automated system offers the availability of a repeatable, standard procedure, assuring experimenters in different laboratories that their methods are identical, and their results truly comparable.

Although early procedures for SOAE detection could be made somewhat objective (see Whitehead *et al.*, 1993, for example), it was not until the implementation of off-line analyses of digitized ear-canal recordings (see Bilger *et al.*, 1990; Rao and Bilger, 1991) that it was possible to automate the process. Having a relatively long sample of the waveform in the outer ear canal made it possible to establish a noise criterion for acceptance of individual segments for spectral analysis, enabling analysis and averaging of only those samples relatively uncontaminated by noise. Talmadge *et al.* (1993) reported an analytic procedure for quantifying SOAEs based on digital recordings, and while their detection procedure was not fully automated, it was highly objective. Penner *et al.* (1993) reported an automated procedure for detection of SOAEs from digital recordings, and she and her colleagues have subsequently explored several approaches to the detection problem (Penner and Zhang, 1997; Zhang and Penner, 1998). Our laboratory began an independent investigation of the problem, motivated by the need to detect and measure SOAEs objectively in a large number of subjects from several different subpopulations (McFadden and Pasanen, 1999). Here, we describe an automated procedure that uses off-line analysis of digital recordings of the output of a commercially available OAE measurement microphone system. An early version of our procedure was presented in

[a)]Author to whom correspondence should be addressed. Electronic mail: mcfadden@psy.utexas.edu

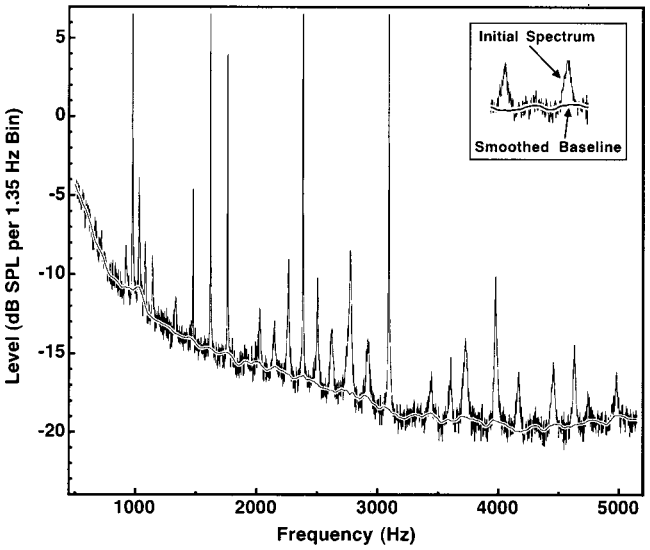


FIG. 1. An example of an initial spectrum constructed by summing the spectra from FFTs performed on the 150 quietest 743-ms segments present in 2 min of acoustic waveform from the external ear canal. The ordinate values have been transformed to decibels sound-pressure level, and to improve clarity, only spectral values 500–5150 Hz have been plotted. The solid line centered in the white band is the smoothed baseline obtained for this spectrum. Inset: a segment of the spectrum beginning at 4100 Hz is magnified to show the smoothed baseline more clearly. Because the FFTs involved 16 384 sample points, the abscissa values are separated by 1.35 Hz.

Pasanen and McFadden (1996), and then described briefly in McFadden and Pasanen (1999). Data from the latter study will be used here to explicate and evaluate our SOAE detection procedure, and will be compared with results from Talmadge *et al.* (1993), Whitehead *et al.* (1993), and Zhang and Penner (1998).

Although our procedure involves a relatively large number of steps, the computations themselves are mathematically simple, easy to implement, and quite rapid. After the ear-canal recording is obtained, the complete analysis for a single ear currently requires less than 30 s of computation time on a relatively fast (in 2000) desktop personal computer (Power Macintosh G4, 400 MHz). It is therefore rapid enough to be accomplished while a subject is still in place, thereby allowing for the collection of replacement data when the need arises.

There are two extreme cases that pose special challenges to any SOAE-detection procedure. First is the problem of detecting weak SOAEs. The SOAE-detection process always involves identification of a peak in a noisy spectrum, and then estimation of the likelihood that the peak in question arose from a process other than the observed noisy background. The value of any SOAE-detection procedure lies to a great extent in the confidence the user has that a relatively small spectral peak can be called an SOAE. The other extreme case challenging any SOAE detector is the ear containing many closely spaced SOAEs. Figure 1 is a spectral plot obtained from such an ear. Identifying the strong SOAEs in this ear would be easy using almost any procedure, but it is a far greater challenge to develop a system for identifying the many weak components present. In the development of the procedure described here, particular attention was devoted to satisfactory treatment of these two extreme conditions of weak SOAEs and multiple, closely spaced SOAEs.

Whether the SOAEs in question are weak or strong, their detection always involves a criterion—implicitly even when not explicitly. There are in fact situations where it is advantageous to allow the detection criterion to vary in a systematic and quantifiable way. For example, it is known that SOAEs are more prevalent, more numerous, and more powerful in female ears than in male ears (e.g., Talmadge *et al.*, 1993). The use of an overly strict criterion when comparing SOAEs in the two sexes could exaggerate the true sex difference because male ears have weaker and less easily detected SOAEs. If an experimenter were able to vary systematically the criterion for accepting a spectral peak as an SOAE while making comparisons between males and females, any group differences found could be stated more precisely (e.g., the group difference existed independent of criterion, only with a strict criterion, or whatever). Although traditionally it has been desirable to minimize the likelihood of erroneous identification of a spectral peak as an SOAE when it is not (a false alarm), it might sometimes be preferable to accept a very slight increase in false alarms as a small price to pay for the proportionally greater number of SOAEs detected, and (particularly in male subjects, for example) the greater number of ears contributing to SOAE statistics for the group. In regard to the criterion issue, our desire was to develop an automated detection procedure that was ''free-standing'' or ''self-contained,'' in that the criterion would be based on the data at hand rather than on normative data collected elsewhere.

Figure 1 illustrates several characteristics of SOAEs: the minimum spacing between SOAEs is approximately proportional to emission frequency (see, e.g., Talmadge *et al.*, 1993; Zwicker, 1990, for reviews) and the bandwidths of emissions at their bases show a similar variation with frequency.[1] Accordingly, the procedure described here involves some parameters that change proportionally with frequency. We believe that this yields a detection procedure that is more effective in detecting weak emissions than any of the more general-purpose smoothing and peak-detection algorithms that are available.

I. PROCEDURE

This section is intentionally lengthy out of the desire to provide sufficient detail for readers to implement or change the detection procedure in their own laboratories. For what follows, the entire spectrum-averaging and SOAE-detection procedure was implemented off-line. However, with advances in computing speed, the procedure is implementable with minimal delay after the recordings are obtained. All of the data shown are from subjects in an experiment reported by McFadden and Pasanen (1999) that was concerned with differences in the SOAEs in people of differing sexual orientations. A complete description of the data-collection procedures can be found in that paper. The raw data were in the form of digitized records of the acoustic waveform obtained from a microphone (Etymotic ER-10) placed in the ear canal of a human subject lying quietly in a soundproofed room. The waveform was amplified and high-pass filtered at 400

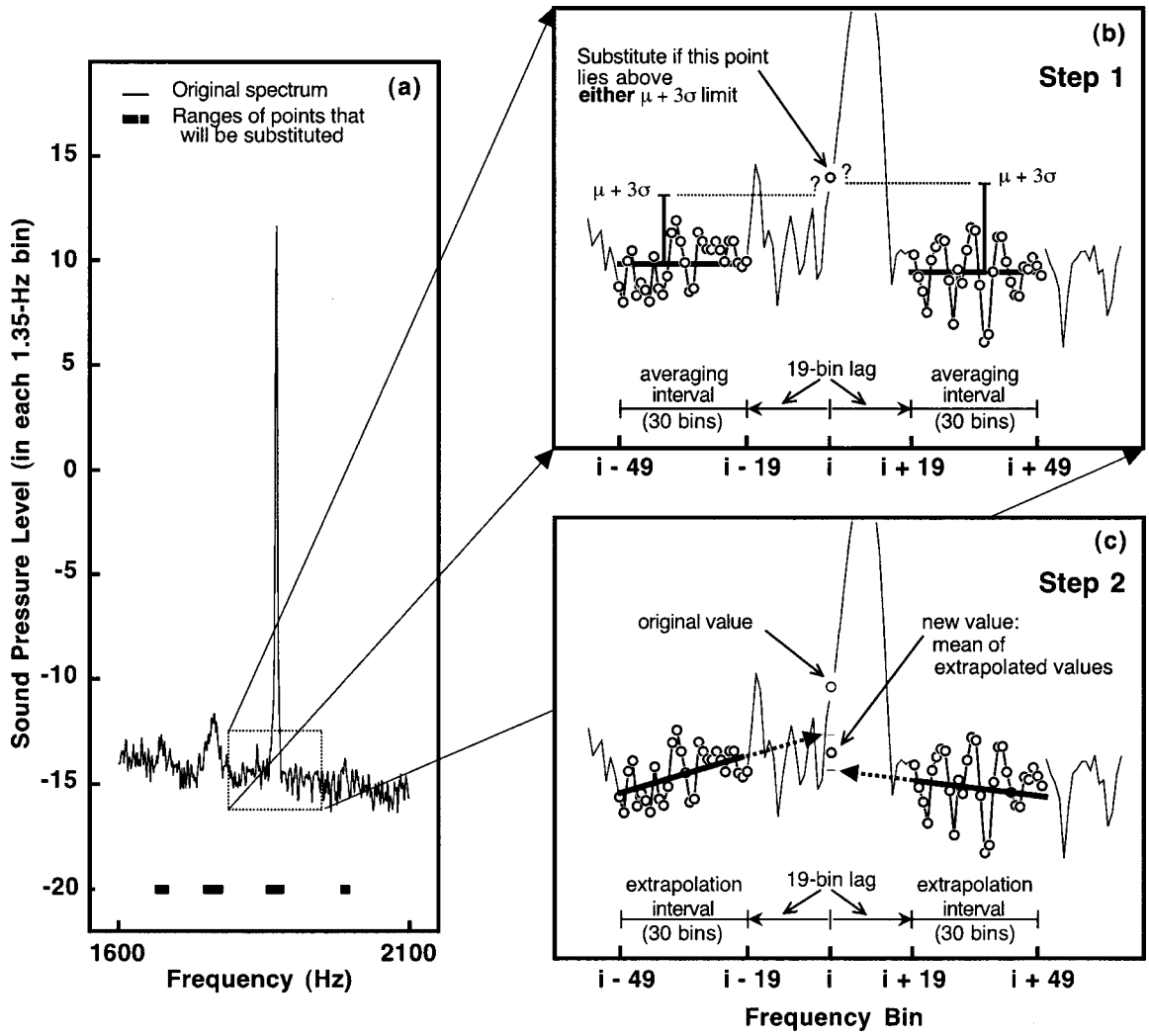


FIG. 2. An illustration of the processes of deleting and replacing extreme ordinate values. (a) An excerpt from an initial spectrum containing several local peaks. The black bars under the spectrum mark the ranges of values deleted because they exceeded the local mean plus 3 s.d.'s. (b) An expanded view of a portion of the initial spectrum showing the sliding ranges over which the local means and standard deviations were calculated, and a point satisfying the criteria for deletion. (c) An example of how a new value was calculated for a deleted value by averaging two linear extrapolations, one from a frequency range below the deleted value and one from a range above it. The lag, in number of bins, is proportional to frequency and is 19 bins for this illustration, at 1900 Hz.

Hz in order to remove the noise from bodily movements, etc. The sampling frequency was 22.05 kHz. Typically, the data were obtained as four 30-s recordings taken from each ear, and, typically, the time interval between successive recordings was 10–60 s. All data collection and subsequent analyses were performed on a Macintosh personal computer, using the LabVIEW® programming environment (National Instruments), and this platform-independent software is available from the authors upon request. The data acquisition and digital control interface hardware were plug-in boards also manufactured by National Instruments.

The SOAE-detection procedure is broadly divided into five phases. First is the extraction of the initial spectrum from the waveforms recorded in the ear canal. Second is the determination of the smooth, ''frequency response'' function (called the smoothed baseline) from that initial spectrum. Third is the expression of spectral peaks as deviations from this smoothed baseline. Fourth is identification of SOAEs, and fifth is the assignment of frequency and level to the peaks identified as SOAEs.

A. Phase 1: The initial spectrum

The first goal of the procedure was to eliminate the noisiest time periods in the recordings prior to performing spectral analysis. The entire 2-min record was divided up into successive, overlapping segments of approximately 743-ms duration each, with the onsets of successive segments being separated by approximately 186 ms. Given the sampling rate used to make the digital recordings (22.05 kHz), these values correspond to 16 384 sample points per segment, and 75% overlap of successive segments. The rms level was computed for each time segment, and a histogram of rms values was compiled across all the time segments for that ear. For each of the 150 time segments having the lowest rms values, fast Fourier transforms (16k FFTs, Hanning window) were computed and averaged. (That is, only the quietest 25% of the available time segments was used for the analysis.) The resulting averaged spectrum will be called the *initial spectrum*; the estimates of level (in decibels) are in frequency bins 1.35 Hz in width. An example is shown in

Fig. 1. This spectrum can be conceptualized as a relatively uniform noise process across the frequency range, with certain positive peaks deviating significantly from that noise background. The general aim of the automated SOAE-detection procedure is to determine for each local peak in the spectrum a measure of the likelihood that it resulted from the broadband noise process that comprised the background. When that computed likelihood is very small, it is assumed that the peak arose from a *different* process, namely the process responsible for producing an SOAE.

It should be noted that all smoothing and detection operations described here were performed in semilog space; i.e., ordinate units of decibels and abscissa units of frequency in linearly spaced bins. Linear units of power were used only for the final calculation of the total power[2] for an individual SOAE [see Sec. E (phase 5) below] and for the simulation of an emission (Sec. II C).

B. Phase 2: The smoothed baseline

Ignoring the SOAEs for a moment, it is clear from Fig. 1 that the initial spectrum is not flat. The frequency response of the recording system and the spectrum of the noise in the ear canal combine to shape the overall contour seen in Fig. 1. It is useful therefore to model the initial spectrum as a slowly varying, smooth function (the frequency response), upon which is superimposed a rapidly varying, fine-grain process consisting of noise and SOAEs. (The character of this noise process was established to be essentially constant over the entire frequency range under study, and to be similar for male and female ears.[3])

Obtaining the smoothed baseline involves (a) scanning the initial spectrum for peaks that deviate substantially from the rest of the spectrum in that local region; (b) removing those peaks plus any surrounding deviant values, and substituting estimated values that are in better agreement with the data in that immediate vicinity; and (c) smoothing using conventional rectangular and median filters on the spectrum containing the substituted values. The substitution procedure is illustrated in Fig. 2.

To begin this process, a pass is made through the entire spectrum, bin by bin, and each spectral value is compared with two acceptance limits calculated from neighboring data in order to determine whether the value is sufficiently extreme to be deleted and replaced. The two acceptance limits are obtained by calculating means and standard deviations for two 30-bin frequency regions flanking the spectral value of interest, one on its low-frequency side and one on its high-frequency side. If the spectral value under consideration is greater than the mean plus 3 standard deviations (SDs) for *either* 30-bin region, that frequency bin is flagged for replacement; otherwise, it is kept, and the algorithm slides up one bin and repeats the process. The original spectral value is replaced by a calculated value in the following way (see Fig. 2). A linear regression line is fitted to the points in each of the sliding 30-bin frequency regions that were used to calculate the means and standard deviations mentioned above, and the two regression lines are extrapolated to the frequency bin containing the flagged point. One extrapolation is upward from the frequency region below the spectral value of interest, and the other extrapolation is downward from the corresponding region above the flagged point. The two extrapolated values are averaged, and that average is substituted for the flagged point. Then, the algorithm slides up one bin and repeats the above process. To ensure convergence of this process (see below), the entire first pass of checking and extrapolation is performed using only the original spectral values. The new spectral values arising from extrapolation are not used until the next step in the procedure.

The smoothing process begins with the 300-Hz bin. For that bin, the 30-bin acceptance and extrapolation intervals begin 3 frequency bins below and above the 300-Hz bin. This 3-bin interval (3.04 Hz) is called the *lag*, and is varied proportionally to the frequency of the spectral point under evaluation (as is the frequency separation of actual SOAEs—see Zwicker, 1990; Probst *et al.*, 1991 for reviews). For example, the lag at 900 Hz is 9 bins (12.1 Hz), and at 9000 Hz is 90 bins (121.1 Hz). As noted above, the output of the ear-canal microphone was high-pass filtered at 400 Hz before being amplified and digitized, to reduce the problem of low-frequency noise overloading the recording system. Because there is often substantial biological noise remaining, SOAEs are difficult to detect reliably at frequencies immediately above 400 Hz. Here, only SOAEs above 550 Hz are considered. Smoothing begins at 300 Hz to ensure that no transient artifacts from the initiation of the smoothing process are present above 550 Hz.

During this initial smoothing stage of the procedure, every point in the original spectrum is evaluated for its consistency with neighboring values, and a new, somewhat smoother spectrum is generated as a result, with the larger peaks absent or at least reduced considerably. Note that a log is kept of those bins in the new spectrum that now contain substituted values.

Next, several additional passes are made through the new spectrum in order to smooth it further. The procedure used is slightly different from the preceding stage in order to avoid the problem of recursive substitution. In our algorithm, newly substituted values are never immediately used for extrapolation during the current pass. (A recursive substitution procedure can, in some instances, generate a function that diverges substantially from the shape of the initial spectrum.) Following the initial substitution pass, the spectrum is smoothed using a 19-point median filter.[4] A second substitution pass is then made upon the new, smoothed spectrum, again starting at the lowest frequency. All previously flagged and substituted points (but only these points) are revisited, and each flagged point is again replaced, this time using only the linear extrapolation from the 30-bin interval *above* the point. Thus, each logged bin is again assigned a new value, based on the (recently smoothed) neighboring data points. A third substitution pass is then performed, but it proceeds from high frequency to low, and uses only the 30-bin extrapolation interval *below* each logged point. The 19-point median filtering is then repeated. The two one-sided substitution passes just described are then performed again, followed by a final smoothing of the entire spectrum using a 51-point rectangular window[4] that works from low to high frequency. (This procedure of

alternating substitution from above and below was adopted because in some spectra containing many closely spaced emissions, a ''clean'' region for extrapolation is sometimes difficult to find.) It should be emphasized that, during all the stages described in this paragraph, only the originally flagged bins are reexamined and recomputed. All other bins are changed only by the median- and rectangular-window smoothing processes. The result of these multiple replacement and smoothing operations is a relatively slowly varying function passing through the middle of the ''fine grain'' of the initial spectrum. This final function is called the *smoothed baseline*. In Fig. 1, the smoothed baseline is the solid line centered in the white band that follows the contour of the spectrum.

C. Phase 3: Spectral peaks

The next stage in the analysis is to identify all positive peaks in the initial spectrum and obtain an estimate of each peak's deviation from the noise floor. For the purposes of this stage, a spectral peak is defined simply as any point in the initial spectrum where the slope changes from positive to negative with increasing frequency. Typically, this is equivalent to a point having a value greater than either of the two immediately flanking points. Under such a broad definition, it follows that there will be many peaks detected in the initial spectrum, typically more than 1500 across the 9-kHz range of frequencies considered here. All such peaks are identified, their locations on the frequency axis noted, and for each peak, a *peak deviation* is calculated. The peak deviation is the mean of the five spectral values centered at the identified peak in the initial spectrum, minus the mean of the corresponding five values in the smoothed baseline. The set of all peak-deviation values, each corresponding to a unique spectral peak, is assembled into an array. The elements in this array of peak-deviation values are checked to determine whether the spectral bin containing the peak had been flagged during the smoothing operation above, indicating that the peak was sufficiently deviant to give rise to a substituted value during the smoothing process. The set of all *unflagged* peaks, those *not* corresponding to substituted points, is now assembled into a second array, and the mean and standard deviation are computed for this set of peak-deviation values. Note that these latter peaks can be traced to regions in the initial spectrum where, in the first smoothing pass, it was determined that an SOAE was *not* likely to be found. Put in signal-detection terms, this set of unflagged peak-deviation values forms an estimate of the ''noise-alone'' distribution of deviation values, against which deviation values of observed peaks can be compared.

D. Phase 4: Identification of SOAEs

The next step is to determine whether a given peak in the initial spectrum is to be called an SOAE. Using the mean and standard deviation of the set of unflagged peaks (just described above), each entry in the array of peak-deviation values for *all* spectral peaks (flagged and unflagged) is now expressed as a multiple of standard-deviation units above the mean. That is, 5-bin peak-deviation values for the initial spectrum are expressed as multiples of the standard deviation of the peak deviations obtained from spectral regions judged likely not to contain SOAEs. It should be clear that the greater the magnitude of the standard-deviation multiple for a particular 5-bin region surrounding a peak in the initial spectrum, the greater the likelihood that the values in that region arose from a process other than that which generated the set of unflagged peaks. Said differently, the larger the standard-deviation multiple associated with any peak, the greater the confidence one has in declaring that peak to be an SOAE. In signal-detection terms, the standard-deviation multiple is the decision variable. Once a criterion value has been selected for the standard-deviation multiple, any peak in the initial spectrum can be tested against that criterion value and an objective, binary decision made about its status as an SOAE.[5]

Because the definition of a peak was simply ''a change in slope from positive to negative,'' there may be many more peaks identified as SOAEs than truly are. For example, a spectral peak associated with a true SOAE may well have secondary peaks along the skirts of the main peak, and some of these might also satisfy the decision criterion. Hence, a final cleanup is necessary. This final stage also takes into account the general belief that two true, independent SOAEs cannot be located within about 0.1 octave of each other (Zwicker, 1990). The cleanup begins by ranking the set of peaks identified as SOAEs in order of their standard-deviation multiples. The peak having the highest multiple of all is accepted finally as an SOAE, and, along with any other peaks within 0.1 octave of that peak, is deleted from the set. From the remaining peaks, the one having the highest multiple is also finally accepted as an SOAE, it and those within 0.1 octave are deleted, and so on, until the entire set is checked in this way. This cleanup eliminates secondary peaks occurring by chance in the immediate vicinity of a real SOAE, some intermittent SOAEs, and a portion of the SOAEs that are actually distortion products generated by the interaction of other SOAEs (see Burns *et al.*, 1984; Probst *et al.*, 1991; van Dijk and Wit, 1998). (Note that, in this version of our algorithm, no explicit attempt is made to determine whether identified SOAEs might be distortion products.)

E. Phase 5: Assignment of frequency and level to identified SOAEs

The frequency actually assigned to a peak identified as an SOAE is simply the frequency bin that contained the peak in the initial spectrum. The level of each SOAE detected in a given ear is estimated using the following summing procedure performed on the initial spectrum. Beginning with the spectral peak itself and moving toward higher frequencies, values in successive bins are summed (in units of power) as long as the addition of each additional point increases the sum (total power) by at least 1.0%. Then, the process is continued beginning with the bin immediately below the spectral peak and moving downward. Finally, the sum is converted from units of power back to SPL.

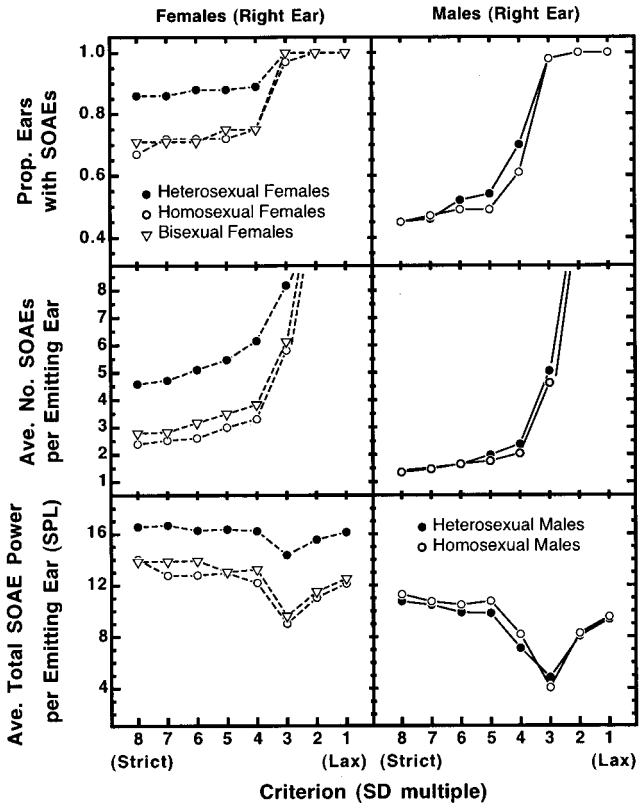


FIG. 3. The proportion of ears exhibiting SOAEs (top panels), the average number of SOAEs per emitting ear (middle panels), and the strength of the SOAEs identified (bottom panels) plotted as a function of the decision criterion adopted (expressed as a multiple of the SD of peak deviation scores) and shown separately for males and females (right ears only). Symbols denote the data from different subject groups. Clearly, all experimental conclusions would be the same for any decision criterion greater than about 4. Data are from subjects in McFadden and Pasanen (1999).

II. EVALUATION OF THE PROCEDURE

Two different objective approaches have been used to evaluate our SOAE-detection algorithm. First, default values for the numerous parameters in the smoothing and detection processes were determined; then the various parameters in the algorithm were varied, the data for a fixed group of subjects were reanalyzed, and the effect of that parameter variation on the summary statistics for that group was observed. The summary statistics used consisted of one global measure, the proportion of ears (or subjects) in the sample having detectable SOAEs, and two individual ear measures—number of SOAEs detected in an ear and total SOAE power (summed across all SOAEs present in an ear), both averaged over all ears in the sample. This first approach is treated in detail in Secs. II A and II B. The second approach (Sec. II C) was a simulation, where synthesized SOAEs in the form of Lorentzian peaks (see Talmadge *et al.*, 1993) of known height and bandwidth were added to spectra otherwise devoid of detectable emissions. By submitting the spectra with the added Lorentzian to the detection algorithm, detection performance could be estimated as a function of known SOAE magnitude and bandwidth. Both of these approaches were based on real-ear data, the same 2-min recordings from ears of subjects reported in McFadden and Pasanen (1999). (In that paper, we reported that the SOAEs of homosexual and bisexual females differed in number and strength from those of heterosexual females, but the SOAEs of homosexual and bisexual males did not differ from those of heterosexual males.)

A. Selection of decision criterion

The most salient of all the parameters in the detection procedure is the standard-deviation multiple used as the decision criterion for acceptance of a spectral peak as an SOAE. That criterion will have a direct effect on almost any summary statistic involving SOAEs. Figure 3 shows three summary measures plotted as a function of decision criterion level, for the right ears of male and female subjects from McFadden and Pasanen (1999). The top panels show the effect of SOAE decision criterion on the proportion of right ears having at least one SOAE detected by the algorithm for three groups of male and female subjects: heterosexual, homosexual, and bisexual (bisexual males, a very small sample, was omitted for clarity). For criterion values in the range 8.0 to 5.0 (strict to lax), the proportion of emitting ears is quite stable, with a sharp break in the curves as criterion was set lower than 5.0 (4.0 for females). For the three groups of females shown in the left-hand panel, the curves are very nearly parallel, suggesting that any conclusions drawn about differences between groups were not dependent on the value of decision criterion selected (provided it was not too lax, of course). The pattern for the middle panels, showing average number of SOAEs per emitting ear, is also parallel and very stable, with the sharp break in the curves occurring at the same criterion value (4.0 SDs) for both sexes. The bottom panels show the effect of decision criterion on the average total SOAE power per emitting ear, and the pattern is somewhat different. Although the absolute values of total power vary (and often substantially) as criterion is varied, the plots are roughly parallel. The difference in total SOAE power between heterosexual and nonheterosexual females, for example, is relatively constant for all values of the decision criterion, even for the most lax values at the extreme right. The data from these three summary measures suggest that a decision criterion somewhere above 4.0 would be desirable.

The decision criterion selected by McFadden and Pasanen (1999) was a standard-deviation multiple of 5.0. That value was chosen by comparing, for a large number of ears, the SOAEs identified by the algorithm to those obtained by visual scoring by experienced observers, in ignorance of the subject group in each instance, and also in ignorance of the plots of Fig. 3. The value of 5.0 was chosen as a slightly conservative compromise; SOAEs so identified by the algorithm were readily accepted as SOAEs after visual inspection of the spectra. When the criterion was set to 4.0, there were a few peaks identified as SOAEs that a human observer might question, but none that an observer would reject outright. However, at a criterion level of 3.0, there were peaks identified as SOAEs that, on inspection, would be considered false alarms. As the criterion value was raised to 7.0 and higher, there were many spectral peaks rejected by the algorithm as SOAEs that were considered to be real SOAEs upon visual inspection. The criterion and false-alarm issues are examined further in Secs. C and D below.

B. Smoothing parameters

In the spectral smoothing stage of the algorithm, the size of the extrapolation and smoothing windows, and the lag in bins separating the extrapolation window and the spectral point under consideration, can in some cases affect the fit of the smoothed baseline to the spectrum in those regions where no SOAEs are present. The success of the procedure depends upon a good fit in those regions because an SOAE is indicated by the very *lack* of a good fit in a small local region. Any procedure must balance the competing needs to fit well in the broad sense but not so well locally that the smoothed baseline function follows the contours of relatively broadband, weak SOAEs, making them impossible to detect. If the smoothed baseline fits the spectrum poorly over a wide range of frequencies, the mean peak-deviation score for the unflagged points will be inflated, resulting in an artificially strict decision criterion. If the region of poor fit is small, a moderate but broad elevation in that spectral region might be incorrectly identified as an SOAE.

To test the robustness of the algorithm to variations in certain parameter values, a group of heterosexual female subjects from McFadden and Pasanen (1999) was analyzed several times, with a single parameter varied from its default value each time. Four parameters were varied: extrapolation window (20, 30, or 40 bins), lag (2, 3, or 4 bins at 300 Hz and the corresponding proportions at higher frequencies), smoothing window (41, 51, or 61 bins), and the acceptance criterion for substitution and extrapolation in the initial spectrum (2.5-, 3.0-, or 3.5-s.d. multiples). The default value for each parameter was the middle one of each triplet. Three summary statistics for the group—proportion of ears having SOAEs, average number of SOAEs per ear, and average total SOAE power per emitting ear—were computed for each analysis. As each parameter was varied, the overall change in the three summary statistics averaged about $\pm$ 1%, with a maximum of 1.8% for lag value, smoothing, and extrapolation windows. The analysis was somewhat more sensitive to changes in the criterion for substitution and extrapolation, where the summary statistics varied as much as 2.9%. Because this latter parameter determines directly the set of flagged peaks in the initial spectrum, thereby directly affecting the decision criterion, it is not surprising that selecting the value for that parameter is so critical. The important point is that the summary statistics changed relatively little over fairly wide ranges for these parameters. Unfortunately, there is no ''gold standard'' for goodness of fit that would guide the parameter-selection process. As noted, the values finally chosen as default for these parameters came from examining spectra from a large number of ears having many SOAEs, and visual assessment of goodness of fit in the neighborhood of strong, closely spaced SOAEs.

It could be argued that, given a value for the decision criterion which results in very infrequent false alarms, the procedure should be optimized for the highest rate of SOAE detection possible. In fact, the default set of parameter values did yield the maximum number of SOAEs detected per ear and the maximum number of ears having at least one emission (or very nearly so), when looking at summary statistics for the various subject groups in McFadden and Pasanen (1999).

C. Detection of simulated SOAEs

SOAE detection can be viewed as an example of the classic Neyman–Pearson decision problem (Van Trees, 1968). In any given ear, the *a priori* probability of an SOAE occurring at any selected frequency is not known. The best that an observer can do is establish a criterion for identification of an SOAE such that the probability of a false detection is acceptably low. One way to evaluate a detection algorithm under such conditions is by simulation. For example, spectra that by other analyses, or by inspection, observers agree contain no SOAEs can be submitted to the detection algorithm for analysis, and any peaks identified as SOAEs by the algorithm can be viewed as false detections. For an additional test, a spectral peak can be synthesized and added to such spectra to mimic a real SOAE. The sensitivity of the algorithm can be inferred from the likelihood of that known added peak being detected as its peak amplitude is varied. Both of these approaches were used.

Because the noise floor of the recording system is not flat as a function of frequency, any characterization of the probability of detection of an SOAE simply as a function of its SPL is meaningless. However, as noted above, local fluctuations in the noise floor about the smoothed baseline are very similar in magnitude over the entire frequency range under study.[3] It follows that the likelihood of detecting a given SOAE, expressed as a fixed *increment* above the noise floor (or smoothed baseline) will be approximately constant.

The ears of subjects from McFadden and Pasanen (1999) having no SOAEs detected by the algorithm when the decision criterion was set to 5.0 s.d.'s were rechecked by eye to obtain 60 (16 female, 44 male) having no identifiable SOAEs. Averaged spectra were obtained, based on the 150 quietest time intervals (phase 1 described above). A Lorentzian peak of fixed height and bandwidth was added, in units of power, to various regions of the spectra obtained from each ear. For each ear a starting location was selected at random in the vicinity of 550 Hz. The peak was added to the spectrum at that location, and the resulting spectrum was submitted to the smoothing and detection process. Any peak identified as an SOAE within a 50-bin window (67.5 Hz) centered at the point of addition was scored as a ''detection.'' The center point of addition of the Lorentzian was then moved 50 bins higher in frequency than the initial random location, and the smoothing and detection procedure was repeated. This 50-bin stepping procedure was repeated so that the addition occurred at locations covering the entire spectrum up to 9 kHz in each of the 60 ears. Each spectral bin was therefore sampled only once on any individual simulation run, ensuring that any aberrant peak in the initial spectrum that might be (incorrectly) identified as an SOAE would contribute only once to the total count of detections. This simulation was repeated for a number of amplitudes of the added Lorentzian peak, at bandwidths typical of actual SOAEs. Repeated over all 60 spectra, the total number of peak additions was about 7500 for each condition for each combination of bandwidth and amplitude. The number of

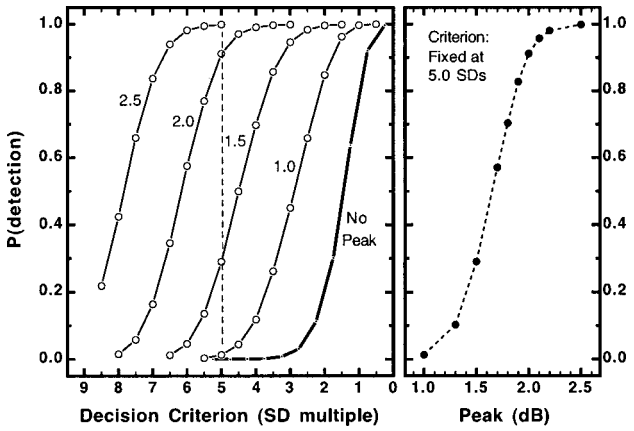


FIG. 4. (Left): The probability of a Lorentzian peak added to the initial spectrum being classified as an SOAE as a function of the decision criterion adopted (expressed as a multiple of the SD of peak-deviation scores). The parameter on the curves is the height of the added peak (dB), and the curve marked ''no peak'' corresponds to a peak height of 0 dB. (Right): The probability of a peak being classified as an SOAE as a function of peak magnitude for a fixed decision criterion of 5.0 SDs. The bandwidth of the added peak at half-maximum was 10.8 Hz for each curve.

times the addition resulted in a detection divided by the total number of peak additions provided an estimate of the probability of detection of an SOAE of the selected amplitude and bandwidth.[6]

The peak addition was performed in the following way. Our desire was to find a Lorentzian peak that when added to the spectrum produced a simulated SOAE rising a fixed number of decibels above the noise floor. The noise floor in the range of peak addition was estimated by computing the mean of the spectral points in the smoothed baseline over that range, in units of power. The height of the desired peak in decibels was converted to a linear (power) increment above the noise floor. This provided the height of the resulting peak in power units, called the *target peak power*. The noise floor power was subtracted from the target peak power to obtain the height of the actual Lorentzian peak to be added to the initial spectrum. The Lorentzian at that height and a selected bandwidth was computed, over a range of 200 bins, and added to the initial spectrum, in units of power. After conversion back to decibels, the resulting initial spectrum was submitted to the smoothing and peak-detection algorithm. The 200-bin interval over which the Lorentzian was computed was selected to be large relative to the peak bandwidth in order to eliminate any transient at either end of the interval resulting from peak addition. For the points most remote from the peak's maximum, the increment to the initial spectrum was at most 0.01 dB, for the peak height and bandwidth values used in the simulation.

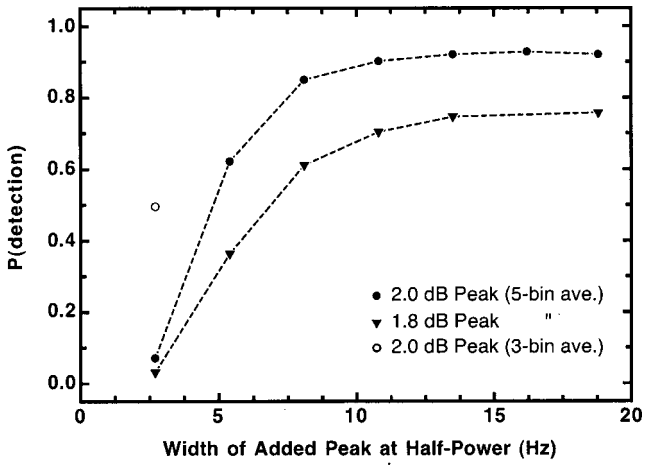


FIG. 5. The probability of an added Lorentzian peak being classified as an SOAE as a function of its bandwidth at half-maximum. Peak height was either 1.8 or 2.0 dB as indicated. Decision criterion was 5.0 SDs. The single open circle is the probability for a 2.0-dB peak when the peak-deviation score is based on 3 spectral bins centered on the peak, instead of 5, as normally used.

For every added Lorentzian peak (and activation of the detection algorithm), the decision criterion for acceptance of a peak as an SOAE was varied, from very lax to strict. Assuming that at least one peak per ear was identified when the criterion was set to its most lax value, the criterion was then adjusted upward in increments of 0.5 SD, with detected peaks recorded at each stage. The process was stopped when the decision criterion value became so strict that no peaks were identified, and the values of the standard-deviation multiples for each such detected peak were recorded.

Figure 4 shows the rate of detection as a function of the magnitude of the decision criterion for Lorentzian peaks of several heights. The curve at the far right in the left panel is the limiting case, where the peak height was zero, and it offers a qualitative estimate of the false-alarm rate. The vertical dotted line indicates a criterion level of 5.0 SDs, noted above to be judged a good conservative compromise for the data of McFadden and Pasanen (1999). Figure 4 shows that the probability of detecting a peak when none in fact was added is asymptotically low for a criterion value of 5.0 SDs, but unfortunately, using real data, it is not practical to obtain a reliable estimate for a false-alarm rate that is so vanishingly small. A different approach to estimating false alarms is described in Sec. D below.

It is useful to compare the plots of detection rates for the range of peak heights where the detection rate is nonzero but not perfect for a selected value of the decision criterion, such as 5.0. A 1.0-dB peak is only rarely identified at that criterion, while a 2.5-dB peak is identified nearly every time. Simulations were run at several peak heights in that range, and the detection rate for a criterion of 5.0 s.d.'s was noted for each peak height. These data are summarized as a probability of detection function in the right-hand panel of Fig. 4.

The likelihood of the added Lorentzian peak being detected depended upon its bandwidth as well as its height, and the effect of this variable is shown in Fig. 5. For the simulation summarized in Fig. 4, the Lorentzian bandwidth was 10.8 Hz (8 bins). For this exercise, the detection rate was measured as the bandwidth at half-power was varied from 2.7 to 18.8 Hz (2 to 14 spectral bins), with the peak height fixed at 1.8 or 2.0 dB. Detection improved as bandwidth increased, and was nearly asymptotic with a bandwidth of 13.5 Hz. Detection performance was clearly poor for the narrowest peak width (2.7 Hz, or 2 bins). However, for the group of heterosexual female subjects studied in McFadden and Pasanen (1999), the mean bandwidth for SOAEs that were less than 3 dB above the noise floor was 13.8 Hz,

TABLE I. Comparison of prevalence of SOAEs and number (standard error) of SOAEs per emitting ear across several studies. Prevalence is expressed as the percent of subjects having at least one emission. The number of subjects in each group is in parentheses. Mean and median numbers of emissions are per emitting ear, with left- and right-ear data pooled. SE (provided when known) is the standard deviation divided by the square root of the number of emitting ears. Data for males and females in the present study are for the heterosexual groups only. The two previous studies summarized are Talmadge *et al.* (1993) and Zhang and Penner (1998).

	Females			Males		
		# Emissions per emitting ear			# Emissions per emitting ear	
	Prevalence (No. subjects)	Mean (SE)	Median	Prevalence (No. subjects)	Mean (SE)	Median
Present study (16K FFTs)	91% ($N=57$)	6.0 (0.40)	6.0	59% ($N=56$)	3.4 (0.38)	2.0
Present study (4k FFTs)	89% (57)	5.9 (0.39)	6.0	59% (56)	3.4 (0.37)	2.0
Present study (32k FFTs)	91% (57)	5.6 (0.38)	5.0	62% (56)	3.2 (0.35)	2.0
Talmadge *et al.* (32k FFTs)	83% (36)	8.3	6.5	63% (40)	4.7	3.0
Zhang and Penner (32k FFTs)	82% (44)	4.1	3.0	68% (39)	2.7	2.0

suggesting that weak SOAEs are rarely as narrow as 2.7 Hz. Of course, it is possible that this observed mean bandwidth was elevated simply because the narrowest emissions of low amplitude were not being detected by the algorithm. Recall that identification is based on the deviation score for a spectral peak, defined as the mean of the 5 spectral bins centered at the peak, and this definition works against the detection of narrow-band peaks. Therefore, the data for that group of female subjects were reanalyzed, using peak-deviation scores based on only 3 points (instead of 5) centered at the spectral peak, with detected peaks fitted by a Lorentzian function to estimate bandwidth. As a further test, the simulation was repeated for a 2-dB narrow-band peak (2.7-Hz bandwidth), using this 3-point definition for the deviation score, and the detection rate is plotted as the single open circle in Fig. 5. In the simulation, detection performance improved, from about 5% to nearly 50%. However, the reanalysis of real subjects' data revealed that, for narrow-band (5.4 Hz or less) SOAEs less than 3 dB above the noise floor, only one additional emission (out of 14) was detected. This reanalysis confirms that very weak SOAEs having extremely narrow bandwidths are not common. Furthermore, basing the peak deviation score on 3 points instead of 5 resulted in poorer overall detection of SOAEs (692 vs 712 total emissions in 114 ears).

D. Estimation of false-alarm rate

It is clear from Fig. 4 that the likelihood of falsely identifying a spectral peak as an SOAE was very low for decision criteria greater than about 3.5 SDs. Unfortunately, the estimation of these extremely low probabilities is not a simple matter. In the simulation, the rate of false detections was less than 0.001 for criterion values of 4.25 or higher (and of course was zero for a criterion value of 5.0 SDs) for the sample of 60 nonemitting ears. A theoretical approach using the distribution of peak-deviation scores for unflagged peaks does lead to a reasonable estimate. This distribution of peak deviations is surely not normal because the range of possible negative values is not unbounded. However, the set of deviations lying above the mean did comprise very nearly 50% of all unflagged peaks, and when this upper half of the distribution of deviation scores was compared to the upper half of a normal distribution, the two were found to be very similar indeed. Specifically, the distributions of peak-deviation scores for the set of 60 nonemitting ears used in the simulation above were normalized and compiled for 16 values of standard-deviation multiple above the mean, in steps of 0.25. The results from all 60 histograms were pooled, and then compared with a normal density function with mean zero and variance 1.0. For the upper halves of the distributions, after combining the extreme bins to obtain a count of at least 1, chi-square was computed for the remaining 14 abscissa values. The chi-square sum was 0.793 ($p>0.999\,98$, $df=11$), revealing that the upper half of the distribution is very close to the normal distribution. Priestley (1981) pointed out that a spectral estimate of the nature described here is in fact asymptotically normal. Assuming that the upper half of the distribution is normal, an estimate of false-alarm rate can be obtained from the area in the tail of the normal density function, for values greater than 5 standard deviations. This area is approximately 2.87E-7, the probability that any single spectral peak would be falsely identified as an SOAE. This is close to the estimate Talmadge *et al.* (1993) reported for their method of requiring five consecutive spectral points lying at least 2 SDs above the noise floor. In our procedure, there are about 1500 peaks in a typical spectrum (16k FFTs), meaning that the expected false-acceptance rate is about 0.000 43 emissions per ear (using a decision criterion of 5.0

SD). If the highest tolerable probability of misidentifying an ear as emitting when in fact it is not were set to 0.005, for example, it follows from the above argument that the SD multiple used for the decision criterion would have to be set at very nearly 4.50.

E. Number of points per FFT

The data, analyses, and parameter selections reported so far in this paper were based on FFTs of 16k points, but full analyses were also performed using FFTs of 4k and 32k points. Summary data comparing the three analyses for the male and female heterosexual subjects from McFadden and Pasanen (1999) are shown in Table I. The prevalence estimates and the number of SOAEs per emitting ear tabled suggest a slightly better detection performance for the 16k FFT analysis than for the 4k analysis, at least for the female group. In addition, the total number of SOAEs detected for both male and female groups (not shown) increased somewhat when the larger window was used. For a recording of fixed length (in this case, 2 min), there is an obvious tradeoff between FFT window size and maximum number of FFTs available to be averaged. However, even when the number of spectra averaged was a fixed proportion of available records (the quietest 25%), the 16k FFT analysis (150 records) afforded a slight advantage over the 4k FFT analysis (300 records). For the group of heterosexual female subjects in McFadden and Pasanen (1999), the 16k analysis found emissions in 84% of the left ears and 88% of the right ears; the 4k analysis found emissions in 79% of the left and 81% of the right ears. The average number of emissions per emitting ear was very slightly higher for the 4k analysis, with 5.87 (4k, left ear) and 5.67 (16k, left ear), with a similar pattern for the right ear. Of course, the additional emitting ears detected by the 16k analysis would most likely have had relatively few emissions per ear, resulting in a reduced average number per emitting ear.

In the third row of Table I are shown the results of the 32k analysis of the data. For comparability of noise criterion, the quietest 75 samples of 32k-point duration were used for the averaged FFTs. The 16k analysis yielded a total of 584 emissions detected for this subgroup of female subjects, and the 32k analysis yielded a total of 555; for males, the emissions totaled 200 and 185, respectively. On the other hand, the prevalence estimate for males increased slightly with the 32k-FFT analysis. There are certainly a number of factors accounting for the reduced sensitivity of the 32k analysis, but a major one must be the reduction of the number of spectra averaged from 150 to 75. It must be added that a full exploration of the parameter space was not performed for the 32k analysis, nor was the simulation involving the added Lorentzian peak conducted. It is possible that parameter settings could be optimized for slight further improvement in SOAE detection.

Many authors favor analyses based on FFTs of 32k points or more, and Talmadge *et al.* (1993) pointed out the clear necessity for the finer frequency resolution afforded by such large window sizes when estimating SOAE bandwidth. Estimates of prevalence of SOAEs and number of SOAEs per emitting ear obtained by Zhang and Penner (1998) and Talmadge *et al.* (1993), both based on 32k FFTs, are also shown in Table I for comparison with our results. The sampling rate used for collecting our data was 22.05 kHz, providing a frequency resolution of 1.35 Hz (16k FFT), identical to that used by Talmadge *et al.* (1993), and similar to that of Zhang and Penner (1998). Although there would be a theoretical advantage from finer spectral resolution in ability to detect weak SOAEs of very narrow bandwidth, it is the strong emissions, not the weak, that typically have the narrowest bandwidths, and their detectability is never in question. In fact, from our experience, very weak narrow-band spectral peaks are often artifacts (harmonics of 60 Hz, for example), and they are fortuitously rejected by our algorithm. Inspection of the averaged spectra from the 32k FFTs revealed that the lower-frequency regions (below about 1 kHz) were often noticeably noisier than the corresponding 16k-FFT averages. Coupled with the higher prevalence estimate for males shown in Table I, this observation suggests that, where feasible, the 32k-FFT analysis might well be preferable, provided that the time record available for analysis is long enough to provide sufficiently many quiet samples (see Sec. G below). From the above we conclude that, for purposes of SOAE *detection*, 16k-FFT analysis using our algorithm performs at least as well as the other procedures cited, and the savings in computation time and memory when using the smaller FFT window may prove to be an advantage in some situations.

The number of emissions found per emitting ear shown in Table I varies somewhat across the three studies, and differences in the rules for SOAE acceptance probably account for some of that variation. Talmadge *et al.* (1993) reported all spectral peaks meeting their detection criterion, irrespective of minimum SOAE frequency spacing or frequency relationships with other SOAEs. Zhang and Penner (1998) apparently discarded emissions which could be distortion products of other emissions. In the present study, we did not omit possible distortion products, but we did eliminate those spectral peaks that were within 0.1 octave of another peak of greater magnitude (about 0.44-mm separation along the basilar membrane using the Greenwood map: Greenwood, 1990, 1991). When we relaxed this requirement (using 16k FFTs), to a minimum frequency spacing of only 3% (0.043 octave), the mean number of emissions increased to 7.3 per emitting ear (median: 6.5) for females, and 4.3 (median: 3.0) for males, in close agreement with Talmadge *et al.* (1993). Relaxing the minimum spacing to 3% provided for elimination of any secondary spectral peaks riding on the flanks of a stronger SOAE, but allowed for inclusion of all ''freestanding'' peaks strong enough to satisfy the decision criterion.

F. Elimination of noisy samples

Detection performance of our algorithm was slightly poorer when analyzing 32k-FFT averages than when analyzing 16k FFTs (see Table I), and possible reasons for this difference were mentioned in the previous section. Halving the number of FFTs averaged produced the expected doubling of the variability of the unflagged spectral peaks about the smoothed baseline, probably obscuring some of the smallest SOAE peaks. However, the elevation in noise floor

observed in the low-frequency regions of many spectra suggests that the pool of sufficiently quiet recording segments was exhausted for some ears. This situation could possibly be improved if a more sophisticated noisy-sample rejection scheme were adopted, such as that of Zhang and Penner (1998) or Talmadge *et al.* (1993). The Talmadge *et al.* (1993) procedure is based on peaks in the original time sample, and uses a different rejection criterion for each of three spectral regions, resulting in a ''composite spectrum,'' where the lowest frequency range is based on the fewest samples averaged (the strictest criterion, the quietest 25%), and the other two higher-frequency ranges are based on successively more samples. The quietest 150 samples (for 16k FFTs) described above corresponds very closely to 25% of the available samples. For the purposes of the McFadden and Pasanen (1999) study, it was not considered appropriate to treat different spectral regions differently because evidence already exists that SOAE prevalence may vary across spectral regions for different populations (Whitehead *et al.*, 1993). Zhang and Penner (1998) used an iterative procedure based on both number of zero crossings and the expected Gaussian distribution of the energy in short segments in their recordings. It is possible that their procedure might produce an improvement over ours, but again, for the purposes of our study, it was considered important to have the averaged spectrum for every subject based on the same number of samples; theirs does not appear to do so.

The requirement of finding windows of 32k contiguous sample points (almost 1.5-s duration) that are sufficiently quiet will always pose a serious difficulty, but an improvement could be achieved by the placement of the middle part of sample windows within those periods in the time record where there is relatively lower noise, and allowing noisier parts to fall in the first or last quarter, where they would be attenuated by the window function. The noise criterion we used for rejection of noisy samples was based on the rms value calculated for each time sample, and this is probably adequate when the samples are reasonably short (0.74 s in our 16k analysis) because the noise generated by a human subject is not likely to occur in extremely short bursts. When the sample duration is much longer, however, the rms value may underestimate the degree of contamination of that sample. Based on our experience, it appears that a recording sample somewhat longer than 2 min would be the simplest solution for those wanting to use 32k FFTs.

Our data were collected mainly from cooperative college-age subjects, and therefore were about as quiet as could be expected. Other subject populations might warrant recorded samples totaling more than 2 min. Our procedure was to record continuously for only 30 s, with a break of 15 s or so between recordings. The subject was told via intercom when a 30-s segment was finished, given about a 10-s break, and then a 5-s warning when a new recording was about to begin (although the new recording could be delayed as long as necessary if the subject was not ready when the warning was given). We believe this procedure reduced the noisy episodes in the recordings considerably. For other subject populations, such as children, shorter recording segments might be advisable.

III. SUMMARY

We have described here an automated procedure for detection of SOAEs. The intention was, first, to provide a more complete description of the procedure used by McFadden and Pasanen (1999) [initially reported by Pasanen and McFadden (1996)], and second, to provide sufficient detail to allow implementation of the procedure by others. Adoption of a standard SOAE-detection procedure, whether ours or not, would be quite helpful in allowing comparisons of data from different laboratories, and would enable collection of SOAE data by nonexpert researchers or clinicians. The procedure described here has several advantages:

(i) It has high sensitivity combined with a low false-acceptance rate.
(ii) It allows direct manipulation of SOAE acceptance criterion over a meaningful range.
(iii) It is computationally simple, and rapid enough to be accomplished while a subject is still in place, allowing for collection of replacement data when the need arises.
(iv) The minimum height needed for a peak to be characterized as an SOAE at a specific level of statistical confidence is set according to the quality of the data collected from that ear.
(v) The smoothing process is tailored to reflect the bandwidth and frequency spacing unique to SOAEs.
(vi) Although there are a number of parameter values to be set, the procedure is quite robust to those choices. Nevertheless, users interested in optimizing the algorithm's performance for specific conditions can change parameter values easily, and exploration of the parameter space to optimize the algorithm's performance can be done in a simple and straightforward manner.

The entire algorithm and other data analyses were programed using the LabVIEW® graphical programming language (National Instruments), but the computations are conceptually simple, and could be easily implemented in other programming environments.

ACKNOWLEDGMENTS

This work was supported by a research grant from the National Institute on Deafness and other Communication Disorders (No. NIDCD 00153). The authors thank C. A. Champlin, L. K. Cormack, and J. C. Loehlin for their helpful comments on an earlier version of the manuscript. Interested readers are invited to contact the authors directly to obtain a LabVIEW® version of the SOAE-detection software.

[1]The bandwidth of an SOAE has typically been taken as the width at half-power of a Lorentzian peak fitted to the SOAE (e.g., Talmadge *et al.*, 1993; Van Dijk and Wit, 1998), but this estimate does not necessarily reflect accurately the width of the SOAE near the base of the peak. The SOAEs of the group of heterosexual female subjects in McFadden and Pasanen (1999) were fitted with Lorentzian functions to estimate bandwidth. Narrow-band SOAEs (5.4 Hz or less) were ignored because the spectral frequency resolution was inadequate for accurate bandwidth estimates. The correlation of SOAE frequency with bandwidth at half-power for the remaining SOAEs was 0.40.

[0-2]Technically, level is the correct term for describing values expressed in decibels sound-pressure level. The term power is used here intentionally as a reminder that the SOAE detection algorithm operated on the power spectrum, and the calculations of SOAE strength were done entirely in power units, with only the final results being converted into decibels.

[3]The 60 spectra described in Sec. II C (each having no SOAEs) were used to verify that, except for absolute values, the low- and high-frequency regions of the initial spectrum were of similar character. The spectrum was split at 3.0 and 5.0 kHz, and the mean and standard deviation of the peak-deviation scores were computed separately for the ranges above and below the split. In both cases, the mean and standard deviation of the peak-deviation scores for the upper- and lower-frequency regions differed by less than 0.02 dB. It is recognized that spectra for individual ears may show greater fluctuations in the low-frequency region as a result of body noise, etc., but this difference is minimized when recordings of sufficient duration are available to allow rejection of noisy samples.

[4]The 19-point median filter replaces each point of the spectrum with the median value of the 19 original bin values centered on the selected point. The 51-point rectangular smoothing window is a filter which replaces each point of the spectrum with the arithmetic mean of the 51 original bin values centered on the selected point.

[5]The actual value of criterion for accepting a peak as an SOAE was obtained separately for each of the six subject groups in McFadden and Pasanen (1999). The mean plus 5 s.d. values ranged across groups from about 1.33 to 1.40 in the left ear, and 1.32 to 1.41 in the right, implying that the distributions of nonflagged peaks were very similar across groups.

[6]To select appropriate bandwidths for simulation, the bandwidth of each SOAE was estimated by fitting a Lorentzian function to that set of spectral points contributing to the total power described above, and using the half-power bandwidth of the fitted function as the estimate of SOAE bandwidth. As Talmadge *et al.* (1993) pointed out, such a bandwidth estimate is only appropriate if it is at least four times the frequency resolution of the spectrum. For the present analysis, with a frequency resolution of 1.35 Hz, it follows that bandwidth estimates below about 5.4 Hz are not accurate. However, for evaluating the detection performance of the algorithm, this bandwidth inaccuracy is not a problem because such narrow-band SOAEs are typically quite strong and easily detected. Our data (see Sec. II C) and the data of Talmadge *et al.* (1993) indicate that weak emissions (within 10 dB of the noise floor) typically have bandwidths substantially greater than 5 or 6 Hz.

Bilger, R., Matthies, M. L., Hammel, D. R., and Demorest, M. E. (**1990**). ''Genetic implications of gender differences in the prevalence of spontaneous otoacoustic emissions,'' J. Speech Hear. Res. **33**, 418–432.

Burns, E. M., Strickland, E. A., Tubis, A., and Jones, K. L. (**1984**). ''Interactions among spontaneous otoacoustic emissions. I. Distortion products and linked emissions,'' Hear. Res. **16**, 271–278.

Greenwood D.D. (**1990**) ''A cochlear frequency-position function for several species–29 years later.'' J. Acoust. Soc. Am. **87**, 2592–2605.

Greenwood, D.D. (**1991**). ''Critical bandwidth and consonance in relation to cochlear frequency-position corrdinates.'' Hear. Res. **54**, 164-208.

Kemp, D.T. (**1979**). ''Evidence of mechanical nonlinearity and frequency selective wave amplificaiton in the cochlea.'' Arch. Oto-Rhino-Larnygol. **224**, 37-45.

McFadden, D., and Pasanen, E. G. (**1999**). ''Spontaneous otoacoustic emissions in heterosexuals, homosexuals, and bisexuals,'' J. Acoust. Soc. Am. **105**, 2401–2413.

Pasanen, E. G., and McFadden, D. (**1996**). ''An automated off-line procedure for detecting SOAEs from a power spectrum,'' J. Acoust. Soc. Am. **100**, 2628(A).

Penner, M. J., Glotzbach, L., and Huang, T. (**1993**). ''Spontaneous otoacoustic emissions: Measurement and data,'' Hear. Res. **68**, 229–237.

Penner, M. J., and Zhang, T. (**1997**). ''Prevalence of spontaneous otoacoustic emissions in adults revisited,'' Hear. Res. **103**, 28–34.

Priestley, M. B. (**1981**). *Spectral Analysis and Time Series* (Academic, London), Vol. 1, pp. 449–471.

Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). ''A review of otoacoustic emissions,'' J. Acoust. Soc. Am. **89**, 2027–2067.

Rao, P., and Bilger, R. (**1991**). ''Spectral instabilities of spontaneous otoacoustic emissions,'' J. Acoust. Soc. Am. **90**, 2289(A).

Talmadge, C. L., Long, G. R., Murphy, W. J., and Tubis, A. (**1993**). ''New off-line method for detecting spontaneous otoacoustic emissions in human subjects,'' Hear. Res. **71**, 170–182.

Van Dijk, P., and Wit, H. P. (**1998**). ''Correlated amplitude fluctuations of spontaneous otoacoustic emissions,'' J. Acoust. Soc. Am. **104**, 336–343.

Van Trees, H. L. (**1968**). *Detection, Estimation, and Modulation Theory* (Wiley, New York), pp. 33–46.

Whitehead, M. L., Kamal, N., Lonsbury-Martin, B. L., and Martin, G. K. (**1993**). ''Spontaneous otoacoustic emissions in different racial groups,'' Scand. Audiol. **22**, 3–10.

Zhang, T., and Penner, M. J. (**1998**). ''A new method for the automated detection of spontaneous otoacoustic emissions embedded in noisy data,'' Hear. Res. **117**, 107–113.

Zwicker, E. (**1990**). ''On the frequency separation of simultaneously evoked otoacoustic emissions' consecutive extrema and its relation to cochlear traveling waves,'' J. Acoust. Soc. Am. **88**, 1639–1641.

On the spectral periodicity of transient-evoked otoacoustic emissions from normal and damaged cochleas[a)]

Paul Avan[b)]
Laboratory of Sensory Biophysics, School of Medicine, Clermont-Ferrand, France

Hero P. Wit
Audiology Department, Akademisch Ziekenhuis Groningen, The Netherlands

Matthieu Guitton
UPRESA CNRS 7060 ''Sensorimotor systems'' and Formation Associée Claude Bernard, Paris, France

Thierry Mom
Laboratory of Sensory Biophysics, School of Medicine, Clermont-Ferrand, France

Pierre Bonfils
UPRESA CNRS 7060 ''Sensorimotor systems'' and Formation Associée Claude Bernard, Paris, France

(Received 3 January 2000; revised 7 June 2000; accepted 19 June 2000)

The spectral quasi-periodicity of transient-evoked otoacoustic emissions (TEOAE) is well acknowledged since Zwicker described a preferred spacing of 0.4 bark between consecutive peaks in the spectrum of otoacoustic emissions from normal ears. While there is scarce evidence of any anatomical reason for this regularity, several functional models of the cochlea have predicted that the structure of emission spectra reflects important characteristics of cochlear filters. In an attempt to check such predictions, the average regularity of TEOAE spectra was studied in three groups of human subjects, normally hearing adults, healthy neonates, and adults suffering from noise-induced hearing loss. Significant differences in emission periodicities were found. Around 1 kHz, the preferred spacing was close to 130 Hz in normally hearing adult ears and neonates. In contrast, no clear periodicity was found in the group of damaged ears, even though they had clinically normal pure-tone audiometry below 2 kHz. Around 4 kHz, the preferred spacing was close to 240 Hz in normal adults and neonates, whereas TEOAEs were absent in many impaired ears. A phenomenological model assuming that TEOAEs stem from the responses of a slightly disarrayed bank of highly tuned filters predicts that the filter width would be the same in healthy young adults and neonates. In contrast, ears suffering from high-frequency hearing loss could exhibit early damaged filters. The proposed method might provide an objective assessment of parameters otherwise difficult to evaluate, especially in neonatal cochleas. [S0001-4966(00)04209-0]

PACS numbers: 43.64.Jb, 43.64.Bt [BLM]

I. INTRODUCTION

The description of the cochlea as a bank of filters is more than a century old (Helmholtz, 1862). Gold (1948) pointed out that an active mechanical feedback loop is indispensable for the high sensitivity and tuning of the normal cochlea to be achieved. After Davis (1983) coined the term ''cochlear amplifier'' to take into account the idea of activity, the basic feedback mechanism proved to originate from the bidirectional transduction occurring in the electromotile outer hair cells (OHCs) (Weiss *et al.*, 1985; Brownell *et al.*, 1985). Length and stiffness changes of OHCs occur when their membrane potential varies and likely enhance basilar membrane vibrations in a frequency-selective manner (Dallos and Evans, 1995). Thus, the bank of filters representing a healthy cochlea is a highly tuned one. Otoacoustic emissions (Kemp, 1978) are thought to represent an acoustical by-product of the feedback processes, and particularly of the electromechanical stage of OHC transduction, because this stage entails injection of mechanical energy in the organ of Corti and cochlear fluids (Brownell, 1990). From this standpoint, transient-evoked otoacoustic emissions (TEOAEs) elicited by broadband clicks can be thought of as a combination of ringing, delayed responses from the excited cochlear filters. TEOAE recordings have thus become a popular clinical tool to assess the function of cochlear OHCs, i.e., the key elements of cochlear sensitivity and filtering, especially in neonates (Vohr *et al.*, 1998; Aidan *et al.*, 1999).

However, Gold (1989) raised an intriguing paradox by pointing out that, for simple mathematical reasons developed by Wit *et al.* (1994b), the overall output of a bank of overlapping filters with similar shapes and regularly spaced center frequencies is zero; in other words, such a system cannot emit TEOAEs even though every individual filter does generate a delayed emission. This phenomenon is due to mutual cancellations of the oscillating contributions of neighboring

[a)]The methodology of this study was presented as a poster at the 21st ARO Midwinter meeting in St. Petersburg Beach, FL, in February 1998.

[b)]Author to whom correspondence should be addressed. Electronic mail: paul.avan@u-clermont1.fr

filters. However, if some degree of disorder was introduced in the array of filters, TEOAE-looking oscillations would be observed at the output of the filter bank. As it is a common observation that the temporal and spectral characteristics of human TEOAEs do exhibit a conspicuous degree of irregularity and idiosyncrasy, the idea that anatomical or functional irregularities are required in addition to activity has been investigated by Kemp (1979), Wilson (1980), Sutton and Wilson (1981), Ruggero *et al.* (1983), Zwicker (1986), Kemp (1986), and Hilger *et al.* (1995).

In apparent contradiction with the previous line of reasoning, several features of TEOAEs exhibit a striking regularity. It is well acknowledged that the peaks and troughs in the complex frequency spectrum of TEOAEs are regularly spaced (Shera and Zweig, 1995). A similar quasi-periodicity is found in stimulus-frequency otoacoustic emissions: about 0.4 bark according to Zwicker and Schloth (1984), i.e., 65 Hz around 1000 Hz. Manley (1983) pointed out that anatomical irregularities in OHC patterns must exist in a coiled cochlea and may be somewhat periodic. Alternatively, Shera and Zweig (1995) proposed an elaborate analytical model of cochlear mechanics that reconciled the two ideas of regularity and disorder. They showed that a ''corrugated'' cochlea can generate otoacoustic emissions with quasi-periodic spectra provided some cochlear amplifier is at work. A simpler phenomenological model was proposed by Wit *et al.* (1994a, b), who considered TEOAEs as the result of combinations of ringing outputs from a bank of highly tuned, linear, and independent filters with randomly misaligned center frequencies. They showed that a disarrayed filter bank generates realistic TEOAE-looking responses to transients and that these responses do exhibit a quasi-periodic spectrum as observed in real ears. Interestingly, these models predict that the characteristics of TEOAE spectral periodicity are closely related to the performance of the cochlear filter. It comes out that the preferred frequency spacing of TEOAE spectral peaks should increase when the filter width decreases, in other words, when the tuning is better.

Thus, the goal of the present study was to perform a detailed study of TEOAE spectra in sizeable homogeneous samples of human ears in an attempt to derive information concerning cochlear-filter characteristics in these samples. A constraint of the foregoing models is that the filter characteristics are more easily extracted from *preferred* frequency spacings, and not from individual ones. In other words, the intended analysis must ultimately process TEOAE data after averaging across ears. Three groups, as homogeneous as possible, were defined; young adults with normal pure-tone audiograms ($n=28$); healthy neonates ($n=61$): and adults with mild occupational noise-induced hearing loss at high frequencies ($n=40$). Individual frequency spacings of peaks in TEOAE spectra were determined around 1 and 3 kHz, then averaged across the ears of each group. Neonates were compared to normal young adults around 3 kHz, whereas many ears with noise-induced hearing loss did not have TEOAEs at this frequency. Comparisons around 1 kHz were performed across all three groups since all the ears with noise-induced hearing loss had strictly normal pure-tone thresholds at 1 kHz and their TEOAEs were present in this frequency range.

II. EXPERIMENTAL METHOD

The right and left ears of three series of subjects were examined. Audiometric tests and TEOAE recordings were done, then a selection was made to exclude some ears, according to the outcome of TEOAEs. Let N be the initial number of ears in each group, and n the remaining number after selection. Young, normally hearing adults (age ranging from 22 to 30 y; $N=37$ ears) composed the first series, labeled AD. Their pure-tone audiogram was controlled and their auditory thresholds were $\leqslant$15 dB HL at all octave frequencies $\leqslant$8 kHz. The second series, NN, included healthy neonates born in due term who were tested in a well-baby unit at day 2 or 3 after birth as part of a systematic hearing-screening procedure ($N=200$ ears). None of them had a known risk factor for hearing, and none of them has been reported to have sensorineural hearing loss since the time of data collection in 1996. The third series, NIHL, was made of adults (age ranging from 29 to 54 y) referred for occupational noise-induced hearing loss ($N=56$ ears). Adult subjects were tested in a sound-treated booth. They all underwent otoscopy, tympanometry, and pure-tone audiometry at octave frequencies from 0.125 to 8 kHz, and at 6 kHz. The tests compiled with the French regulations of the Department of Health and with the principles of the declaration of Helsinki.

In all subjects, TEOAEs were recorded with a ILO88 Otodynamics® equipment, as described by Kemp *et al.* (1990), using clicks at 70 dB SPL peak-equivalent. The so-called linear mode was used, i.e., all clicks were sent with the same level and polarity, with a repetition rate of 33/s, and 260 noiseless epochs were averaged. The recording window was set at 2.5–30 ms in order to better assess the presence of long-lasting TEOAE components, and the sampling rate was such that 512 bins were collected [example in Fig. 1(a)]. The complex frequency spectrum of TEOAEs was computed by fast Fourier transform (FFT) and the same magnitude spectrum as displayed by the ILO88 screen could be obtained offline [Fig. 1(b)]. In addition, the real and imaginary parts were displayed separately [Fig. 1(c), for the real part]. The periodicity of their peaks and troughs was analyzed by computing another FFT [Fig. 1(d)]. Figure 1(d) shows that the resulting spectrum exhibited peaks, reflecting the predominant frequency spacings visible in the TEOAE spectrum of Fig. 1(c). Recall that the scale of the horizontal axis of TEOAE spectra is in Hz, hence the scale of the horizontal axis coming out of the second FFT is directly obtained in Hz^{-1}. For the sake of clarity, an inverse transformation was applied to horizontal-axis labels so that the spacing between adjacent TEOAE components is given in Hz [Fig. 1(d)]. The smallest possible spacing is the Nyquist limit 66 Hz, i.e., twice the 33-Hz frequency resolution of the first FFT spectrum. Notice that the cepstrum analysis, a standard method for detecting sidebands and harmonic patterns in complex signals, also involves the computation of two FFTs in a row since the cepstrum is defined as the power spectrum of the logarithm of a power spectrum (Bogert *et al.*, 1963).

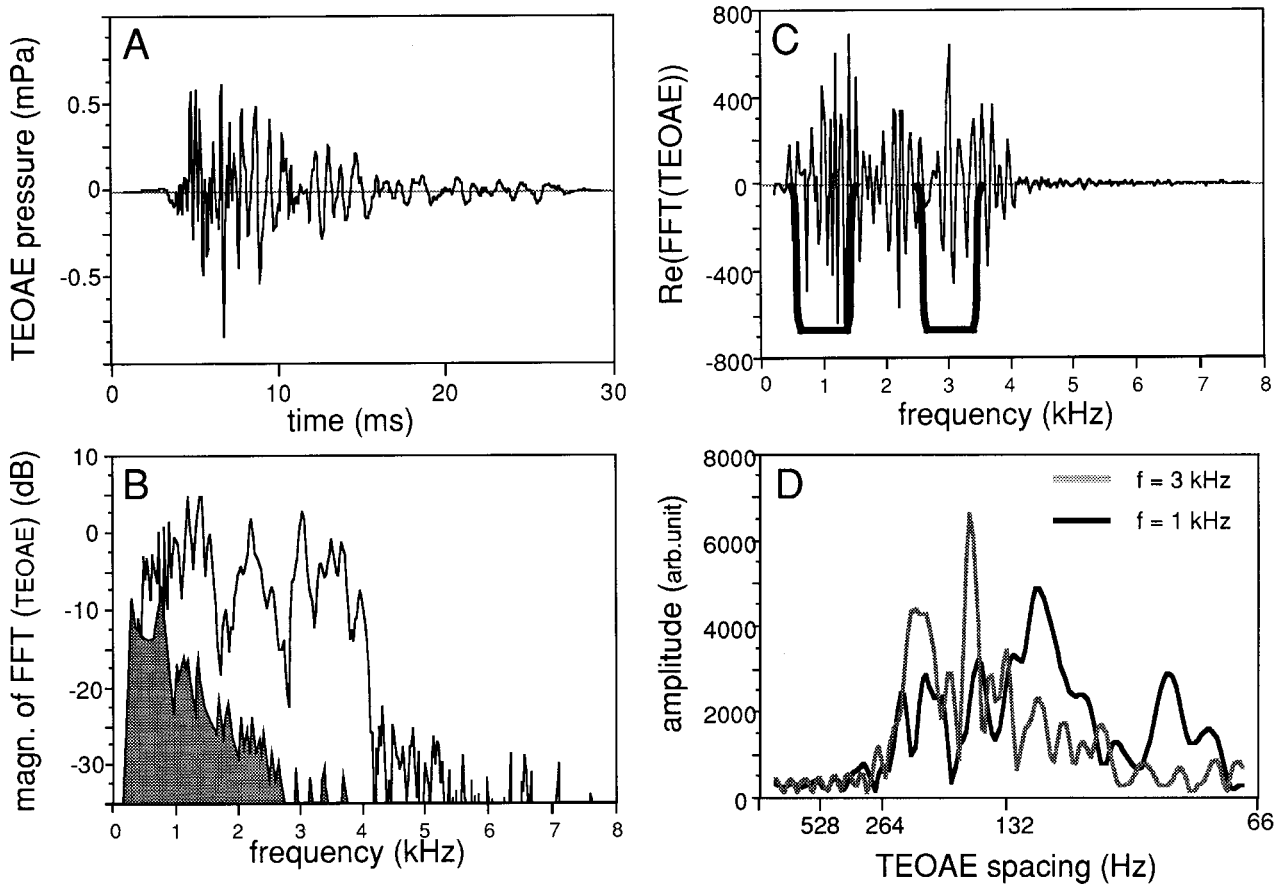


FIG. 1. (a) Example of TEOAE in the time domain, in a healthy adult ear. Little or no energy is present after 20 ms with respect to click onset. (b) Standard magnitude spectrum of this emission obtained by FFT (noise background: shaded area). (c) Real part of the FFT of this TEOAE (vertical scale: arbitrary unit), revealing a regular pattern of peaks and troughs. The periodicity of this pattern was analyzed within the two windows drawn in bold lines. (d) FFTs of spectrum C after its windowing around $f=1$ kHz (bold line) and $f=3$ kHz (shaded line). The horizontal axis derived from the second FFT was directly scaled in Hz^{-1} and its labels have been converted into Hz values to indicate the periodicity of diagram C in a more straightforward manner.

The late temporal components of TEOAEs were considered as synchronized spontaneous otoacoustic emissions (Probst *et al.*, 1986). They were identified as one or several reproducible peaks that remained in the spectrum of the late time-windowed TEOAE (trapezoidal window, 0.5-ms rise and fall, 5-ms plateau, between 24 and 30 ms after click onset—thick lines in Fig. 4, top diagram). It is likely that the case of ears with spontaneous otoacoustic emissions (SOAE) cannot be described correctly by a bank of linear, independent filters: since SOAEs exist in the absence of any external stimulus, they represent a highly nonlinear regime of the cochlear filters. Linearity is required by the model of Wit *et al.* (1994b) that served as a background for quantitative subsequent interpretations. Furthermore, in Sec. V we will show that the presence of SOAEs embedded in a TEOAE spectrum generates large, atypical irregularities in its quasi-periodical pattern. For all these reasons, we chose to discard all the ears presenting synchronized SOAEs. The final analysis of TEOAE periodicities was carried out on $n=28$ ears of series AD, $n=61$ ears of series NN, and $n=40$ ears of series NIHL. Only the results from these ears will be presented thoroughly in Sec. V. A few results will be shown from series NNS+ made of the $n=139$ ears excluded from series NN owing to the presence of SOAEs.

III. THEORETICAL BACKGROUND

Though we acknowledge that the model proposed by Wit *et al.* (1994b) is by no means taken for granted, we used it as a support for quantitative analysis of TEOAE data, whereas the qualitative analysis was restricted to determining the preferred frequency spacing of TEOAE spectral peaks in every ear and comparing it from one series to another. It is therefore useful to briefly recall the procedures proposed by Wit *et al.* (1994b) to perform a simple analytical derivation of the shape of filters from TEOAE spectra. Let $z(f_k,f)$ be the response of the kth filter of the bank with center frequency f_k, to a signal at frequency f. Actually z is assumed to be only a function of $f-f_k$, which means that all filters are similar. Assume that the center frequency of filter $\#k$ is shifted by $\varepsilon(f_k)$, with ε being a small random amount, so that $\varepsilon(f_k)$ represents the irregularity function of the filter bank. It is easy to show (Wit *et al.*, 1994b) that the output $\mathcal{O}(f)$ of the irregular filter bank excited at frequency f is

$$\mathcal{O}(f)=\int z'(f-x)\cdot\varepsilon(x)\,dx. \tag{1}$$

This formula represents a convolution (symbol $\otimes$), i.e., $\mathcal{O}(f)=z'\otimes\varepsilon=z\otimes\varepsilon'$.

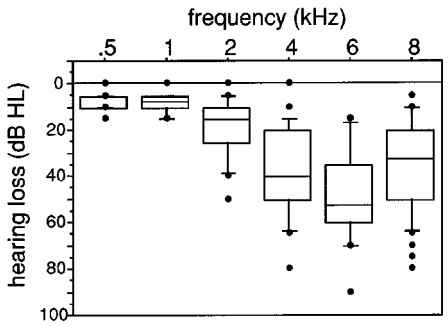


FIG. 2. Boxplot distribution of pure-tone auditory thresholds in the ears with mild noise-exposed hearing loss of occupational origin (sample of 40 ears). The upper and lower limits of the rectangular boxes represent the 25th and 75th percentiles, the inner horizontal line marks the median. The lines projecting from the boxes refer to 10th and 90th percentiles and the dots mark outlying data.

Thus, the output can be interpreted as being the result of filtering the derivative of the irregularity function, ε', by the Fourier transform of the filter shape z, or alternatively, filtering the irregularity function by the Fourier transform of the derivative of the filter function. As a consequence of this filtering, the Fourier transforms of either the real or the imaginary part of the filter output (i.e., the TEOAE) are more or less periodical: preferred frequency spacings between consecutive peaks or troughs should be observed even though the existence of ε blurs the individual frequency patterns. Another way to present the same result consists in computing the Fourier transform of FFT $(\mathrm{Re}(\mathcal{O}))$ and checking that it tends to present a peak, as done in the present work. The data analysis has to account for a serious drawback of this theory. Individual functions $\varepsilon(f_k)$ are unknown, thus individual patterns FFT(FFT(Re($\mathcal{O}$))) should look ''noisy'' and cannot be expected to reveal clearly the filter shape z. However, the unknown contribution of ε may be eliminated by averaging FFT(FFT(Re($\mathcal{O}$))) across many ears, provided z is the same for all the ears in the sample. If this assumption is valid, the averaging process should reveal a smooth estimation of FFT(FFT(Re(z'))).

IV. DATA PROCESSING

According to the previously exposed model, although individual patterns of TEOAE spectra are ''noisy'' because of the influence of ε, some important features of filter z' are embedded therein. To gain information on z', data around 1 and 3 kHz were processed separately after windowing the TEOAE spectrum FFT(Re(TEOAE)) by a 1-kHz-wide window centered on either frequency [cosine-squared edges, 0.2 kHz wide; flat plateau, 0.6 kHz wide; window profiles in Fig. 1(c)]. Next, the second FFT was computed within each window. The frequency spacing fp corresponding to the highest peak in FFT(FFT(Re(TEOAE))) was measured around 1 and 3 kHz. Assuming that the shape of this last Fourier transform was roughly that of a bandpass-filter transfer function with peak amplitude Ap, the low- and high-frequency-spacing limits corresponding to a 6-dB bandwidth, i.e., the smallest and largest frequency spacings $f$min and $f$max with amplitude Ap/2, were measured. The goal of this individual data processing was therefore to bring out a few hints of the existence of a bandpass filter at the origin of TEOAEs: for example, the distributions of $f$max, fp, and $f$min would be expected to get shifted or broadened in a group with impaired bandpass filters due to cochlear damage. Unfortunately, the presence of ε precluded any more direct study of the filter. In order to eliminate ε, FFT(FFT(Re(TEOAE))) was averaged in every group around 1 and 3 kHz, and the obtained smoothed patterns were compared. Across-ear invariance of z inside a group had to be assumed for the procedure to be valid. Although likely in normal ears NN and AD, this may not have been true in NIHL ears because the degree and frequency range of impairment varied from ear to ear. Instead of revealing FFT(FFT(Re(z'))), improper averaging would have destroyed periodicity information.

V. RESULTS

A. Audiometric data

In the two groups of adult ears, otoscopy and tympanometry were always normal. In the group of normally hearing subjects AD, all pure-tone auditory thresholds were $\leqslant$15 dB HL at every tested frequency up to 8 kHz. In the noise-exposed ears, audiometric notches of varying depth and extent were found. All these ears presented clinically normal pure-tone auditory thresholds up to 1 kHz included (i.e., $\leqslant$15 dB HL). The maximum notch depth varied between 20 and 90 dB HL (average 48 dB HL, median 45 dB HL), and was reached around 4 or 6 kHz in all cases. The distributions of pure-tone auditory thresholds are plotted in Fig. 2 between 0.5 and 8 kHz. They reflect the typical pattern expected for mild noise-induced hearing loss.

B. Basic TEOAE data

The percentages of ears that had to be discarded due to the presence of one or several synchronized SOAEs were 19% in group AD, 69% in group NN, and 0% in group NIHL. The characteristics of TEOAEs in these three groups presented distinctive features concerning the shape of frequency spectra and the levels in the different frequency bands. In the first place, the TEOAEs found in the group of noise-impaired ears either had a low amplitude (in 18 ears) or were absent (in 22 ears) at frequencies $>$2.5 kHz, as expected for ears with high-frequency hearing losses. Furthermore, the spectra tended to peak at higher frequencies in neonates (Fig. 3, top), namely around 2.2 kHz on average (s.d. 0.79 kHz), than in normally hearing adult ears that peaked around 1.16 kHz (s.d. 0.30). Noise-damaged ears did not differ from AD ones in this respect, since the best TEOAE component was at 1.16 kHz on average (s.d. 0.50). An ANOVA test indicated that this difference across groups was significant ($p>0.0001$) and the *post hoc* test of Scheffé suggested that the neonate group was responsible for this difference ($p<0.0001$).

The levels of TEOAEs also differed across groups, with NN ears exhibiting a larger average level (19.1 dB SPL peak eq.; s.d. 3.7). The average TEOAE level in AD ears was intermediate (10.7 dB SPL peak eq.; s.d. 2.9), and the average TEOAE level in NIHL ears was the smallest (5.5 dB SPL peak eq.; s.d. 4.3). An ANOVA showed that these dif-

ferences were significant ($p<0.0001$) and the *post hoc* test of Scheffé suggested that all pairs of groups differed ($p<0.0001$). The same type of comparison was extended to bandpass-filtered TEOAEs, around $f=1$ and 3 kHz (Butterworth eight-pole filter, 1 kHz width in order to match the preferred frequency spacing analysis presented further). The results are presented in Fig. 3 (middle diagram) for $f=1$ kHz and Fig. 3 (bottom) for $f=3$ kHz. Neonatal ears gave rise to the largest average TEOAEs in the two frequency bands, while the average TEOAEs in AD ears were 5.8 dB smaller at $f=1$ kHz (and 13.3 dB smaller at $f=3$ kHz). The average TEOAEs in NIHL ears were the smallest, i.e., 4.9 dB smaller than normal adults at fc=1 kHz, whereas TEOAEs were absent at fc=3 kHz in too many ears for the analysis to be meaningful. An ANOVA showed that the differences were significant for the two fc values ($p<0.0001$), and the *post hoc* tests of Scheffé showed that all pairs of groups differed ($p<0.0001$).

C. Preferred frequency spacings in TEOAE spectra

The preferred frequency spacing analysis was performed with group AD serving as reference. A typical example of AD TEOAEs recorded in a 2.5–30-ms window is represented in Fig. 1(a). No long-lasting ringing was visible after about 20 ms. Figure 1(b) depicts the magnitude spectrum resulting from the first FFT, as traditionally displayed in clinical tests. In this ear, many spectral peaks were found between 0.7 and 4 kHz. The spectrum of background noise was derived from the comparison of buffers A and B provided by the ILO88 software [shaded area in Fig. 1(b)], following Kemp *et al.* (1990). Signal-to-noise ratios largely exceeded 10 dB for most TEOAE spectral peaks in all ears. The real part of the FFT is plotted in Fig. 1(c). A clear periodicity of the peaks and troughs was visible on this diagram, with the interval between two consecutive peaks increasing rather steadily from 1 to 4 kHz. Similar frequency-dependent periodicities were observed in the TEOAE of the other ears. The spectral periodicities of the peaks found in the real part of the spectrum were disclosed in a more straightforward manner by computing a second FFT, after windowing the diagram of Fig. 1(c) by a 1-kHz-wide window centered either on $f=1$ or 3 kHz [bold lines, Fig. 1(c); outcome in Fig. 1(d), 1-kHz center frequency, bold line, 3 kHz, shaded line]. On this example, the FFT analysis confirmed that the frequency spacing between peaks was quite regular, and the spectrum computed around 1 kHz exhibited one dominant peak [bold line, Fig. 1(d); Ap=4980 arbitrary units, corresponding to a preferred spacing fp=118 Hz]. However, other smaller peaks were also visible so that the pattern of this ear had a ''noisy'' aspect. The bandpass filter limits $f$max and $f$min as defined in Sec. II were 244 and 82 Hz, respectively. Around 3 kHz, two main peaks of preferred spacing were found around 160 and 250 Hz.

Figure 4 depicts another example of a TEOAE in a normal adult ear, showing the presence of a long-lasting component (around 1.2 kHz; top diagram). This component most likely corresponded to a synchronized SOAE. The structure of peaks and troughs of the real part of the Fourier transform

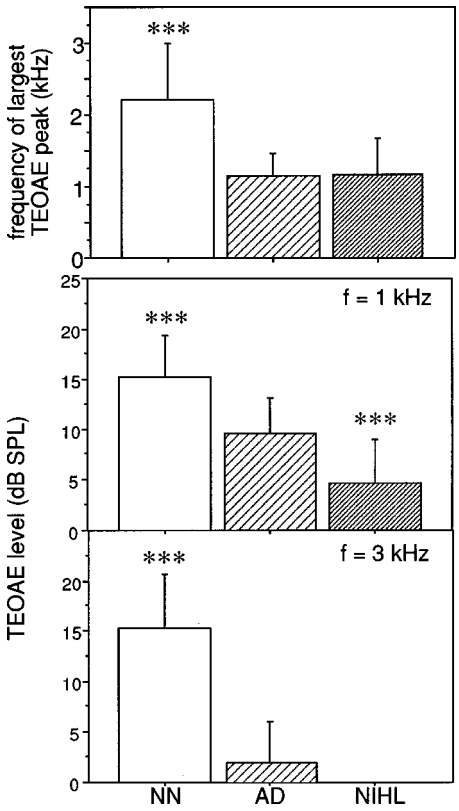


FIG. 3. (top) Average frequency of the highest TEOAE spectral peak across groups NN, AD, and NIHL (length of bars: one s.d.). The stars indicate the groups identified as different from AD according to ANOVA and *post hoc* Scheffé analysis ($p<0.0001$). (middle and bottom diagrams) TEOAE levels across groups in 1-kHz frequency bands centered around $f=1$ and $f=3$ kHz (same symbols as in top diagram). The stars indicate the groups that significantly differed from AD after ANOVA and Scheffé tests ($p<0.0001$).

of the TEOAE seemed disturbed in the vicinity of 1.2 kHz (middle diagram). Its FFT around 1 kHz confirmed that the obtained pattern of preferred frequency spacings was highly disturbed with several peaks of similar magnitude around 70, 80, 100, and 150 Hz.

For AD ears, the distributions of $f$max, fp, and $f$min around 1 kHz are represented in Fig. 5 as boxplots depicting the medians, 25th and 75th percentiles, borderline and outlying values. NN results showed for comparison tended to have similar $f$max and fp distributions, whereas $f$min tended to be systematically a little smaller. Around 3 kHz, all three distributions were very similar in NN versus AD groups. The quantitative analysis of these distributions was performed with an ANOVA after including NIHL ears and will be described next.

After averaging the results obtained from the different samples of ears, the contributions of the irregularity functions were smoothed out. According to section methods, the results with thought to provide an estimate of the Fourier transform of the derivative of the filter function in the frequency domain. For the frequencies around 1 kHz, the result of this average procedure is represented in Fig. 6 for neonates without SOAE (top) and adults (bottom) (average, thick line, average ± one s.d., thin lines). For the frequencies around 3 kHz, the results are presented in Fig. 7. One peak

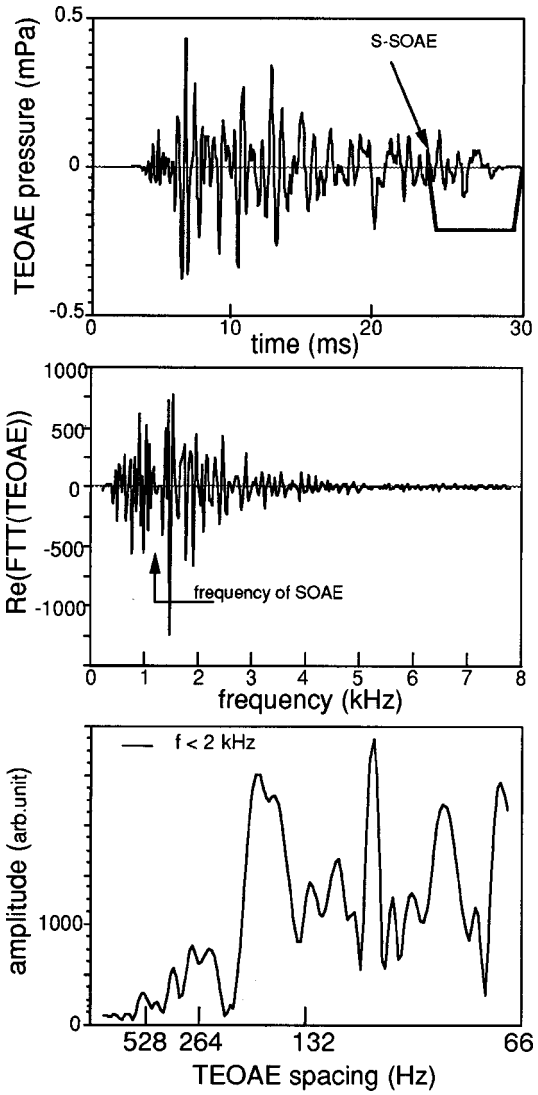


FIG. 4. (top) Example of sustained TEOAE, with a component around 1.2 kHz lasting more than 28 ms. This frequency likely corresponds to a SOAE in this ear, synchronized by the click stimulus. Trapeze in bold lines, 24–30 ms: time window for synchronized-SOAE research. (middle) The real part of the FFT of this TEOAE shows large irregularities in the pattern of peaks and troughs around the frequency of the SOAE. (bottom) The resulting pattern of spacings between adjacent TEOAE peaks around 1 kHz seems highly disorganized. This ear was not included in the analysis because the proposed model could not take into account the presence of SOAEs.

corresponding to the preferred frequency spacing of TEOAE components is visible on each diagram. For example, at 1 kHz, the preferred frequency spacing was close to 130 Hz in adults and neonates. A secondary peak was visible around 70 Hz, in keeping with the results reported by Zwicker and Schloth (1984). At 3 kHz, the preferred spacing increased to about 230 Hz. The plots corresponding to average ± one s.d. tended to peak at the same frequencies as the average plots, except at 1 kHz for AD ears: the highest peak of the plot ''average + one s.d.'' was at 260 Hz, whereas the peak close to 130 Hz was only the second highest one. The average results of group NNS+ ($n=139$ ears with synchronized SOAEs) are superimposed on those of group NN in Figs. 6 and 7 so as to check the influence of one or a few synchronized SOAEs on the estimation of FFT(FFT(Re(z'))). Apart from an increased emission amplitude in NNS+ (+3.0 and

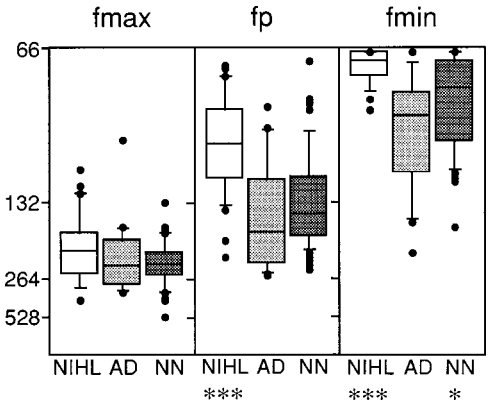


FIG. 5. Distributions of fp, frequency of the peak in the spacing pattern of individual TEOAEs (center diagram), and $f$max (left panel) and $f$min (right panel), bounding the ''6-dB bandwidth'' of the filter allegedly accounting for the shape of the spacing pattern. Each panel compares the distributions across groups (labels of horizontal axis) using boxplots (see legend of Fig. 3). The stars identify the groups that significantly differed from AD after ANOVA and Scheffé tests.

+3.6 dB relative to NN, see the different vertical scale on the right part of Figs. 6 and 7, top diagrams), the averaged FFTs of NNS+ ears were virtually identical to those of NN ears, suggesting that the presence of SOAEs, overall, bore little influence on preferred TEOAE spacings.

Figure 8 depicts an example of data processing in a noise-exposed ear. The TEOAE response [Fig. 8(a)] exhibited a small amplitude and no clear high-frequency compo-

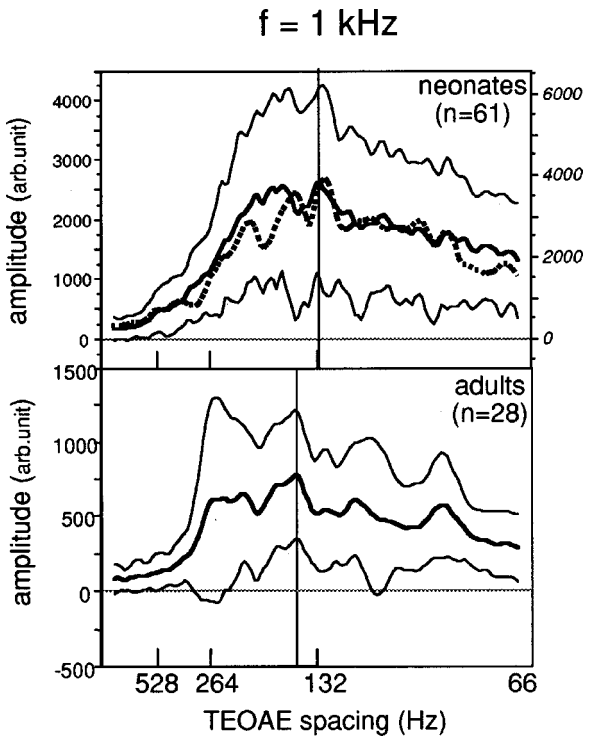


FIG. 6. Average spacing (bold continuous line), and average ± one s.d. (thin lines) between adjacent TEOAE components around 1 kHz, in groups NN (top) and AD (bottom). The corresponding vertical scale is on the left side. A relatively smooth plot is obtained in both cases, although the plots of adult ears tend to present marked secondary peaks. The vertical lines indicate the preferred frequency spacing, around 130 Hz in both samples. Dotted line: average spacing in group NNS+ ($n=139$ ears), corresponding vertical scale on the right side in italics.

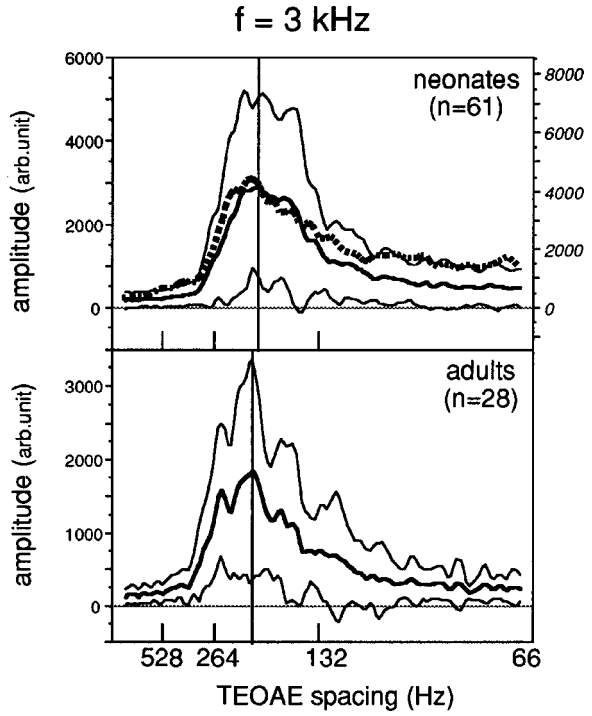


FIG. 7. Similar to Fig. 6, for frequency 3 kHz.

nent. The magnitude spectrum [Fig. 8(b)] confirmed the absence of TEOAE above 3 kHz, whereas the TEOAE components around 1 kHz were well above the noise floor. The real part of FFT [Fig. 8(c)] was highly irregular and its FFT [Fig. 8(d)] did not reveal any dominant frequency spacing. Such a purely ''noisy,'' broadband pattern was found in 25 ears. The audiogram of the ear in Fig. 8(e) was typical of mild noise-induced hearing loss with normal pure-tone thresholds up to 2 kHz and a 50-dB notch centered at 6 kHz. The TEOAEs of noise-exposed ears could not be analyzed around 3 kHz because they were absent at frequencies >2.5 kHz in 55% ears. For NIHL ears, the distributions of $f$max, fp, and $f$min around 1 kHz are plotted in Fig. 5 where they can be compared to AD and NN distributions. While $f$max's differed little, fp's were shifted toward values lower than normal (ANOVA: $p<0.0001$, with Scheffé *post hoc* test suggesting that NIHL ears strongly differed from all others, with $p<0.001$). Even more clearly, $f$min's were almost always very close to the smallest measurable value of 66 Hz. The ANOVA was significant ($p<0.0001$), due mainly to NIHL ears ($p<0.0001$) and also partly to NN ears ($p=0.03$) as compared with AD ears, according to the Scheffé *post hoc* test.

Around 1 kHz, the result of the averaging procedure hardly revealed any common preferred spacing for the observed TEOAEs (Fig. 9; average, thick line, average ± one s.d., thin lines). According to Fig. 5, individual TEOAEs in NIHL ears seemed to exhibit some degree of preferred spacing, and the plot corresponding to average + one s.d. indeed revealed a peak below 66 Hz (Fig. 9, dashed line). The average and average − one s.d. plots exhibited no clear peak, suggesting two possible interpretations. The preferred spacing may have differed from one NIHL ear to another in such a way that the averaging procedure blurred the resulting plot instead of just smoothing out the ''noisy'' contributions. Alternatively, a number of individual frequency patterns may have been purely ''noisy'' without any underlying bandpass filter shape, thus rendering doubtful the measurement of fp, $f$max, and $f$min in these cases.

VI. DISCUSSION

For the time being, the overall significance of TEOAEs is widely acknowledged, together with the relation of TEOAE absence to hearing sensitivity. TEOAEs tend to disappear in and around frequency bands where a sensorineural hearing loss >30 dB is found, provided the loss is due to cochlear outer hair cell impairment (review in Probst *et al.*, 1991). This finding strongly supports the use of TEOAEs for neonatal hearing screening (Vohr *et al.*, 1998; Aidan *et al.*, 1999). Our current knowledge of cochlear physiology implies that hearing sensitivity is largely due to the same feed back mechanisms ensuring frequency tuning. Hence, the two key functions of the cochlea, namely sensitivity and filtering, are impaired in a consistent manner whenever some part of the cochlear feedback loop is damaged (e.g., Liberman and Dodds, 1984; Ruggero and Rich, 1991; Patuzzi, 1996). Since TEOAEs result from precisely the same feedback loop, they should provide a fine insight, not only into the hearing sensitivity, but also into the attendant filtering processes. Unfortunately, few noninvasive methods are available to test the performance of cochlear filters (e.g., Pick *et al.*, 1977; Glasberg and Moore, 1986; Tyler, 1986) and their complexity precludes their use in neonates. Thus it would be useful to extend the analysis of TEOAE characteristics and strive to find at least indirect clues regarding the quality of cochlear filters.

Following Zwicker and Schloth (1984), several authors have reported that normal stimulus frequency otoacoustic emissions evoked by pure tones present a striking spectral regularity with a preferred spacing of about 0.4 bark, that is, 65 Hz around 1 kHz, between two consecutive peaks or troughs (Lonsbury-Martin *et al.*, 1990; Zwicker, 1990; Shera and Zweig, 1995). Scant data are available for TEOAEs, but they are expected to follow a similar rule. Indeed, Zwicker and Schloth (1984) presented the results from one ear where the same frequency spacing of 120 Hz was found between the spectral maxima of stimulus frequency emissions and TEOAEs around 1 kHz. Our results for normal adult ears do not contradict Zwicker and Schloth (1984), who reported large interquartile intervals extending from 50 to 140 Hz at 1 kHz, and the data variance visible in Fig. 6 confirms their observation (see secondary peaks around 70 and 100 Hz). Later, Lonsbury-Martin *et al.* (1990) confirmed the variability of frequency spacings in stimulus frequency otoacoustic emissions. No data has been published so far concerning nonadult or abnormal cochleas. Neonatal cochleas might prove not to be mature in some respects, and even though residual TEOAEs can remain present in noise-damaged cochleas, their properties are likely abnormal, because noise-induced hearing loss primarily involves damage to outer hair cells (Patuzzi, 1991).

The origin of TEOAE quasi-periodicity is still questioned. Impedance discontinuities along the organ of Corti

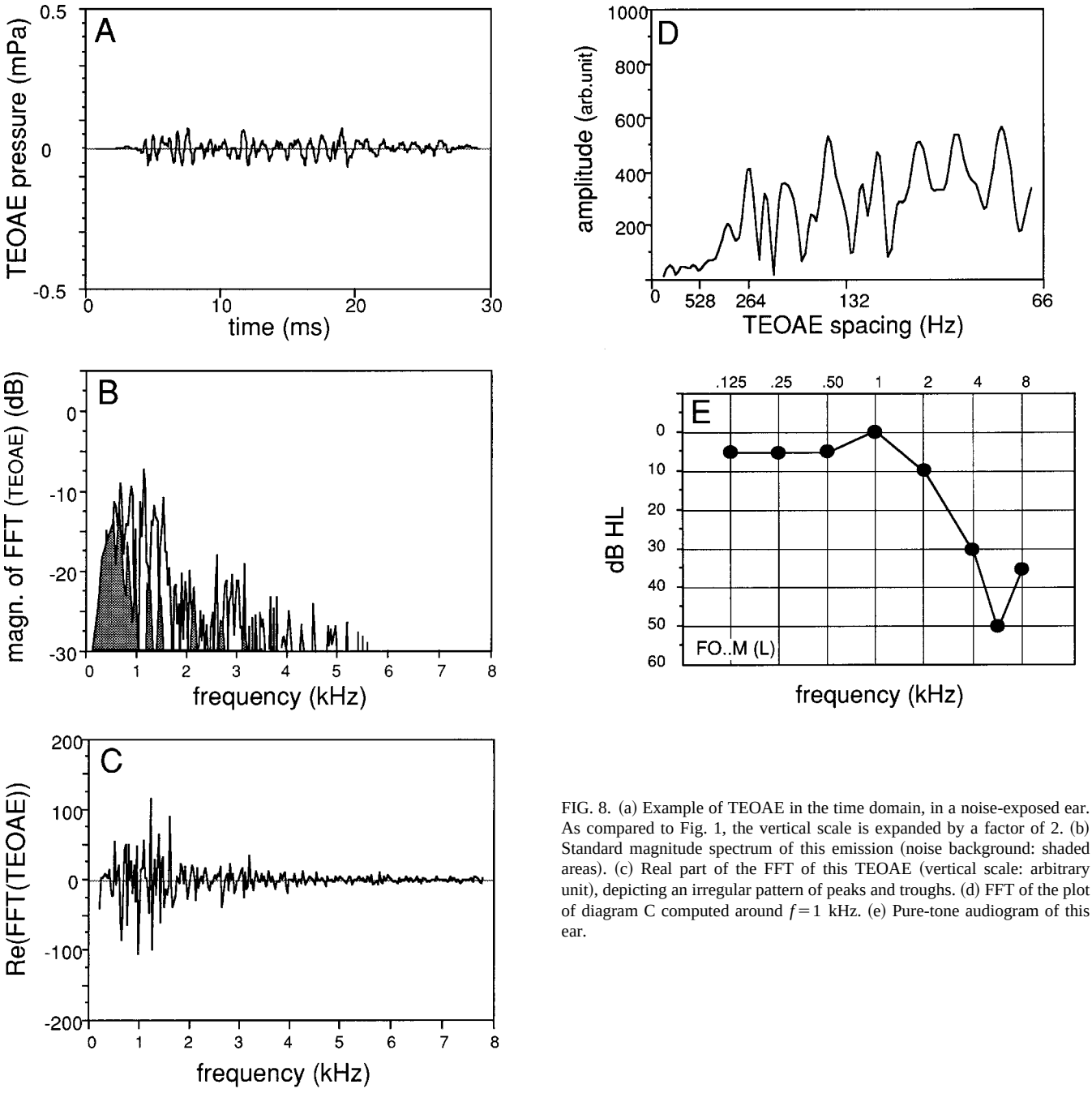


FIG. 8. (a) Example of TEOAE in the time domain, in a noise-exposed ear. As compared to Fig. 1, the vertical scale is expanded by a factor of 2. (b) Standard magnitude spectrum of this emission (noise background: shaded areas). (c) Real part of the FFT of this TEOAE (vertical scale: arbitrary unit), depicting an irregular pattern of peaks and troughs. (d) FFT of the plot of diagram C computed around $f=1$ kHz. (e) Pure-tone audiogram of this ear.

might favor TEOAE production at the corresponding frequencies (Wilson, 1980; Sutton and Wilson, 1981), and Manley (1983) proposed that, for geometrical reasons, regularly spaced discontinuities should exist in a coiled cochlea. Later, the attempts to find relationships between TEOAE spectral peaks and anatomical irregularities in the cochlea have been far from conclusive. Ruggero *et al.* (1983) and Lonsbury-Martin *et al.* (1988) found some evidence of SOAE presence at the best frequency of irregular cochlear places. On the other hand, Hilger *et al.* (1995) failed to observe clear correlations between sensory cell patterns and TEOAE spectra in guinea pig ears. Using a thorough theoretical approach of cochlear micromechanics, Shera and Zweig (1995) proved that what they called ''dynamic symmetry creation'' is possible in an active system exhibiting some degree of random irregularities. Wit *et al.* (1994b) implemented a simpler model starting from Gold's original idea (1948, 1989) and found a good agreement between model predictions and real-ear data. It assumed that the filter shape was invariant along the x axis, with x being a monotonic function of frequency. Actually, the procedure of the present work had to suppose that this function was linear, owing to the lack of knowledge of the required transformation from the frequency axis to the x axis before the FFT of the real part of the FFT of the TEOAE signal was computed. This approximation was probably correct here because after filtering around 1 or 3 kHz, only a small part of the frequency axis was taken into account, and whatever the required f to x transformation, it was expected to deviate little from a linear function in the restricted range of interest. This numerical extension of Wit *et al.*'s model was used to extract what we speculate to be physiologically relevant information

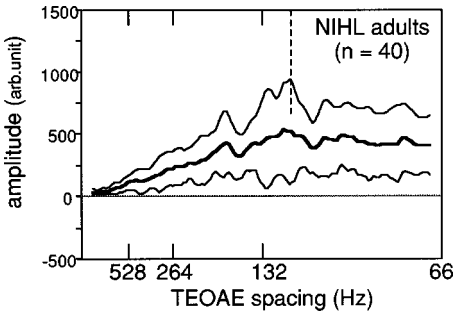


FIG. 9. Average spacing (bold line), and average ± one s.d. (thin lines) between adjacent TEOAE components around 1 kHz, in group NIHL. Plots corresponding to average and average − one s.d. do not present any significant peak. The plot representing average + one s.d. tends to peak around 110 Hz (dashed vertical line).

from the highly complex patterns of TEOAEs, and the three samples of ears were likely to present different physiological characteristics. Regardless of any theoretical interpretation, the present work shows that the preferred spacings for TEOAE spectral components were similar in normal adult ears and normal neonatal ears, and quite different in mildly noise-impaired adult ears.

The model we used here suggests further interpretation of these differences. When the function z characterizing the TEOAE-emitting filters gets broadened, its derivative z' is blunter, thus its Fourier transform is expected to become narrower and to peak at a lower value. Therefore, the observation that the preferred frequency spacing is similar in neonates and normal adults suggests that their filter function is similar, whereas a poorer performance of cochlear filters in the presence of NIHL may be inferred, even in the areas of apparently normal sensitivity. It is difficult to corroborate with psychoacoustical data the outcome of this model predicting that neonatal cochlear filters are similar to adult ones. Most of the properties of the peripheral auditory system are said to be mature at birth, except the resonance properties of the external and middle ear (e.g., Morlet *et al.*, 1996). As far as TEOAEs are concerned, the cochlear maturity has been called into question at least in preterm neonates, e.g., by Morlet *et al.* (1996) and Brienesse *et al.* (1996, 1997). They reported maturation-like TEOAE or SOAE changes after birth that could not be attributed to the growth of the outer and middle ear transmission apparatus. The present data confirm that TEOAE levels are age dependent in normal ears, and that the frequency interval with the largest TEOAEs is also age dependent. These findings are consistent with the idea that infant outer and middle ears favor TEOAEs at higher frequencies than adult ones. However, the ''filter-width'' similarity does not point to any maturation regarding TEOAE fine structure.

In the NIHL group, the differences with normal ears could not involve any difference in the conductive system, and they were somewhat unexpected because all pure-tone auditory thresholds had remained unaffected at 1 kHz. In the first place, the average level of TEOAEs around 1 kHz was significantly smaller than in normal ears, by about 5 dB, despite strictly normal auditory thresholds at 1 kHz. This finding confirms previous reports by Avan *et al.* in humans (1997) and guinea pigs (1995), that the decrease of TEOAEs tends to affect lower frequencies than those suggested by a simple frequency-to-frequency correspondence when inspecting pure-tone audiograms and TEOAE spectra. Recently, Yates and Withnell (1999) attributed this effect to a large contribution to lower frequency emissions from intermodulation processes between high-frequency components of the transient stimulus. Furthermore, in the group of ears with NIHL, the hypothetical characteristics of the cochlear TEOAE-generating filters seemed altered. However, the process of averaging may simply have blurred preferred spacings if they happened to differ markedly from one ear to another. Pick *et al.* (1977), Glasberg and Moore (1986), and Tyler (1986) have indeed described that auditory filters were broadened in damaged cochleas, but they did not specifically report filter broadening at low frequencies in the presence of damage restricted to higher frequencies. Instead, filter bandwidths were altered only when the damage exceeded 30–40 dB around the tested frequencies.

The need to exclude the ears with SOAEs from model-derived computations may have biased the recruitment of ears. There is evidence that ears with SOAEs have a better sensitivity and perhaps, an overall better hearing than ears without SOAEs (McFadden and Mishra, 1993). Most of the ears that were excluded in this study were from neonates. Detailed clinical studies have reported that the 70% prevalence of SOAEs do not vary much with age if hearing loss is controlled as a variable (Burns *et al.*, 1992; Talmadge *et al.*, 1993). The prevalence in the present sample of normal adult ears was only 19%; however, recall that long-lasting TEOAEs were assumed to be synchronized SOAEs, while SOAEs were not directly detected. This technical difference may explain the discrepancy in prevalence rates (Wable and Collet, 1994). Nonetheless, when the model-derived computation was applied to NNS+ ears despite its questionable validity, the influence of one or few SOAEs embedded in the overall TEOAE responses seemed negligible as illustrated by Figs. 6 and 7.

In conclusion, the present study confirms that the spectrum of TEOAEs is made of more or less regularly spaced peaks and troughs, as described, for instance, by Zwicker and Schloth (1984). This observation held both for healthy adults and neonates, whereas less regularity could be found in a sample of subjects with mild noise-induced hearing loss, although their hearing loss was restricted to frequencies >2 kHz. The origin of TEOAE spectral regularity and its meaning are not yet fully understood, although several models have predicted its existence and related it to the fine cochlear filtering (Wit *et al.*, 1994b; Shera and Zweig, 1995). The application of Wit's model to averaging the TEOAE patterns from as homogeneous as possible samples of ears seemed to provide consistent results. The average width of the TEOAE-generating cochlear filters turned out to be similar in groups of neonates and normally hearing young adults. We speculate that TEOAEs provide an objective reflection of the tuning of their sources, and that they may be used for a noninvasive comparison of the average performance of tuning across groups of ears. Several questions remain open, e.g., how do these results relate to the ear's performance, how to

account for the case of ears with SOAEs, is there any early signature of cochlear impairment in the individual TEOAE spectra from ears with near-normal auditory thresholds, and, if this were confirmed, could the proposed method be useful to address the as yet unsolved case of adults with poor TEOAEs despite normal pure-tone auditory (Kapadia and Lutman, 1997).

ACKNOWLEDGMENTS

The comments of Jacek Smurzynski and Rudolf Probst on an earlier version of this text and the editorial help of Fran Harris are gratefully acknowledged.

Aidan, D., Avan, P., and Bonfils, P. (**1999**). ''Auditory screening in neonates by means of transient evoked otoacoustic emissions: A report of 2,842 recordings,'' Ann. Otol. Rhinol. Laryngol. **108**, 525–531.

Avan, P., Bonfils, P., Loth, D., Elbez, M., and Erminy, M. (**1995**). ''Transient-evoked otoacoustic emissions and high-frequency acoustic trauma in the guinea pig,'' J. Acoust. Soc. Am. **97**, 3012–3020.

Avan, P., Elbez, M., and Bonfils, P. (**1997**). ''Click-evoked otoacoustic emissions and the influence of high-frequency hearing losses in humans,'' J. Acoust. Soc. Am. **101**, 2771–2777.

Bogert, B. P., Healy, M. J. R., and Tukey, J. W. (**1963**). ''The quefrency analysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking,'' in *Proceedings of the Symposium on Time Series Analysis*, edited by M. Rosenblatt (Wiley, New York), pp. 209–243.

Brienesse, P., Anteunis, L. J., Wit, H. P., Gavilanes, D., and Maetzdorf, W. J. (**1996**). ''Otoacoustic emissions in preterm infants: indications for cochlear development?'' Audiology **35**, 296–306.

Brienesse, P., Anteunis, L. J., Maetzdorf, W. J., Blanco, C. E., and Manni, J. J. (**1997**). ''Frequency shift of individual spontaneous otoacoustic emissions in preterm infants,'' Pediatr. Res. **42**, 478–483.

Brownell, W. E. (**1990**). ''Outer hair cell electromotility and otoacoustic emissions,'' Ear Hear. **11**, 82–92.

Brownell, W. E., Bader, C. R., Bertrand, D., de Ribaupierre, Y. (**1985**). ''Evoked mechanical responses of isolated cochlear outer hair cells,'' Science **227,** 194–196.

Burns, E. M., Arehart, K. H., and Campbell, S. L. (**1992**). ''Prevalence of spontaneous otoacoustic emissions in neonates,'' J. Acoust. Soc. Am. **91**, 1571–1575.

Dallos, P., and Evans, B. N. (**1995**). ''High frequency motility of outer hair cells and the cochlear amplifier,'' Science **267**, 2006–2009.

Davis H. (**1983**). ''An active process in cochlear mechanics,'' Hearing Res. **9**, 79–90.

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.

Gold, T. (**1948**). ''Hearing II. The physical basis of the action of the cochlea,'' Proc. R. Soc. London B **135**, 492–498.

Gold, T. (**1989**). ''Historical background to the proposal, 40 years ago, of an active model for cochlear frequency analysis,'' in *Cochlear Mechanisms, Structures, Functions and Models*, edited by J. P. Wilson and D. T. Kemp (Plenum, New York), pp. 299–305.

Helmholtz, H. von (**1862**). *Die Lehre von den Tonempfindungen als Physiologische Grundlage für die Theorie der Musik* (F. Wieveg, Braunschweig).

Hilger, A. W., Furness, D. N., and Wilson, J. P. (**1995**). ''The possible relationship between transient evoked otoacoustic emissions and organ of Corti irregularities in the guinea pig,'' Hear. Res. **84**, 1–11.

Kapadia, S., and Lutman, M. (**1997**). ''Are normal hearing thresholds a sufficient condition for click-evoked otoacoustic emissions?'' J. Acoust. Soc. Am. **101**, 3566–3567.

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.

Kemp, D. T. (**1986**). ''Otoacoustic emissions, travelling waves and cochlear mechanisms,'' Hear. Res. **22**, 95–104.

Kemp, D. T., Ryan, S., and Bray, P. (**1990**). ''A guide to the effective use of otoacoustic emissions,'' Ear Hear. **11**, 93–105.

Liberman, M. C., and Dodds, L. W. (**1984**). ''Single neuron labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves,'' Hear. Res. **16**, 55–74.

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.

Lonsbury-Martin, B. L., Harris, F. P., Stagner, B. B., Hawkins, M. D., and Martin, G. K. (**1990**). ''Distortion product emissions in humans: II. Relations to acoustic immittance and stimulus frequency and spontaneous otoacoustic emissions in normally hearing subjects,'' Ann. Otol. Rhinol. Laryngol. **99**, 15–29.

Manley, G. A. (**1983**). ''Frequency spacing of acoustic emissions: a possible explanation,'' in *Mechanisms of Hearing*, edited by W. R. Webster and L. M. Aitken (Monash U.P., Clayton), pp. 36–39.

McFadden, D., and Mishra, R. (**1993**). ''On the relation between hearing sensitivity and otoacoustic emissions,'' Hear. Res. **71**, 208–213.

Morlet, T., Perrin, E., Durrant, J. D., Lapillonne, A., Ferber, C., Duclaux, R., Putet, G., and Collet, L. (**1996**). ''Development of active cochlear mechanisms in humans differs between gender,'' Neurosci. Lett. **220**, 49–52.

Patuzzi, R. B. (**1991**). ''Effect of noise on auditory nerve responses,'' in *Noise-induced Hearing Loss*, edited by A. L. Dancer, D. Henderson, R. J. Salvi, and R. P. Hemernik (Mosby Year Book, St. Louis), pp. 45–59.

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.

Pick, G., Evans, E. F., and Wilson, J. P. (**1977**). ''Frequency resolution in patients with hearing loss of cochlear origin,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 273–281.

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., Coats, A. C., Martin, G. K., and Lonsbury-Martin, B. L. (**1986**). ''Spontaneous, click- and toneburt-evoked otoacoustic emissions from normal ears,'' Hear. Res. **21**, 261–275.

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 Freyman, R. (**1983**). ''Spontaneous and impulsively evoked otoacoustic emissions: Indicators of cochlear pathology?'' Hear. Res. **19**, 283–300.

Shear, C., and Zweig, G. (**1995**). ''The origin of periodicity in the spectrum of otoacoustic emissions,'' J. Acoust. Soc. Am. **98**, 2018–2047.

Sutton, G. J., and Wilson, J. P. (**1981**). ''Modelling cochlear echoes: The influence of irregularities in frequency mapping on summed cochlear activity,'' in *Mechanisms of Hearing*, edited by E. de Boer and M. A. Viergever (Delft U.P., Delft), pp. 83–90.

Talmadge, C. L., Long, G. R., Murphy, W. J., and Tubis, A. (**1993**). ''New off-line method for detecting spontaneous otoacoustic emissions in human subjects,'' Hear. Res. **71**, 170–182.

Tyler, R. S. (**1986**). ''Frequency resolution in hearing-impaired listeners,'' in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London), pp. 309–371

Vohr, B. R., Carty, L. M., Moore, P. E., and Letourneau, K. (**1998**). ''The Rhode Island Hearing Assessment Program: Experience with statewide hearing screening (1993–1996),'' J. Pediatr. (St. Louis) **133**, 353–357.

Wable, J., and Collet, L. (**1994**). ''Can synchronized otoacoustic emissions really be attributed to SOAEs?'' Hear. Res. **80**, 141–145.

Weiss, T. F., Peake, W. T., and Rosowski, J. J. (**1985**). ''A model for signal transmission in an ear having hair cells with free-standing stereocilia. I. Empirical basis for model structure,'' Hear. Res. **20**, 131–138.

Wilson, J. P. (**1980**). ''Model of cochlear function and acoustic re-emission,'' in *Psychophysical, Physiological and Behavioural Studies in Hearing*, edited by G. van den Brink and F. A. Bilsen (Delft U. P., Delft), pp. 72–73.

Wit, H. P., van Dijk, P., and Avan, P. (**1994a**). ''Wavelet analysis of real ear and synthesized click evoked otoacoustic emissions,'' Hear. Res. **73**, 141–147.

Wit, H. P., van Dijk, P., and Avan, P. (**1994b**). ''On the shape of evoked otoacoustic emission spectra,'' Hear. Res. **81**, 208–214.

Yates, G. K., and Withnell, R. H. (**1999**). ''The role of intermodulation distortion in transient-evoked otoacoustic emissions,'' Hear. Res. **136**, 49–64.
Zwicker, E. (**1986**). '' 'Otoacoustic emissions' in a nonlinear cochlear hardware model with feedback,'' J. Acoust. Soc. Am. **80**, 154–162.
Zwicker, E. (**1990**). ''On the frequency separation of simultaneously evoked otoacoustic emissions' consecutive extrema and its relation to cochlear traveling waves,'' J. Acoust. Soc. Am. **88**, 1639–1641.
Zwicker, E., and Schloth, E. (**1984**). ''Interrelation of different oto-acoustic emissions,'' J. Acoust. Soc. Am. **75**, 1148–1154.

A noninvasive method for estimating acoustic admittance at the tympanic membrane

Gregory T. Huang, John J. Rosowski, Sunil Puria,[a] and William T. Peake
Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114 and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 5 August 1999; revised 27 April 2000; accepted 1 May 2000)

The acoustic admittance at the tympanic membrane (TM), Y_{TM}, describes the linear acoustic properties of the ear. Here, a noninvasive measurement procedure is developed for estimating Y_{TM} in intact ears. The method consists of (1) measuring the admittance in the ear canal Y_{EC} with a commercially available earphone-and-microphone system, and (2) estimating Y_{TM} via a uniform-tube approximation of the space between the measurement point and the TM. The dimensions of this space are estimated from Y_{EC} via an area-estimation algorithm [Keefe *et al.*, J. Acoust. Soc. Am. **91**, 470 (1992)] and measurements made with controlled static pressures in the canal. Measurements in artificial loads are used to test the accuracy of the measurement system and to determine sources of error. For accurate admittance measurements: (1) extension of the microphone tube medially beyond the earphone's port is necessary for frequencies above 2 kHz; (2) the acoustic system must be calibrated in known loads with diameters within 15% of the canal diameter, because the source's output characteristics vary with load diameter. The method is applied to intact ears of anesthetized domestic cats; for frequencies below 5 kHz, the estimated Y_{TM} in four ears have features that are similar to those of previous measurements made at the cat TM. Sources of error include nonuniform waves generated at the earphone's narrow port, inaccuracy in estimation of canal dimensions, irregular geometry of the canal, and earphone-microphone cross talk. [S0001-4966(00)00409-4]

PACS numbers: 43.64.Yp, 43.64.Ha, 43.64.Tk, 43.58.Bh [BLM]

INTRODUCTION

A. Motivation

The acoustic admittance at the tympanic membrane (TM) Y_{TM}—the ratio of the volume velocity of the TM to the sound pressure just lateral to the TM—is a useful description of an ear's acoustic properties for scientific and clinical purposes.[1] Also known as the middle-ear input admittance, Y_{TM} characterizes the acoustic response of the middle ear at the TM (Onchi, 1961; Zwislocki, 1962; Mundie, 1963; Møller, 1965); it affects the sound-pressure distribution along the ear canal and the flow of acoustic energy into the ear (Rosowski *et al.*, 1986; Stinson and Khanna, 1994). Clinical measurements of Y_{TM} aid in the diagnosis of pathologies such as fluid in the middle ear and ossicular fixation (Jerger, 1975; Wilber and Feldman, 1976; Margolis and Shanks, 1985).

Despite improvements in measurement methods over the last 40 years, accounting for the effects of the ear canal space between the measurement point and the TM remains a troublesome issue. For an intact human ear, the acoustic probe is usually placed such that its tip is 8–15 mm lateral to the TM, for the comfort and safety of the subject (e.g., Keefe *et al.*, 1993; Voss and Allen, 1994). In clinical admittance measurements, which are usually restricted to frequencies below 1 kHz, the intervening air volume of the canal is determined from the acoustic response with applied static pressure in the canal that stiffens the TM; the compliance of the canal volume can be subtracted from the zero-static-pressure admittance to give the admittance at the TM (Margolis and Shanks, 1985). For higher frequencies, where acoustic behavior is complicated by spatial variations, the simplest approach is to approximate the canal space by a rigid cylindrical tube of uniform cross section in which sound pressure varies only along the canal (i.e., a uniform plane-wave model); the measured admittance is then transformed via a transmission-line model to give an estimate of Y_{TM}. With this method, Møller (1960) and Zwislocki (1970) were able to estimate Y_{TM} in humans for frequencies up to 2 kHz. Rabinowitz (1981) estimated Y_{TM} for frequencies up to 4 kHz with a more elaborate scheme.[2]

Acoustic behavior in the ear canal has also been described in terms of acoustic pressure reflectance, the ratio of the pressure component associated with the outward-going wave to the component associated with the inward-going wave (Stinson *et al.*, 1982; Keefe *et al.*, 1992; Voss and Allen, 1994; Stinson and Khanna, 1994; Puria and Allen, 1998; Feeney and Keefe, 1999). With a uniform plane-wave model, reflectance is determined by the admittance in the canal and the cross-sectional area of the canal, so that admittance and reflectance are interchangeable descriptions of acoustic behavior; the measurement difficulties described in this paper also apply to reflectance. With the assumption of lossless sound propagation in the canal, reflectance magnitude is constant with position along the canal, and spatial

[a]Current address: Stanford University, 801 Welch Road, Palo Alto, CA 94304.

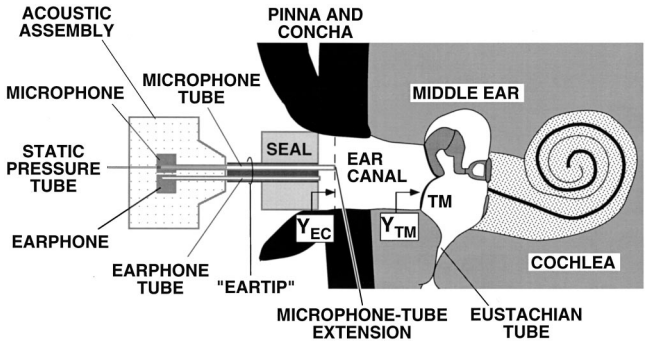


FIG. 1. Schematic diagram of the acoustic probe coupled to an external ear canal. The black area represents the pinna, concha, and cartilaginous ear canal. The unshaded (white) areas represent air filled spaces—ear canal, middle-ear cavity, Eustachian tube, and the microphone and earphone tubes of the acoustic probe. Y_{EC} is the acoustic admittance measured at the microphone-tube tip's location (dashed line) in the ear canal. Y_{TM} is the acoustic admittance at the tympanic membrane (TM), i.e., the middle-ear input admittance. The acoustic assembly (Etymotic Research ER-10C) is coupled to the ear canal through a flexible plastic ''eartip'' that houses both the microphone and earphone tubes. A compressible seal surrounds the eartip in the canal. The outer diameter of the eartip is 3.3 mm. Inner diameters: microphone tube 1.9 mm, earphone tube 0.5 mm. The static-pressure tube and the extension of the microphone tube (3–4 mm beyond the earphone-tube tip) are custom alterations of the ER-10C.

variations are described by a linear dependence of reflectance angle on position. Although the reflectance magnitude describes some important features of power flow, it does not uniquely characterize the mechanics of the middle ear. For example, knowledge of Y_{TM} allows one to distinguish between masslike and compliancelike middle-ear responses, whereas reflectance magnitude does not. We choose to focus on admittance because interpretation of its properties in terms of middle-ear structural contributions has been fruitful.

The goal of this paper is to develop a noninvasive method for measuring ear canal admittance and for estimating Y_{TM}, with a commercially available acoustic system that is compatible with a range of external-ear geometries. Within this context: (1) we test the measurement method in known acoustic loads to determine the size of errors and their frequency dependence; (2) we demonstrate the measurement procedure in the intact ears of anesthetized cats; and (3) we report estimates of Y_{TM} in these ears.

Applications of the methodology of this paper include: (1) broadband clinical measurements (e.g., tympanometry) that account for the ear canal (Stevens *et al.*, 1987) or employ static pressures (Wada *et al.*, 1989; Keefe and Levi, 1996; Margolis *et al.*, 1999); (2) design of earphones and stimuli for virtual auditory-space systems (Blauert, 1983, pp. 358–367; Chan *et al.*, 1993); (3) studies of middle- and external-ear function that require noninvasive measurements, e.g., for interspecies (Huang *et al.*, 2000a) or developmental (Keefe *et al.*, 1993) comparisons.

B. Approach

The basic method involves connecting the acoustic system, consisting of an earphone (sound source) and a microphone, to an ear canal (Fig. 1). The earphone produces a broadband ''chirp'' stimulus, and the microphone measures the sound pressure in the canal. We treat the ear as a passive, linear system that can be characterized by an acoustic input admittance. This approach restricts the range of stimulus levels to roughly 60–90 dB SPL, which is high enough such that otoacoustic emissions are negligible [less than 20 dB SPL in cats (Guinan, 1990)] and low enough such that nonlinearities (e.g., middle-ear muscle contractions) can be ignored (Guinan and Peake, 1967; Rabinowitz, 1981). With the source's output characteristics determined from measurements made in known calibration loads, the admittance in the canal at the probe's tip, Y_{EC}, can be determined as a function of frequency from the measured sound pressure. The desired quantity Y_{TM}, the admittance at the TM, can differ greatly from Y_{EC} (e.g., Rabinowitz, 1981, Figs. 5 and 7; Huang *et al.*, 1997, Fig. 4).

The coupling of the acoustic system to the ear canal affects the accuracy of Y_{EC} measurements in two ways. (1) The position of the microphone-tube tip relative to the earphone-tube tip is important, because nonuniform waves are generated at the earphone-tube tip. Extension of the microphone tube a short distance beyond the earphone-tube tip (as in Fig. 1) is thought to reduce the contribution of nonuniform components in the measured pressure (Rabinowitz, 1981; Keefe *et al.*, 1992; Brass and Locke, 1997). We demonstrate the effects of the microphone-tube extension on admittance measurements. (2) The cross-sectional area of the canal at the measurement location can vary substantially among ears. We show that variations in canal area affect the measured admittance; a key mechanism is that properties of the source are affected by the configuration of the region where the acoustic probe couples to the canal (Fig. 1).

The effect of the ear canal's geometry on our ability to estimate Y_{TM} is a second key issue. The shape of the canal can be complex, with changes in both cross-sectional area and axial direction (Shaw, 1974; Stinson, 1985; Rosowski *et al.*, 1988; Rosowski, 1994, pp. 183–185). The human ear canal has been approximated as a uniform tube (Stinson *et al.*, 1982) and as a horn (Stinson, 1985; Rabbitt, 1988; Stinson, 1990), for measuring and modeling the reflectance and pressure distribution along the canal. To estimate the acoustic compliance and resistance of the middle ear from measurements in the human canal, Keefe *et al.* (1993, p. 2628) subtracted a conical approximation of the canal's volume. In domestic cat, Rosowski *et al.* (1988) used a uniform-tube model of the canal to represent its contribution to external-ear radiation-impedance measurements. In noninvasive procedures, we are not able to measure dimensions of the ear canal. Therefore, in this work we test a uniform-tube (transmission-line) model, which requires the estimation of only the ''effective'' length and cross-sectional area of the canal. With these estimates we calculate the ''admittance-location transformation.''

C. Overview

The results of this paper are presented in two sections. In Sec. I, the admittance-measurement method is developed and tested in known acoustic loads, with special attention to two issues: (1) the position of the microphone-tube tip; and (2) differences in canal diameter among subjects. Also, the admittance-location transformation is tested in known loads.

In Sec. II, the method is applied to intact ears of anesthetized cats. The estimates of Y_{TM} are compared to previous measurements of admittance at the TM in cats. Lastly, the repeatability of the method and the sensitivity of the location transformation to changes in the estimated canal dimensions are determined.

I. TESTS OF METHOD IN KNOWN ACOUSTIC LOADS

A. Admittance-measurement technique

1. Hardware

We chose a commercially available acoustic-probe system—the Etymotic Research (Elk Grove Village, IL) ER-10C—which has four main advantages: (1) The earphone and microphone are encased in a lightweight acoustic assembly (Fig. 1); this arrangement reduces the likelihood that the source's characteristics may change as a result of movement of its components or entry of extraneous material. (2) The coupling from the acoustic assembly to the ear uses flexible components, which are crucial for measurements in ear canals with bends. (3) The driver pre-amp and microphone amplifier are housed in a small battery-powered box, so the system is portable. (4) Etymotic Research has (at our request) added a tube in the acoustic assembly through which a static-pressure source can be coupled to the ear canal (Fig. 1); this feature allows acoustic measurements with controlled static-pressure variations in the canal.

The acoustic-probe system is used to stimulate the ear and to measure the acoustic admittance in the ear canal.[3] The basic arrangement is that the earphone and microphone are connected to tubes that are sealed into the canal. In Fig. 1, the acoustic assembly is coupled to an Etymotic ''eartip'' that contains both the earphone and microphone tubes. The system is acoustically sealed into the canal with a compressible foam or rubber plug that surrounds the eartip. The plugs are available in different sizes to allow insertion into a wide range of human ear canals, from infant to adult. An Etymotic eartip with a foam-plug seal was used for all measurements in Sec. I; however, this material does not provide a static-pressure seal. Other types of eartips and seals, designed for use in cats, are described in Sec. II A.

2. Source calibration

The acoustic source is characterized by a Norton equivalent circuit consisting of a volume velocity source U_S in parallel with a source admittance Y_S (Flanders, 1932; Møller, 1960; Rabinowitz, 1981; Allen, 1986; Rosowski *et al.*, 1986; Keefe *et al.*, 1992; Ravicz *et al.*, 1992; Lynch *et al.*, 1994; Neely and Gorga, 1998). We determine these source parameters from measurements in two types of cylindrical calibration loads of known admittance: (1) 15-m-long tubes (effectively ''infinite'' transmission lines) made of flexible-plastic Tygon; (2) short, closed cavities 8–12 mm in length made of Plexiglas. The length of a closed-cavity load, measured with calipers to within 0.5-mm accuracy, is defined as the distance between the microphone-tube tip and the load's termination. The calibration-load theoretical admittances are calculated from a lossy transmission-line model (Egolf, 1977; White *et al.*, 1980; Keefe, 1984; Allen, 1986; Zuercher *et al.*, 1988). For a given calibration-load diameter, Y_S and U_S are determined uniquely from the sound pressures measured in the two loads (e.g., Lynch *et al.*, 1994, p. 2186); for frequencies below 7 kHz, the source parameters are not very sensitive to small changes in the theoretical lengths of these calibration loads.[4]

For an unknown acoustic load, the sound pressure P_L produced by the source in the load is measured and the admittance Y_L calculated as

$$Y_L = (U_S/P_L) - Y_S. \tag{1}$$

The measurement system can be tested on other acoustic loads of known admittance. In the next section, we test the accuracy of the admittance-measurement system within the context of two issues that complicate the method.

B. Complications

1. Position of the microphone-tube tip

a. Background. In the standard Etymotic ER-10C system, both the earphone- and microphone-tube openings are flush with the eartip termination and the surface of a foam or rubber seal. Trial measurements with this configuration in test loads had significant errors in admittance above 1 kHz, particularly for loads having large diameters (10–12 mm). We found that accuracy was greatly improved by extending the microphone tube beyond the eartip termination (and earphone-tube tip) via a short section of steel tubing. The desirability of a microphone-tube extension has been reported for other acoustic-admittance probes (Rabinowitz, 1981; Keefe *et al.*, 1992).

The admittance-measurement scheme assumes uniform acoustic plane waves at the measurement point. However, the discontinuity in diameter from the earphone tube (0.5 mm) to the load (5–12 mm) introduces radial spreading of the wave, which requires nonuniform sound-pressure waves near the source-load interface (Miles, 1946; Ingard, 1948; Karal, 1953; Morse and Ingard, 1968). In theory, nonuniform waves decay exponentially with distance along the canal for frequencies of interest ($<$10 kHz) (Keefe and Benade, 1981; Rabbitt and Friedrich, 1991; Brass and Locke, 1997), so extension of the microphone tube should allow measurement of the sound pressure in a region of more nearly uniform plane waves.

b. Choice of microphone-tube extension length. An appropriate length for the microphone-tube extension can be estimated by considering how a sound-pressure wave's components vary along the long axis (call it z) of a rigid cylindrical tube of radius a. In the sinusoidal steady state, solutions of the wave equation in cylindrical coordinates (r, ϕ, z) subject to the rigid-wall boundary condition have the form (adapted from Morse and Ingard, 1968)

$$P_{mn} = A\{\cos(m\phi) \text{ or } \sin(m\phi)\} J_m(\pi q_{mn} r/a) \exp(-jk_{mn} z), \tag{2}$$

where J_m is a Bessel function, πq_{mn} are the zeroes of J'_m (the prime denotes a derivative with respect to the argument), A is a constant determined by the source, m and n are integers, and $j = \sqrt{-1}$. Equation (2) describes one pressure component or ''mode'' whose spatial dependence is characterized

by the two-component mode number (m, n). The total pressure is the summation of these modes over (m, n), where mode (0, 0) is the uniform component that is dominant at locations away from the earphone-tube tip, and is what we wish to measure. The dependence of each mode on axial distance z is controlled by the propagation wave number

$$k_{mn}=[(2\pi f/c)^2-(\pi q_{mn}/a)^2]^{1/2}, \tag{3}$$

where f is frequency and c is the speed of sound in air. The (m, n) mode is ''evanescent'' or ''cut off'' (i.e., it decays exponentially with z) for frequencies below $f_{mn}=q_{mn}c/2a$, where k_{mn} is imaginary.

In choosing an extension-tube length, an arbitrary but reasonable criterion is that the nonuniform mode with the lowest cutoff frequency [the (1, 0) mode] must decay by a factor of e in this distance, at the highest frequency of interest $f_{\max}$. This length or ''space constant'' (in z), l_e, is determined by setting the argument of the exponential in Eq. (2) equal to -1, with $m=1$ and $n=0$, such that $-jk_{10}l_e=-1$. Substitution into Eq. (3) gives

$$l_e=1/(jk_{10})=[(\pi q_{10}/a)^2-(2\pi f_{\max}/c)^2]^{-1/2}, \tag{4}$$

where $q_{10}=0.586$. We choose (1) $f_{\max}=10$ kHz, because above this frequency admittance-measurement accuracy is limited by other factors (see Sec. II A 3), and (2) load diameter $2a=12.5$ mm, the diameter of our largest test loads, which corresponds to a large ear canal. These values yield $l_e=4.4$ mm. For an ear canal or test load with a smaller diameter, or for lower frequencies, the decay of evanescent modes would be *more* rapid with z, and l_e would be smaller.[5]

Although the above analysis assumes a circular ear-canal cross section, a more detailed theoretical model indicates that the results apply to more realistic (i.e., elliptical) canal geometries. For a given cross section, the cutoff frequencies of nonuniform modes are largely determined by the greatest transverse dimension (Rabbitt and Friedrich, 1991). Thus a microphone-tube extension of 4.4 mm should satisfy approximately the same criterion for an elliptical canal cross section with a major axis length of 12.5 mm.

For our measurement system, we created the extension either by (1) insertion of a section of steel tubing (inner diameter 1.9 mm) that fit snugly into the standard eartip's microphone tube, or (2) construction of special eartips consisting of separate earphone and microphone tubes made of Tygon. The second configuration was used in cat ears for reasons given in Sec. II A 3.

c. Effects of extension on measured admittance. Representative comparisons of admittance measurements to theory and the resulting error ratios are shown in Fig. 2 for two test loads. Admittances were measured with and without a 4-mm microphone-tube extension.[6] First, we focus on the left panels [Fig. 2(A)]. The test load is a rigidly terminated tube with inner diameter 9.5 mm and length 30 mm (measured from the microphone-tube tip to the termination). This load provides a test of the system's ability to measure (1) the compliance of an earlike equivalent volume of air (roughly 2 cm^3) and (2) sharp midfrequency features such as the middle-ear cavity antiresonance present in cats near 4 kHz (Møller, 1965; Guinan and Peake, 1967; Lynch, 1981; Peake *et al.*, 1992; Lynch *et al.*, 1994; Huang *et al.*, 1997). The theoretical admittance of the load was calculated with a lossy transmission-line model.

For frequencies below 1 kHz, the microphone-tube extension has little effect, as both source configurations give accurate estimates of the magnitude and angle of the load admittance [Fig. 2(A)]. Above 1 kHz, in contrast, the accuracy in admittance is generally greater with the extension than without the extension [Fig. 2(A), lower]. Without the extension, the errors are quite large above 2 kHz; the measurement underestimates the frequency of the admittance-magnitude maximum (quarter-wave resonance) near 3 kHz by 20%, which is consistent with the addition of a series mass associated with nonuniform wave components (Ingard, 1948; Karal, 1953; Puria, 1991). The no-extension measurement also fails grossly to approximate the shape of the admittance-magnitude minimum (half-wave antiresonance) near 6 kHz. We attribute these errors to nonuniform waves at the microphone-tube tip that lead to errors both in the source parameters (Y_S and U_S) and in the measured pressure (Keefe and Benade, 1981); near 6 kHz, errors in the source admittance Y_S lead to particularly large errors in the measured admittance Y_L, because the load admittance is much smaller in magnitude than Y_S (Lynch *et al.*, 1994).

With the microphone-tube extension, the measured and theoretical curves agree to within 15% (1.2 dB) in magnitude and 0.04 periods in angle up to 8 kHz, except near the frequencies of the extrema (3 and 6 kHz), where small errors in the test-load length can lead to large differences between measurement and theory. For frequencies above 8 kHz, the accuracy may be limited by (1) residual nonuniform-wave contributions to the pressures measured in the test load and in the calibration loads (see Sec. I B 1 d), and (2) cross talk between the earphone and microphone (see Sec. II A 3 b). The signal-to-noise ratio is not limiting for this frequency range; in the 6–10 kHz region, the noise floor of our measurements is generally 20–30 dB below the signal.

For the right panels [Fig. 2(B)], the test load has the same inner diameter (9.5 mm) and is designed to respond more like a human ear for higher frequencies. The load consists of a section of tube (length 12 mm) terminated by an acoustic resistor leading to a section of closed tube (length 10 mm). The resistor is made from four Knowles BF-series acoustic dampers (each nominally 150 mks MΩ) inserted in parallel into a thin teflon disk that fits snugly in the tube. The total resistance is nominally 37.5 mks MΩ, and the volume of the closed-tube section, or backing cavity, is 0.7 cm^3. The resistance and cavity compliance values were chosen to approximate the resistance and compliance of the human middle ear (Zwislocki, 1970; Rabinowitz, 1981), while the tube section into which the probe is inserted approximates the ear canal.

For the theoretical admittance, the tube sections were modeled as lossy transmission lines, and a series resistance-mass lumped model was used for each of the four parallel Knowles dampers. The nominal resistance value (150 MΩ) was used; the mass value was obtained from the dimensions of the small tube that holds each damper (length 2.4 mm, inner diameter 2.0 mm). Accounting for the end corrections,

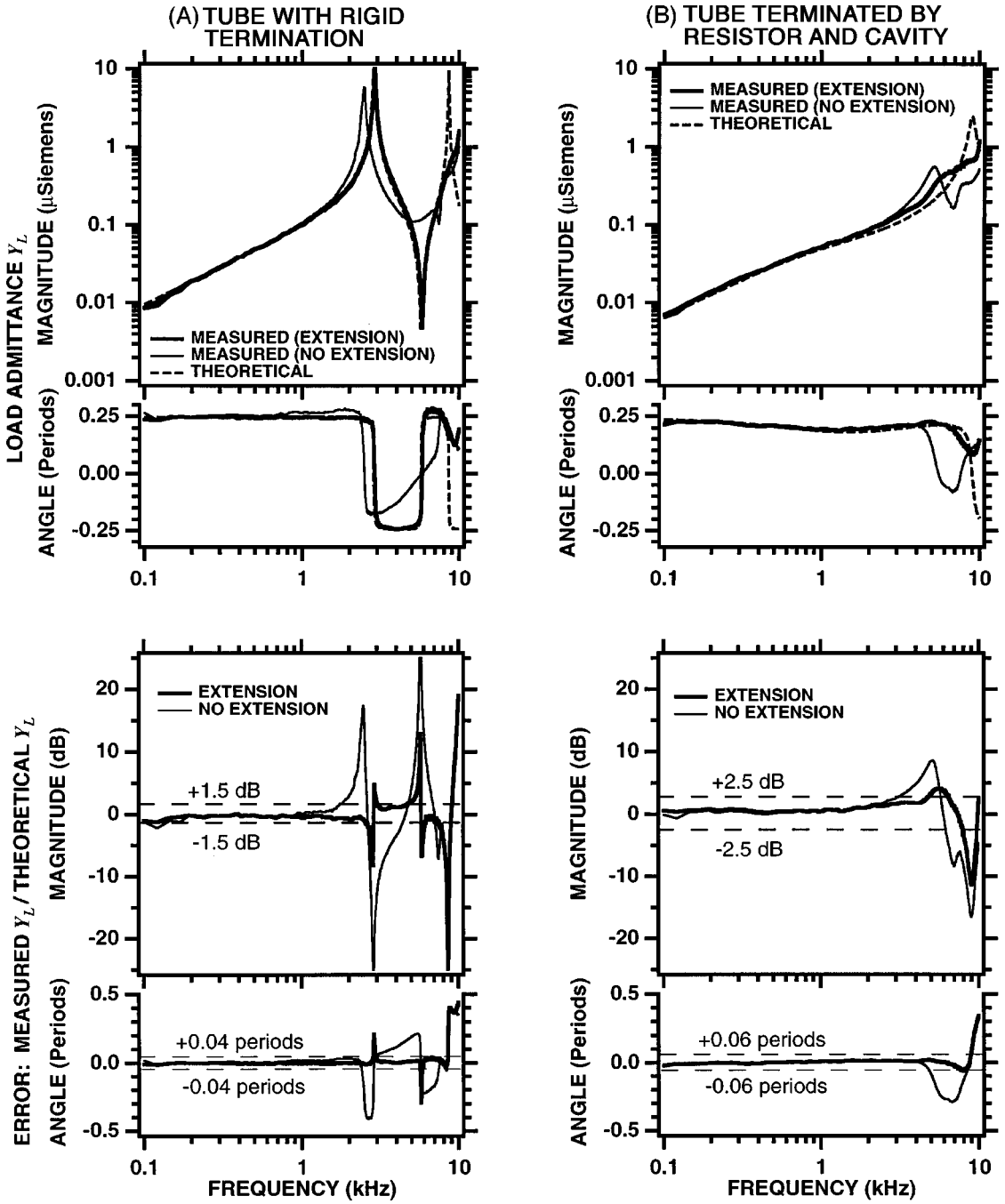


FIG. 2. Effects of the microphone-tube extension on admittance measurements. Results are shown for two test loads: (A) a cylindrical Plexiglas tube having inner diameter 9.5 mm and length 30 mm with a rigid termination; (B) a similar cylindrical tube having inner diameter 9.5 mm and length 12 mm with a resistive-compliant termination—an acoustic resistor (nominally 37.5 mks MΩ) that connects to a closed cavity having the same inner diameter and length 10 mm. Admittances (UPPER panels) were measured with and without the microphone-tube extension; each of these configurations was calibrated using two reference loads. In the upper left panel, the dashed curves nearly overlay the thick curves below 8 kHz. Theoretical admittances were computed from a lossy transmission-line model for the tubes and a lumped resistance-mass model for the acoustic resistor. Errors (LOWER panels) are the ratios of measured admittances to theoretical admittances. The frequency scale is the same in the upper and lower plots. Admittance magnitude units: 1μSiemen $=1/(\text{mks M}\Omega)=10^{-6}$ m^3/(Pa s). Phase angles are plotted in periods (1 period$=2\pi$ radians).

the mass value used was 1606 kg/m^4 for each damper (Beranek, 1986, pp. 137–138). In this resistance-mass model, the discontinuity in area between the closed 9.5-mm-diameter tube and each small damper is accounted for in the end correction.[7] The theoretical admittance is similar to admittances measured in the lateral portion of the ear canal in humans (Keefe *et al.*, 1993; Voss and Allen, 1994)—roughly compliant for a large low-frequency range, with a sharp negative phase transition above 7 kHz [Fig. 2(B), upper].

Above 3 kHz, admittance measurements made in this load without a microphone-tube extension have larger errors than those made with an extension, particularly in phase [Fig. 2(B), lower]. Measurements made with the extension generally agree with the theoretical curves to within 25% (roughly 2 dB) in magnitude and 0.06 periods in angle up to 8 kHz (except for the magnitude between 5 and 6 kHz). In

contrast, measurements made without the extension differ from theory by as much as 7 dB in magnitude and 0.27 periods in angle at 7 kHz.

In summary, the results of Fig. 2 demonstrate that the microphone-tube extension is crucial for accurate admittance measurements above 2 kHz in earlike loads.

d. Effects of extension on inferred admittance at the termination. A goal of this paper is to test a method for inferring the admittance at the tympanic membrane from measurements made more laterally in the ear canal. To correct for the difference in location, the method uses a transmission-line transformation (Sec. I C). Here, we determine the effect of the microphone-tube extension on this location transformation. We compare the transformation of a measurement made with the extension to the same transformation of a measurement made without the extension, in the human-earlike load of Fig. 2(B).

In Fig. 3 we plot the inferred admittance at the ''termination''—the resistor and backing cavity, 12 mm from the measurement point—based on the measurements in the upper panel of Fig. 2(B) made with and without an extension. The estimate with the extension is accurate to higher frequencies than the estimate without the extension. With the extension, the inferred admittance generally agrees with the theoretical within 25% (2.5 dB) in magnitude and 0.06 periods in angle for frequencies up to 6 kHz. With no extension, the inferred admittance has the same accuracy for frequencies up to only 3 kHz. The added frequency range is essential for detecting middle-ear acoustic features in the 3–5 kHz range.

The frequencies at which the measurements diverge from theory are more clearly defined in Fig. 3 than in Fig. 2(B). Also, the errors are generally larger in these inferred admittances. To understand the source of these errors and their implications for measurements in an ear canal of roughly the same size (length 12 mm, diameter 9.5 mm), we ask: are these errors the result of errors in the original (more distant) measurements, or are they introduced by the location transformation?

One hypothesis is that the main source of errors above 3 kHz (see solid curves in Fig. 3) is the presence of residual nonuniform waves associated with the narrow earphone tube, in the original measurements. To test this idea, we made similar measurements with a different acoustic system, referred to here as the ''large source''; this custom-built system is capable of measuring admittance up to 15 kHz (Huang *et al.*, 1997), but is not compatible with most intact external ears because of its large, rigid tip. The volume velocity U_S and source admittance Y_S of the large source are similar in magnitude to those of the ER-10C between and 3 and 10 kHz. Two main differences between the large source and the ER-10C are that (1) the earphone's output port is much larger, having a diameter of 4 mm versus 0.5 mm for the ER-10C, and (2) the microphone-tube tip is centered within the earphone port. The large source should generate smaller nonuniform modes, because the area discontinuity between the source and the load is smaller (e.g., Karal, 1953).

In Fig. 3, the dotted curve is the admittance measured using the large source positioned 8 mm from the termination

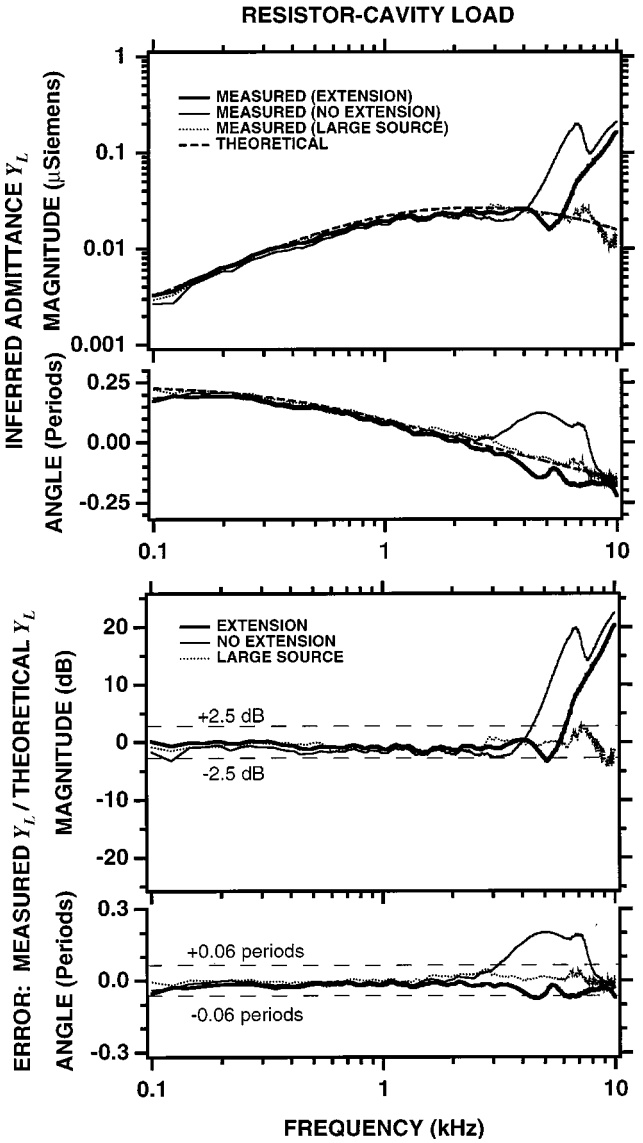


FIG. 3. Effects of variations in measurement method on the inferred admittance of the earlike termination of the test load of Fig. 2(B). The admittance measurements reported in Fig. 2(B) were transformed to the location of the acoustic resistor via a uniform-tube (length 12 mm) plane-wave model. The ''large source'' measurement was made 8 mm from the resistor, using an acoustic source with a large earphone port (e.g., Huang *et al.*, 1997), and transformed to the same location. The theoretical admittance (UPPER panel) and error ratios (LOWER panel) were computed as in Fig. 2. In both plots, the dotted curve is jagged above 6 kHz, as the distance between data points is compressed on the log frequency scale.

and transformed to the termination. The inferred admittance agrees well with the theoretical up to 10 kHz. The greater accuracy of this estimate for higher frequencies indicates that (1) the location transformation is not the main source of errors for this load, and (2) the configuration of the earphone and microphone ports in the ''large source'' greatly reduces the measurement error above 3 kHz. This result is consistent with the hypothesis that the narrow earphone tube of the ER-10C system limits the accuracy of admittance measurements above 3 kHz.

In summary, the data of this section indicate that the microphone-tube extension *reduces* the effects of nonuniform-wave modes generated at the tip of the narrow

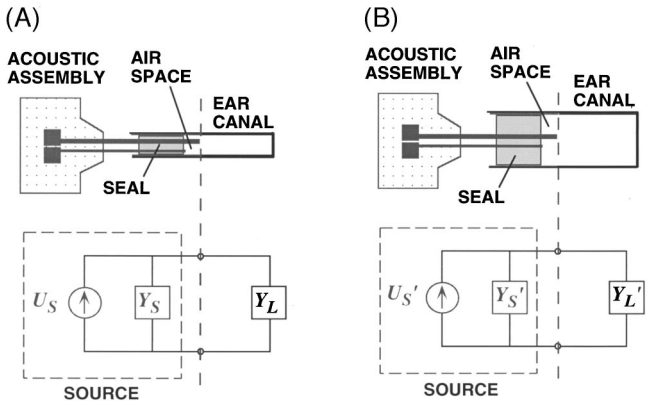


FIG. 4. Diagrams illustrating differences in the ''source'' with the acoustic system (ER-10C) coupled to ear canals of different diameters: (A) human infant or small cat species; (B) human adult or large cat species. To simplify the figure, the microphone and earphone tubes are shown without the eartip's outer material (Fig. 1). The dashed vertical lines at the locations of the microphone-tube tips mark the interfaces between the sources and loads. In the circuit analogs (lower portion), the source is represented by an ideal volume-velocity source (U_S or U_S') in parallel with a source admittance (Y_S or Y_S'). The ear, or test load, is represented by a load admittance (Y_L or Y_L'). Differences in configuration of both the seal and the air space around the microphone-tube extension can affect the source parameters U_S and Y_S.

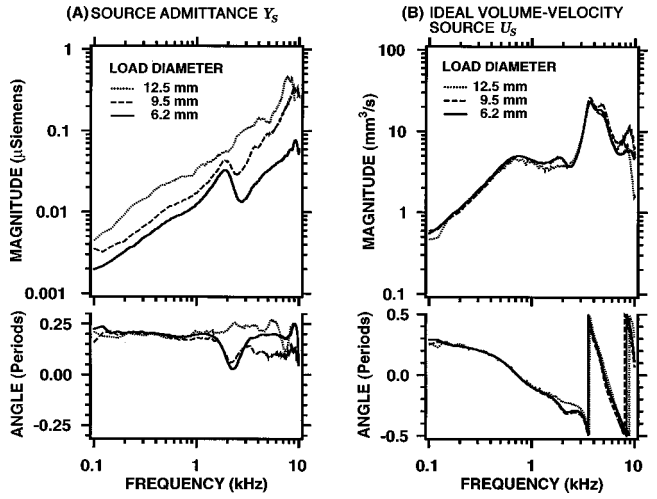


FIG. 5. Effects of load-diameter variations on the source parameters. Norton equivalent source parameters (computed from measurements) are shown for three diameters (6.2 mm, 9.5 mm, and 12.5 mm): (A) source admittance Y_S; (B) ideal-source volume velocity U_S, for 1-volt-peak stimulus level. For each diameter, Y_S and U_S were determined from measurements in two ''reference'' loads: (1) a closed tube of length 8–12 mm, and (2) a tube of length 15 m terminated by an acoustic resistor. To compute U_S, the microphone was calibrated to determine absolute sound-pressure levels.

earphone tube, but *does not eliminate* them for frequencies above 6 kHz. The result that the ''large source'' (with about the same signal-to-noise ratio) measures admittance more accurately than the ER-10C above 6 kHz is consistent with the idea that nonuniform waves are still a problem for our measurement system (ER-10C with microphone-tube extension) at higher frequencies.

2. Intersubject variations in ear-canal diameter

a. General configuration. Human ear canals (including infants and adults) vary in diameter from 5 to 11 mm (Keefe *et al.*, 1993). Figure 4 shows the Etymotic ER-10C system inserted into loads having diameters similar to that of an infant's ear canal [Fig. 4(A)] and an adult's ear canal [Fig. 4(B)]. The sound source undergoes two main changes from load (A) to load (B): (1) the volume of air space around the microphone-tube extension becomes greater; (2) the seal is less compressed because it fills a larger diameter. These two factors can affect the characteristics of the effective source, as all components to the left of the vertical dashed lines are part of the source. Thus the source parameters Y_S and U_S for load (A) are, in general, different from Y_S' and U_S' for load (B). In this section, we explore how diameter variations of the load affect these parameters and the measurement of admittance.

b. Effects of diameter variations on source characteristics. For a given load diameter, the source parameters—admittance Y_S and volume velocity U_S—are determined from measurements in two calibration loads of known admittance. Figure 5 shows (A) Y_S and (B) U_S determined from measurements in calibration loads of three different diameters spanning the (ear canal) size range of interest. The source admittances are all roughly compliant up to 1 kHz, but they vary in frequency dependence for higher frequencies. The source-admittance magnitude increases monotonically with increasing diameter of the load (except for the largest diameter at the highest frequencies). The Y_S curves span a factor of about 3 in low-frequency magnitude with a factor of 4 change in cross-sectional area of the load. This result is qualitatively consistent with the idea that increasing the air volume around the microphone-tube extension and expanding the foam plug result in larger contributions of these structures to $|Y_S|$. The U_S curves, in contrast, are fairly independent of the load diameter. An interpretation is that the volume-velocity source is determined mainly by the characteristics of the earphone and the dimensions of the narrow earphone tube. In summary, as load diameter increases, $|Y_S|$ increases and $|U_S|$ is approximately constant.

c. Effects of diameter variations on measured admittance. The dependence of the source characteristics on the load diameter implies that knowledge of the ear canal diameter is important for accurate admittance measurements in ears. That is, the ''calibration diameter''—the diameter of the calibration loads—should be equal to the canal diameter at the measurement point. In practice, however, the canal diameter is neither known accurately nor constant with position. The goal of this section is to quantify errors in measured admittance that result from differences between calibration diameters and test-load diameters.

Our approach is to compare measured admittances that are computed with source parameters based on different calibration diameters. Admittance measurements are shown in Fig. 6 for a closed cylindrical test cavity of inner diameter 6.2 mm and length 31.3 mm. The measurement with the ''equal-diameter'' calibration agrees well with the theoretical curve up to 9 kHz, whereas with the ''larger-diameter'' calibration the admittance magnitude is low by 15%–20% below 0.6 kHz, and large errors occur in both magnitude and angle above 4 kHz. The low-frequency error results from attributing too great a fraction of the total acoustically ''sensed'' volume to the source and hence attributing too little volume

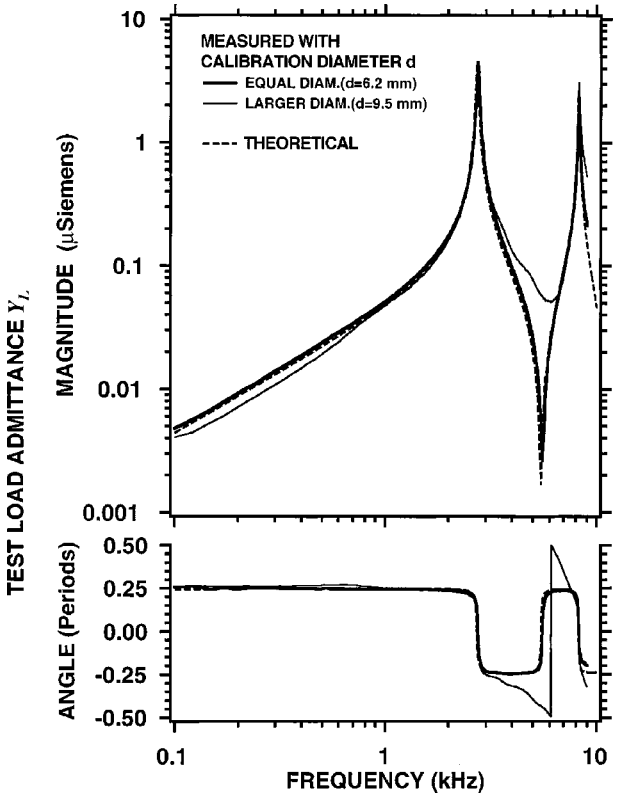


FIG. 6. Demonstration of errors in admittance measurements resulting from unequal test-load and calibration-load diameters. The test load is a closed cylindrical tube of inner diameter 6.2 mm and length 31.3 mm. The ''equal diam.'' curves were computed with source parameters based on calibration measurements in reference loads having the same diameter (6.2 mm). The ''larger diam.'' curves were computed with source parameters based on calibration measurements made in 9.5-mm-diameter reference loads. The data are cut off at 9 kHz because for higher frequencies, admittance measurements in test loads of this diameter (6.2 mm) have large errors.

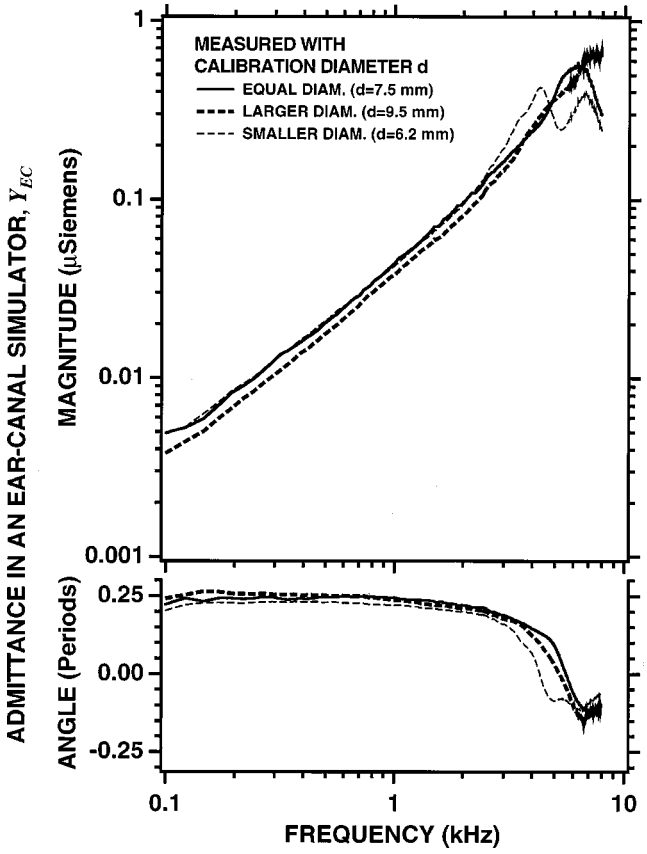


FIG. 7. Effects of different calibration-load diameters on the admittance measured in a commercial ear simulator. The ear simulator is a Zwislocki-type DB-100 coupler (Knowles) having inner diameter 7.5 mm. The measurement was made 15 mm from the inner termination (a microphone diaphragm). The three calibration diameters are the inner diameters of the reference loads used to compute three pairs of source parameters (and hence three admittances). The data are cut off at 8 kHz because for higher frequencies, admittance measurements in test loads of this diameter (7.5 mm) have large errors. The thick dashed line appears jagged and solid above 6 kHz.

(and compliance) to the load. The large higher-frequency errors are due in large part to the presence of the deep, sharp minimum in admittance magnitude near 5.5 kHz. In this frequency region, where $|Y_L| < |Y_S(d = 9.5\,\text{mm})|$, the computed admittance is very sensitive to small errors in Y_S [see Eq. (1)], and the measurement has large errors and fails to capture the shape of the minimum. These results demonstrate that use of a calibration diameter of human-adult size for measuring admittance in an infant-sized canal can produce two types of errors: (1) a low-frequency compliance (or equivalent volume of air) error of 15%–20%; (2) a gross failure to capture the shape of minima in admittance magnitude in the midfrequency range [which is relevant to middle-ear admittance minima reported in cats between 3 and 5 kHz (Møller, 1965; Guinan and Peake, 1967; Lynch *et al.*, 1994; Huang *et al.*, 1997)].

Next, admittance errors introduced by variations in calibration diameter are examined for a test load with a smoother, more human-earlike frequency dependence in the midfrequency range. Figure 7 shows admittances measured in an ear simulator (Zwislocki-type DB-100 coupler, Knowles Electronics, formerly Industrial Research Products). This load consists of a tube (inner diameter 7.5 mm) terminated by a microphone diaphragm, with four lossy, side-branch resonators within 5 mm of the termination.[8] We observe that an increase of 25% in calibration diameter leads to errors of approximately −15% in admittance magnitude below 2 kHz, which is similar to the result in the compliancelike region in Fig. 6. For higher frequencies, the admittance errors are larger, on the order of 40% in magnitude and 0.2 periods in angle for the ''smaller-diameter'' calibration. These results demonstrate that differences of 15%–25% in calibration diameter can lead to sizable errors in both magnitude and angle—particularly above 2 kHz—in an adult-sized simulated ear canal.

To avoid admittance errors of this size in ears and test loads, we have found that the calibration diameter should be within 15% of the load diameter to measure the low-frequency magnitude within 10% and to capture the shape of sharp midfrequency features such as in Fig. 6.

C. Admittance-location transformation

1. Context

With the eartip inserted into the external ear canal, the acoustic system measures the admittance Y_{EC} roughly 10 mm from the TM (Fig. 1). Having addressed above the issues in measuring Y_{EC}, we focus here on the problem of inferring the admittance at the TM, Y_{TM}, via an admittance-

location transformation. The goals of this section are (1) to define the uniform-tube transformation and (2) to test its accuracy with a relatively simple acoustic load.[9] We demonstrate the ''baseline'' accuracy of the location transformation in a load for which the parameters of the transformation can be directly measured. (Application of the transformation to ears is described in Sec. II.)

2. Uniform-tube model

The transformation is defined in terms of locations along the ear canal. To correct for the difference in location between the measured admittance (Y_{EC}) and the desired admittance (Y_{TM}), a lossless cylindrical-tube model of the ear canal space has been employed (Møller, 1965; Zwislocki, 1970; Rabinowitz, 1981; Lynch *et al.*, 1994):

$$Y_{\mathrm{TM}} = Y_0 \frac{Y_{\mathrm{EC}} - jY_0 \tan(kl)}{Y_0 - jY_{\mathrm{EC}} \tan(kl)}, \tag{5}$$

where Y_{EC} is the admittance measured at the microphone-tube tip; $Y_0 = \pi a^2/(\rho_0 c)$ is the characteristic admittance of the tube, l is the length of the tube, a is the radius of the tube, $k = 2\pi f/c$ is the wave number, ρ_0 is the density of air, c is the propagation velocity of sound in air, and f is the frequency.[10] This transmission-line model has two parameters, the length l and radius a of the tube, corresponding to the effective dimensions of the ear-canal space. (A procedure for determining these parameter values in ears is discussed in Sec. II B.)

3. Test of transformation in simulated ear canal

To test the accuracy of the admittance-location transformation with our modified ER-10C system, we used a stable acoustic configuration that *is* a uniform tube, the Zwislocki-type DB-100 ear coupler (see Fig. 7), and made admittance measurements at several positions along the length of the tube. The eartip's position was controlled by a micromanipulator and checked with calipers when possible. In Fig. 8, we report measurements made in the coupler at two positions 10 mm apart, and compare the more ''medial'' measurement (Y_{MEDIAL}) with an estimate obtained by applying the transformation [Eq. (5)] to the more ''lateral'' measurement (Y_{EC}). The measured and estimated Y_{MEDIAL} curves agree within 20% in magnitude and 0.04 periods in angle up to 6 kHz. Above 6 kHz, the magnitudes start to separate, differing by 32% at 8 kHz.

The higher-frequency differences between the measured and estimated Y_{MEDIAL} curves (generally less than 30% in magnitude and 0.05 periods in angle) are similar in size to errors in direct admittance measurements made in other test loads [e.g., the thick curves in Fig. 2(B)]. These errors probably result from nonuniform waves generated at the earphone-tube tip in the test load and the calibration loads. The ''large source'' result in Fig. 3 shows that the location transformation does not necessarily introduce additional errors. Taken together, the results of Fig. 3 and Fig. 8 demonstrate that the ''baseline'' accuracy of the transformation is limited by the accuracy of our admittance-measurement system. That is, with an accurate description of a uniform-canal space, the transformation does not introduce additional errors.

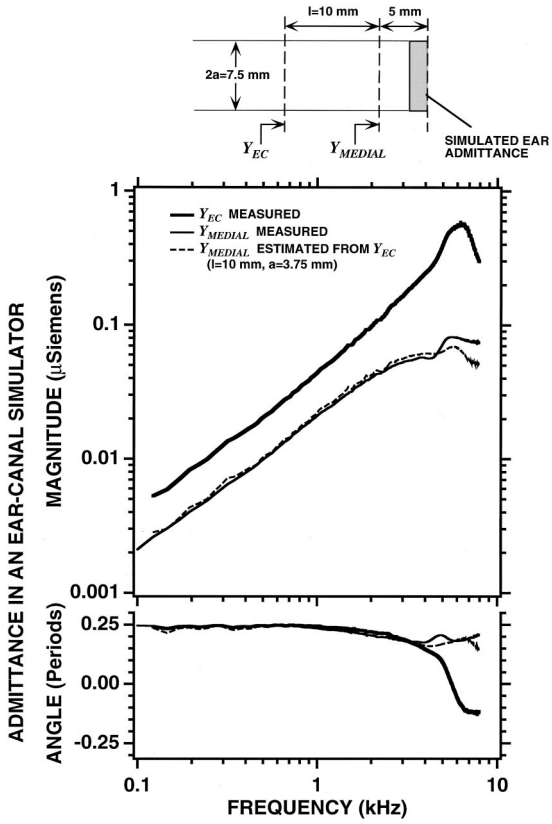


FIG. 8. Test of the admittance-location transformation in an ear simulator (see Fig. 7 caption). The shaded area (''simulated ear admittance'') represents a region of the simulator in which side branches occur, within 2–4 mm of the termination. Y_{EC} was measured 15 mm from the inner termination. Y_{MEDIAL} was measured 10 mm ''medial'' to Y_{EC}, i.e., 5 mm from the inner termination. ''Y_{MEDIAL} estimated'' was computed by transforming Y_{EC} through a uniform tube having diameter 7.5 mm and length 10 mm, via Eq. (5).

II. APPLICATION OF METHOD TO INTACT CAT EARS

A. Materials and methods

1. Subjects

Treatment of experimental animals conformed to the guidelines of the National Institutes of Health, and experimental protocols were approved by the Animal Care committees at the Massachusetts Institute of Technology and at the Massachusetts Eye and Ear Infirmary.

Acoustic measurements were made on eight adult domestic cats weighing between 2.4 and 3.3 kg. In this report, individuals are referred to by chronological order of experiments (1–8), which were performed over the course of one year. Cats 1–2 were subjects of other studies and were used primarily for development of basic methods. Although some measurements were made on all 12 ears of cats 3–8, experimental complications required the exclusion of eight ears from tests of the admittance-location transformation (Huang, 1999, Table 2.1; Huang *et al.*, 2000b); data from these ears are not reported. The most common problems were (1) failure to obtain a stable seal of the eartip in the intact canal

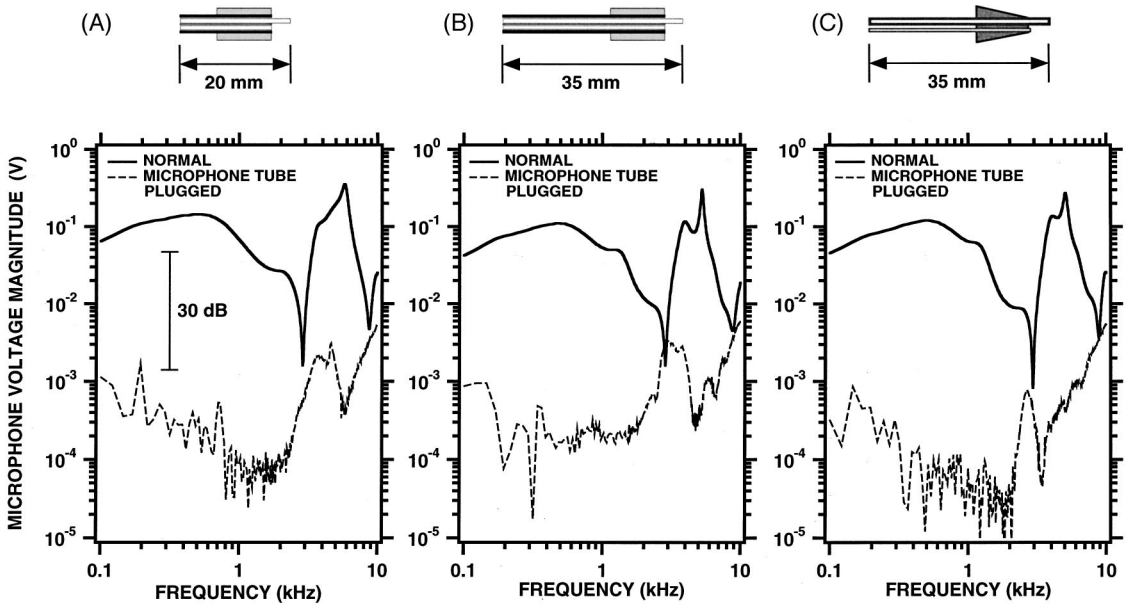


FIG. 9. Comparison of microphone output levels for different types of eartips: (A) standard Etymōtic eartip (length 20 mm), (B) longer Etymōtic eartip (length 35 mm), (C) custom eartip constructed from two separate plastic tubes (length 35 mm): microphone tube's inner diameter 1.6 mm, outer diameter 2.9 mm; earphone tube's inner diameter 0.5 mm, outer diameter 1.5 mm. The test load is the rigid cavity of Fig. 6. To estimate the level of ''artifact'' for each eartip, the voltage response of the microphone was measured before (''normal'') and after plugging the distal end of the microphone tube with a piece of steel wire coated with petroleum jelly. To the extent that the ''microphone tube plugged'' magnitudes are substantially below the ''normal'' measurements, the artifact is not significant.

(cats 3–4) and (2) damage to the TM (cats 5–7).[11] Except where noted, reported measurements were made on ears with clean ear canals, healthy-appearing, translucent TMs, no signs of middle-ear infection (as examined through an otoscope), and no indication of middle-ear static pressure buildup (based on repeated measurements over 5–10 min).

2. Animal preparation

Cats 3–8 were anesthetized with intraperitoneal injections of Dial (75 mg/kg), with booster doses given as indicated by a withdrawal response to a toe pinch. Minimal exposures of both auditory bullae were made, to allow middle-ear cavity venting without disturbing the muscles of the pinnae. A pointed scalpel was used to drill a small hole (diameter 1–2 mm) in the posterior part of each bulla, through which a plastic vent tube (inner diameter 0.3 mm, length 90–100 mm) was inserted and cemented to the bulla. Cats 5 and 6 exhibited some bleeding from the mucosal lining of the bulla; fluid and blood were periodically removed from the cavity by fine absorbent paper points. For cats 7 and 8, the bulla hole was made slightly larger to allow a cleaner incision in the mucosal layer, and the vent was inserted through a small metal nut cemented to the bone. There was no problem with bleeding in these two subjects. During the course of measurements, the vents were periodically removed to check for fluid in either the bullar cavity or the vent and then replaced.

3. Motivation for a custom eartip

a. Geometry of the cat's external ear. Initial attempts to insert and seal the acoustic system (as in Fig. 1) into intact ear canals of cats were unsuccessful because of the geometry of the cat's external ear. The concha narrows toward the entrance of the ear canal, and a sharp bend occurs at the canal-concha border (Rosowski *et al.*, 1988, p. 1699). This geometry makes it very difficult to visualize the eartip and the canal entrance as the eartip is inserted. As a result, in initial attempts the microphone-tube tip was often in the concha and/or a poor acoustic seal occurred.

We found that the standard Etymōtic eartip (length 20 mm), when coupled to the acoustic assembly, was not long enough to reach through the cat's concha and into the ear canal. Even with longer eartips, problems continued because foam and rubber seals (Etymōtic) were too large and/or too inflexible to navigate the sharp bend at the canal entrance and static seals were not possible with the foam material. Consequently, we investigated the use of a longer eartip and a different type of seal for use in cats' ears.

b. Earphone-microphone cross talk. In developing an eartip for use with cats, we found that the length of the eartip affected the accuracy of admittance measurements. Measurements made with longer eartips (30–35 mm) of the same material as the standard Etymōtic device had large errors near the frequency of the admittance-magnitude maximum at the quarter-wave resonance of rigid-walled test loads. It was hypothesized that the errors were the result of low sound-pressure levels, which made acoustic coupling within the eartip from the earphone tube to the microphone tube significant. To test this idea, acoustic measurements were made of the level of this ''cross talk.''

Figure 9 shows microphone output voltages measured in a test load with the microphone tube open (normal) and plugged at its distal tip, for three different types of eartips. The microphone voltage response with the microphone tube plugged is an estimate of the ''artifact'' level, where artifact

refers to the portion of the microphone output resulting from signals other than the sound pressure at the tip of the microphone tube. Possible sources of artifact include mechanical vibration, acoustic coupling through the walls of the earphone and microphone tubes, and electric coupling between the input to the earphone and output of the microphone. Artifact measurements were made with the eartip inserted into different loads (including free space) and were found to be essentially independent of the load. The dependence of artifact level on the type of eartip (demonstrated in Fig. 9) was repeatable and stable over time.

With the standard Etymotic eartip [Fig. 9(A)], the signal-to-artifact ratio is at least 30 dB except near the signal's sharp minimum near 3 kHz (quarter-wave resonance of the load), and above 7 kHz. For measurements in intact cat ears, an eartip with longer-than-''standard'' sound tubes is necessary to reach the ear canal. With a longer eartip of the same configuration (specially ordered from Etymotic), the artifact level increases in the midfrequencies, from 0.8 to 4 kHz [Fig. 9(B)]. In fact, a clear progression of increasing midfrequency artifact level occurs with eartips of increasing intermediate lengths (not shown). The measured signal at the location of the minimum near 3 kHz is primarily artifact, which results in large errors in the measured admittance. Thus in this case, it could be difficult to capture the shape of a sharp midfrequency feature such as the cat's middle-ear-cavity admittance minimum, whose effect in the ear canal can be a pressure *minimum* (Huang *et al.*, 1997).[12]

In response to this demonstration of significant artifact, we constructed ''custom eartips'' from separate earphone and microphone tubes made of Tygon [Fig. 9(C)]. This eartip has a greater wall thickness between the microphone and earphone tubes than the Etymotic eartip and still couples easily to the ER-10C acoustic assembly. With this eartip, the measured artifact level is reduced in the range 0.5–4 kHz (to the noise level for 1–2 kHz), and is generally smaller than that of the standard Etymotic eartip for those frequencies. Our interpretation is that for the midfrequencies (1–6 kHz), the artifact observed with the ''longer eartip'' is the result of acoustic cross talk between the microphone and earphone tubes. The custom eartip reduces the acoustic cross talk and allows the pressure minimum to be resolved.

For frequencies above 7 kHz, all three artifact estimates increase sharply and are quantitatively similar (Fig. 9). This response is observed even when the microphone amplifier is switched off but is not seen when the stimulus is disconnected. These observations are consistent with the idea that in this frequency region, electric cross talk in the measurement system dominates. This electric cross talk, which could limit the accuracy of admittance measurements above 10 kHz, probably originates in our DSP board and (in hindsight) may be reduced by using a power amplifier on the output of the digital-to-analog converter.

4. Admittance-measurement procedure

The coupling of the acoustic system to the intact ear canal of a domestic (or exotic) cat is illustrated schematically in Fig. 10. The custom eartip consists of separate earphone and microphone tubes (length 30–35 mm), both of which

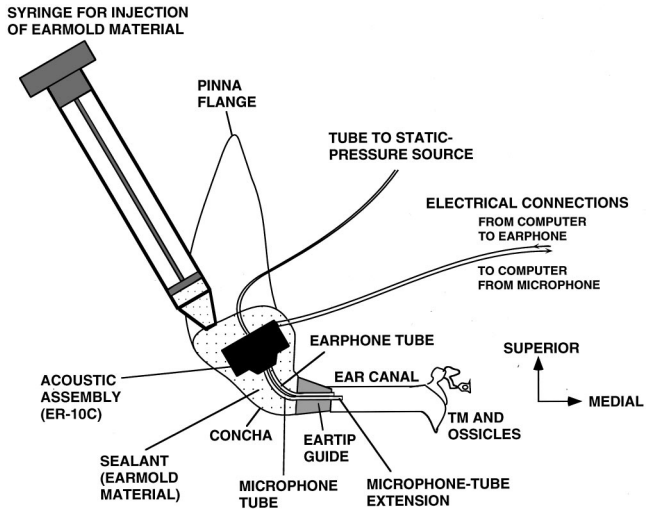


FIG. 10. Method of noninvasive acoustic measurements in an animal ear. This schematic representation does not attempt to portray the ear's anterior–posterior configuration or the irregular shape of the ear canal. The custom-made eartip—two plastic tubes that pass through a silicone ''eartip guide''—is inserted medial to the sharp bend at the canal-concha border, which is a complicating feature in many animal species (e.g., cats). Hearing-aid earmold-impression material is injected behind the eartip guide via syringe to form a static-pressure seal in the ear canal and to hold the eartip in place. The tube in the back of the acoustic assembly (Etymotic ER-10C) couples a static pressure source to the ear canal through the microphone tube. A cable from the acoustic assembly carries the electric stimulus and response signals to and from the computer.

pass through a flexible, conical ''guide'' made from Permatex Blue RTV silicone sealant. The guide holds the tubes together and aids in sealing the eartip into the canal. Eartips with guides that narrow medially are easiest to insert into cat ear canals. A typical guide has length 10 mm, lateral diameter 7 mm, and medial diameter 5 mm.[13] An orthodontic rubber band (diameter 4.8 mm, force value 3.5 ounces, not shown in Fig. 10) doubled around the tubes just lateral to the guide provides cohesion and support. The microphone tube protrudes 3–4 mm past the earphone-tube tip on the medial end.

Our measurement procedure was as follows. First, the concha, ear canal, and TM were examined with an otoscope for wax or other irregularities (e.g., blood or mites). A mold was then made of the concha and the lateral portion of the ear canal in two steps. First, a cotton dam was inserted into the canal with curved forceps to stop the flow of earmold material lateral to the TM. Second, hearing-aid earmold impression material (Westone Laboratories, Colorado Springs, CO) was injected into the concha via a large syringe. The earmold, which hardened in 3–4 min, provided a record of the size and shape of the concha and of the canal near the desired position of the acoustic probe. After examination of the earmold, the acoustic system was placed (by hand) such that the microphone-tube tip was medial to the bend at the canal-concha border. Pulling the pinna flange in a posterodorsal direction tended to straighten the bend and facilitate the insertion of the eartip. Responses to a few acoustic stimulus presentations were viewed on an oscilloscope and the computer as quick indicators of the eartip's position and the patency of the sound tubes.

To form a static-pressure and acoustic seal, earmold im-

pression material (''sealant'' in Fig. 10) was injected into the concha behind the eartip guide. After the sealant hardened, static pressure was introduced into the ear canal manually via an air-filled syringe and measured with a water manometer. If the static pressure was not stable, the acoustic system and sealant were removed from the ear and the insertion procedure repeated. With a functioning static seal, admittances were measured with static pressures in the ear canal for two purposes: (1) to sense any buildup of static pressure in the middle ear; (2) to help to determine the dimensions of the canal space between the measurement point and the TM (see Sec. II B). Generally, the static-pressure level (in mm H_2O) was varied in steps from 0 to -300, 0 to $+200$ ($+300$ was not maintained in most ears), and back to 0. A few (2–4) admittance measurements were taken at each zero-static-pressure setting. Returning the static pressure to zero did not seem to restore the initial condition immediately; rather the admittance tended to ''drift'' back to within 20% of the initial measurement after 30–40 s. After completion of the measurement series, the acoustic system and seal were removed, examined, and compared with the initial earmold to give a rough indication of the eartip's position in the canal during the measurement.

With the custom eartip, the calibration of the acoustic system was slightly different from that reported in Sec. I (with standard Etymotic eartips), in both procedure and source characteristics. Measurements were made in two calibration loads and at least one test load with earmold sealant injected behind the eartip guide to approximate the configuration in the ear. A calibration was performed for each eartip, because of variability in the shape of the guides and the extent of the sealant in the ear.[14] The source admittance Y_S and volume velocity U_S associated with the custom eartips have some quantitative differences from those reported in Fig. 5: (1) Y_S has a different structure of local extrema from 1 to 4 kHz; and (2) $|U_S|$ is lower by 5–8 dB above 4 kHz. These differences are probably due to the greater length of the custom eartip. However, the trend with load diameter is the same, and the accuracy of admittance measurements in closed test loads is similar to the best accuracy shown in Fig. 2(A) and Fig. 6 (see Huang, 1999, Fig. 3.1).

5. Operational difficulties

The process of inserting and sealing the eartip into a cat's intact ear canal was rather difficult. The overall success rate in ears was roughly 50%, where success is defined as a stable and complete acoustic and static-pressure seal of the eartip in the intact canal during admittance measurements. That is, in about half of the attempts, the acoustic system (and sealant) had to be removed, cleaned, and reinserted into the ear. The main reason for this rather low success rate is the restrictive geometry of the cat's concha and canal (not accurately portrayed in Fig. 10). This geometry makes insertion of the eartip into the canal an uncertain task, even with the custom eartip and guide, because movement of the guide is influenced (and impeded) by complexities in the shape of the cartilaginous walls which cannot be seen during the insertion.

The narrow shape of the concha also impedes the placement of the syringe for injection of the sealant. For a complete seal, the sealant material needs to surround the earphone and microphone tubes in the concha to the lateral edge of the guide. Achievement of such a seal often required several attempts. [The consequences of a faulty ear canal seal—in which static pressure holds in the canal but the sealant does not fill in the space lateral to the guide—are discussed in Huang *et al.* 2000b] These operational difficulties introduced large delays in making some of the measurements, as repeated insertions were required.

B. Transformation from Y_{EC} to estimated Y_{TM}

1. Approach and preview

In this section, we describe the procedure for selecting the parameter values [see Eq. (5)] used to transform Y_{EC} into the estimated admittance at the TM, Y_{TM}(est.). The procedure consists of (1) the use of an acoustic area-estimation algorithm (Sec. II B 2 a) and (2) the measurement of admittances with varying static pressures to estimate the volume of the canal space (Sec. II B 2 b). We report admittances measured in the intact ear canals of cats (Y_{EC}) and the results of the admittance-location transformation in these ears [Y_{TM}(est.)] (Sec. II B 3).

2. Choice of transformation parameters

a. Radius. The uniform-tube transformation [Eq. (5)] has two parameters, radius and length. For each ear, the parameter values were determined from acoustic measurements in the ear. First, the radius of the ear canal at the location of the microphone-tube tip was estimated from an algorithm described and tested (in artificial loads) by Keefe *et al.* (1992). Described briefly, the time-domain formulation assumes that the pressure response at ''time zero'' (i.e., the first sample in time) includes only a forward-going wave and is thus determined by the characteristic admittance of the load. This transient response can be expressed in terms of the measured admittance (discretized in frequency), which yields an expression for the radius of the load;

$$a \approx \sqrt{\frac{\rho_0 c}{(\pi/N)\sum_{i=1}^{N} \mathrm{Re}\{1/Y_L(i)\}}}, \tag{6}$$

where Y_L is the measured load admittance, Re{ } denotes the real part, i is the frequency index, N is the number of frequency samples, and the summation gives the inverse discrete Fourier transform evaluated at time zero.

For cat ear canals, we found that the algorithm gave radius estimates that were within 10% of estimates based on earmold impressions, provided that the summation range excluded admittance values with negative real parts (errors which often occurred at the lowest and highest frequencies). Typically we used $N=300$ frequency samples, uniformly spaced over the bandwidth 0.1–7.4 kHz, because for this frequency range admittances measured in test loads were generally as accurate as the thick curves in Fig. 2(A).[15] The final estimate of radius was computed as the mean of estimates from all admittance measurements made with zero static pressure in the canal (7–10 measurements per ear).

For three ears, a calibration-iteration procedure was performed. If the acoustically estimated ear canal diameter differed from the calibration-load diameter by more than 10%, a new acoustic calibration was performed (with calibration diameter equal to the estimate). The ear canal admittances were then recalculated with the new source parameters. From the new admittances, a new diameter was estimated from Eq. (6). In each case, the new diameter estimate ''converged'' to within 10% of the new calibration diameter. Our interpretation is that the new calibration gave a more accurate estimate of the source parameters for these ear canals.

b. Length. Admittances measured with static pressures in the ear canal were used to estimate the volume of the canal space (and hence its length, given the radius above). Figure 11(A, B) shows admittances Y_{EC} from one ear as a function of frequency, for different static-pressure levels in the ear canal. For frequencies below 1 kHz, the admittance magnitude decreases monotonically with increasing levels of static pressure (of either sign) in the canal; in this frequency region, the static pressure reduces the motion of the TM. For higher frequencies, the behavior of Y_{EC} with static pressure is more complicated—strong asymmetries occur with respect to positive and negative pressures, including several sharp features in the positive-pressure admittances. Near 2 kHz, the admittances measured with static pressures are *greater* in magnitude than with zero pressure. These changes with static pressure are repeatable and representative of our data.

The volume of the intact ear canal space V_{EC} is estimated from the admittance Y_{EC} measured at the most negative static pressure. First, from the Y_{EC} data we compute the low-frequency (0.1–0.3 kHz) compliance, or equivalent air volume, as a function of static-pressure level [Fig. 11(C)]. For all ears, this compliance was found to be largest at zero static pressure and smallest at the most negative static pressure (-300 or -400 mm H_2O). We assume that at the most negative static pressure the TM is most nearly rigid, and we take the corresponding equivalent volume as an estimate of the canal volume [e.g., $V_{EC}=0.37\,\mathrm{cm}^3$ in Fig. 11(C)]. This approach is consistent with the results of Shanks and Lilly (1981). The canal volume (V_{EC}) and the radius estimate (a) from Eq. (6) determine the length l of the model tube, $l=V_{EC}/(\pi a^2)$. The values of l and a define the ear canal transformation, via Eq. (5).

In Fig. 11, the relative changes in admittance magnitude below 1 kHz (and in compliance) with static pressure are generally smaller than those previously reported in cats (Møller, 1965, Fig. 11; Margolis *et al.*, 1978, Fig. 4; Lynch, 1981, p. 250). An explanation is that in previous studies, measurements were made within a few millimeters of the TM, whereas in Fig. 11 measurements were made in the intact ear canal at a more lateral position. As the total measured compliance is the sum of the canal compliance and the middle-ear compliance, we would expect the more lateral measurement to be relatively less sensitive to changes in the TM stiffness. Thus the direction of the discrepancy is consistent with the difference in measurement location.

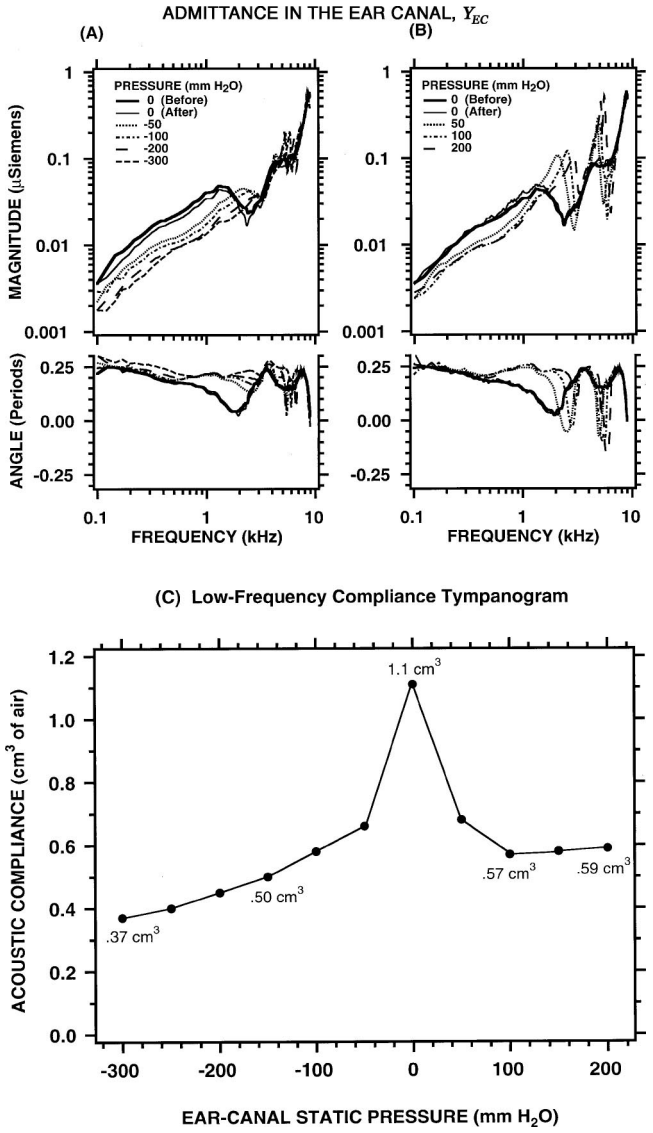


FIG. 11. Effects of static-pressure variations in the ear canal on admittances Y_{EC} measured in an intact ear (cat 3). (A) Negative-static-pressure series. (B) Positive-static-pressure series. (C) Low-frequency compliance tympanogram, computed from (A) and (B). Static pressures in the canal, given in mm H_2O (1-mm $H_2O=0.98$ Dekapascals$=9.8$ Pa), were produced with a syringe and measured with a water manometer. Admittance measurements with zero static pressure were made before and after each pressure series. The duration between the ''Before'' and ''After'' measurements was approximately 5 min. In (C), data points taken at discrete static-pressure levels are connected by straight line segments. For each level, the compliance was calculated as the average of $\mathrm{Im}\{Y_{EC}(f)\}/(2\pi f)$ over 10 frequency points f between 0.1 and 0.3 kHz, where Im{} denotes the imaginary part; compliances are expressed as equivalent volumes of air, computed as $(\rho_0 c^2)C(\rho_0 c^2)$, where ρ_0 is the density of air and c is the speed of sound in air. The compliance with -300-mm H_2O was used as an estimate of the volume of the ear-canal coupling space (0.37 cm^3 for this ear).

3. Estimates of Y_{TM}

In this section, we report estimates of the middle-ear input admittance [Y_{TM}(est.)] in domestic cats, based on measurements in the lateral portion of the intact ear canal (Y_{EC}).

The admittances Y_{EC} measured in four intact ear canals are qualitatively similar [Fig. 12(A)]. They are generally compliancelike, with equivalent volumes between 0.5 and 1.1 cm^3. For frequencies between 2 and 3 kHz, each admittance has a dip in magnitude and a positive angle change.

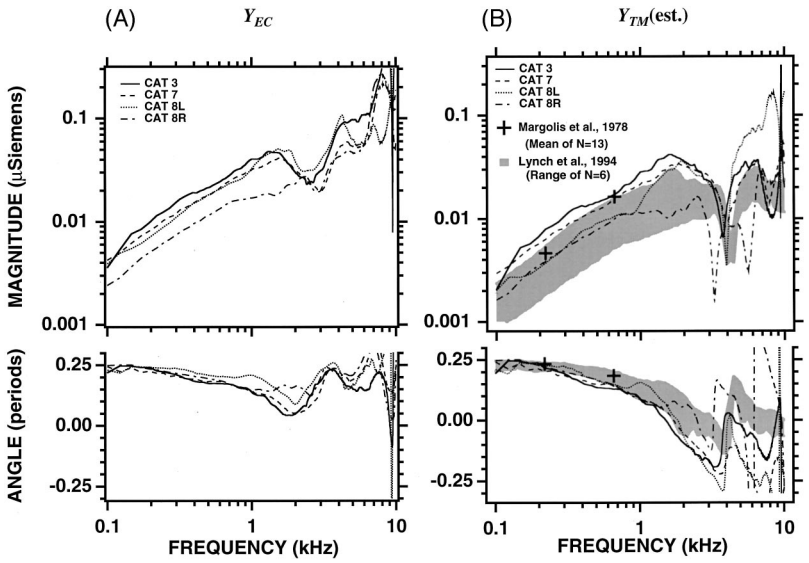


FIG. 12. Transformation from (A) measured admittance in the intact ear canal (Y_{EC}) to (B) estimated admittance at the TM [Y_{TM}(est.)] for four cat ears. The spatial transformation is described by Eq. (5). The Y_{EC} curves were measured with zero static pressure. Parameter values for the transformation (estimated canal length l and radius a) were based on (1) a volume estimate (V) from Y_{EC} measured with large negative static pressure in the canal (see Fig. 11 caption) and (2) a radius estimate from Y_{EC} and Eq. (6) [after Keefe *et al.* (1992)]: cat 3, $l=9.2$ mm, $a=3.5$ mm, $V=0.37$ cm^3; cat 7, $l=10$ mm, $a=3.1$ mm, $V=0.31$ cm^3; cat 8L, $l=11$ mm, $a=3.6$ mm, $V=0.45$ cm^3; cat 8R, $l=8.0$ mm, $a=3.1$ mm, $V=0.25$ cm^3. The crosses are mean Y_{TM} data from Margolis *et al.* (1978, Table I), measured in 13 cat ears at 0.22 kHz and 0.66 kHz. The shaded regions represent the range of Y_{TM} measured by Lynch *et al.* (1994) in six cat ears. The previous data were obtained with most of the external ear removed.

Between 6 and 8 kHz, three of the four measurements (cats 7, 8L, and 8R) have angles greater than 0.25 periods (i.e., negative real parts), which suggest errors in the source calibration for these frequencies.

The admittance-location transformation has three main effects on each of the measured admittances [Fig. 12(B)]: (1) Below 1 kHz, Y_{TM}(est.) remains compliancelike, with $|Y_{TM}(\text{est.})|<|Y_{EC}|$ by an amount corresponding to the estimated compliance of the ear canal space. (2) Between 3 and 4 kHz, Y_{TM}(est.) exhibits a sharp notch in magnitude and a sharp positive angle shift, which are not present in Y_{EC}. (3) Above 4 kHz, Y_{TM}(est.) has large variations with frequency in both magnitude and angle; the angle of Y_{TM}(est.) is generally more negative than that of Y_{EC}.

Previous measurements of Y_{TM} in cats—made within a few millimeters of the TM after removal of the external ear—can be used to assess the gross features of our estimates. In Fig. 12(B), our noninvasive estimates of Y_{TM} are compared to low-frequency mean data from Margolis *et al.* (1978) and to the range of data from Lynch *et al.* (1994). [For further comparisons between the cat data of Lynch *et al.* (1994), Møller (1965), Margolis *et al.* (1978), and Puria and Allen (1998), see Lynch *et al.* (1994, Fig. 19).] The Y_{TM}(est.) curves have several features in common with the previous data. Both are compliant below 1 kHz, become more resistive between 1 and 2 kHz, and exhibit a sharp notch in magnitude and positive transition in angle near 4 kHz. This sharp feature—which we will call the ''cavity notch''—results from a distinctive middle-ear cavity antiresonance in cats (Møller, 1965; Guinan and Peake, 1967; Peake *et al.*, 1992; Huang *et al.*, 1997).

For frequencies below 1 kHz, the equivalent volumes of our estimates (0.35–0.73 cm^3) overlap with the Margolis data and the upper portion of Lynch's range of equivalent volumes (0.17–0.54 cm^3); two of the four estimates (cats 8L and 8R) are completely within Lynch's range, one estimate (cat 7) approximates the upper limit of Lynch's range, and one estimate (cat 3) exceeds the upper limit of Lynch's range by 25%–30% in magnitude. The cavity notches of $|Y_{TM}(\text{est.})|$ are generally within the range of Lynch's data (except for cat 8R), while the associated angles are more negative. The $|Y_{TM}(\text{est.})|$ curve of cat 8R exhibits notches at 3.2 kHz and 5.6 kHz, which differ in frequency from both Lynch's data and our results in other ears. Above 3 kHz, there is greater inter-ear variability in our estimates compared to Lynch's data, especially in angle. Angle magnitudes that are greater than 0.25 periods (occurring in three ears above 6 kHz) must be in error.

In summary, the estimated middle-ear input admittances Y_{TM}(est.) have features that are quantitatively similar to previous measurements of Y_{TM} in domestic cats.

C. Repeatability of Y_{TM} estimates

1. Effects of variations in parameter values

The sensitivity of the ear canal transformation to variations in its parameters determines the effects of errors in these parameters. The effects of small changes in the length l and radius a were computed for three ears. Representative results are reported for one ear (Fig. 13). The changes in Y_{TM}(est.) resulting from changes of $\pm 10\%$ in the parameter values are shown. The shapes of the magnitude and angle of Y_{TM}(est.) remain the same. In general, the admittance is more sensitive to the parameter variations at high frequencies than at low frequencies, and sharp spectral features are affected more than smooth spectral features. The depth of the cavity notch near 4 kHz is strongly affected (by nearly a factor of 4), as is the size of the sharp angle transition at the notch frequency. The frequency of the notch, however, is only changed by $\pm 5\%$, and the low-frequency compliance (or equivalent volume) is changed by less than $\pm 10\%$.

2. Effects of variations in eartip placement

If the admittance-location transformation is accurate, then the inferred admittance at the TM [Y_{TM}(est.)] should be independent of the location at which Y_{EC} is measured. Here, we report measurements that test this theoretical invariance as well as the repeatability of the canal measurement (Fig. 14). Two measurements made in each of two intact ears are plotted in the upper panels. For each ear, two ''insertions''

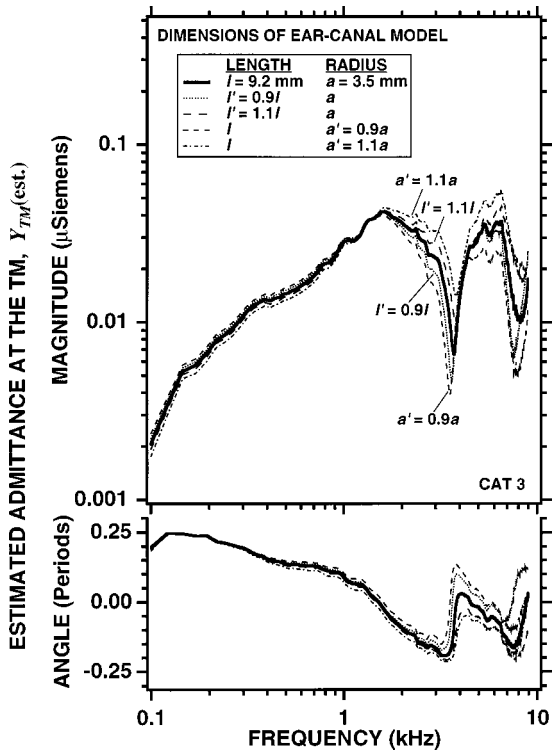


FIG. 13. Effect of variations of ear canal parameters on Y_{TM}(est.). Estimated admittances are shown for one intact ear (cat 3). The thick curve is the estimate from Eq. (5), with ear canal transformation parameters (length l and radius a) selected as described in the caption of Fig. 12. The other curves show the estimated admittances calculated with one ear-canal parameter altered by ±10% and the other parameter unchanged.

were made, where each insertion consisted of (1) coupling the acoustic system to the ear canal, (2) injecting the earmold material to form a static seal, and (3) measuring admittances with static pressures in the canal. The admittances Y_{EC} in the upper panels were measured with zero static pressure. The lower panels show the results of transforming the Y_{EC} curves to the location of the TM, Y_{TM}(est.), with the estimated dimensions of the canal space (l and a) indicated in the legends.

The admittances Y_{EC} from the two insertions in cat 3 are similar in shape [Fig. 14(A), upper]; the curves agree to within 40% in magnitude and 0.08 periods in angle. For cat 8, however, the two admittances Y_{EC} are quite different quantitatively and in their shape [Fig. 14(B), upper]. While the estimated canal dimensions are larger for cat 3 than for cat 8, all of the acoustically determined dimensions are anatomically reasonable. The length estimates from repeated insertions in these two ears range from 4 to 13 mm, while the total ear canal length of cats has been estimated to be between 9.5 and 18.9 mm; the radius estimates of 2.2–3.8 mm are similar to the 2.5–3.4 mm radii estimated from the cross sections of Rosowski *et al.* (1988, p. 1699). The previous anatomical descriptions also suggest a reason for the correlation between radius and length in these four insertions, in that the canal has been noted to be wider at the canal-concha border than near its midpoint.

The admittance measurements have features that are consistent with the estimated canal dimensions (Fig. 14). In cat 3, the probe position appears rather distant from the TM on both insertions and there is little difference between the two measurements. In cat 8, both insertions were closer to the TM than in cat 3, and the second insertion depth is estimated to be only 4 mm from the TM. This depth should lead to an admittance that is little affected by the remaining canal space and would be expected to be different in shape compared to measurements made further away, especially at high frequencies. The decreased effect of the canal in this measurement is seen in the similarity between Y_{EC} and Y_{TM}(est.) in cat 8, insertion 2.

The bottom panels (Fig. 14) demonstrate that the overall method gives repeatable estimates of Y_{TM} for frequencies up to roughly 5 kHz. For both ears, the two Y_{TM}(est.) curves have similar shapes in this frequency range and generally agree within 20% in magnitude and 0.05 periods in angle. The variations in measurement location with different eartip insertions are accounted for by the transformation procedure. For frequencies above the cavity notch (>5 kHz), the differences between the estimates are larger, which suggests that the error in the method is greater for high frequencies.

III. DISCUSSION AND CONCLUSIONS

A. Admittance-measurement issues

1. Microphone-tube extension

(a) With our admittance-measurement approach, if the earphone port is much smaller in diameter than the ear canal (as in the ER-10C system), then extending the microphone tube 3–4 mm past the earphone port is crucial for accurate admittance measurements above 2 kHz (Fig. 2).

(b) The main source of errors above 2 kHz is the large discontinuity in cross-sectional area between the earphone tube and the load, which generates nonuniform waves. The microphone-tube extension improves admittance accuracy for higher frequencies, but it does not eliminate the problem above 6 kHz. Support for these conclusions includes: (1) the increased accuracy of an acoustic system whose earphone port is larger than that of the ER-10C (Fig. 3); (2) the slightly larger errors in admittance for larger load diameters [compare Fig. 2(A) and Fig. 6]; (3) the fact that the total error observed in this frequency range is greater than that due to uncertainties in calibration-load lengths, low signal-to-noise ratio, and earphone-microphone cross talk (Fig. 9).

2. Variations in load diameter

(a) The source admittance varies substantially with the diameter of the ear canal or load; its magnitude increases with increasing diameter [Fig. 5(A)]. Thus knowledge of the canal diameter is essential in the acoustic-calibration procedure for a source of this configuration (with an extended microphone tube and flexible seal).

(b) Differences of 20%–25% between canal diameter and calibration diameter can lead to errors of 25% (2 dB) in admittance magnitude below 2 kHz and greater errors above 2 kHz (Fig. 7). We conclude that calibration measurements should be made in loads having diameter within 10%–15% of the canal diameter.

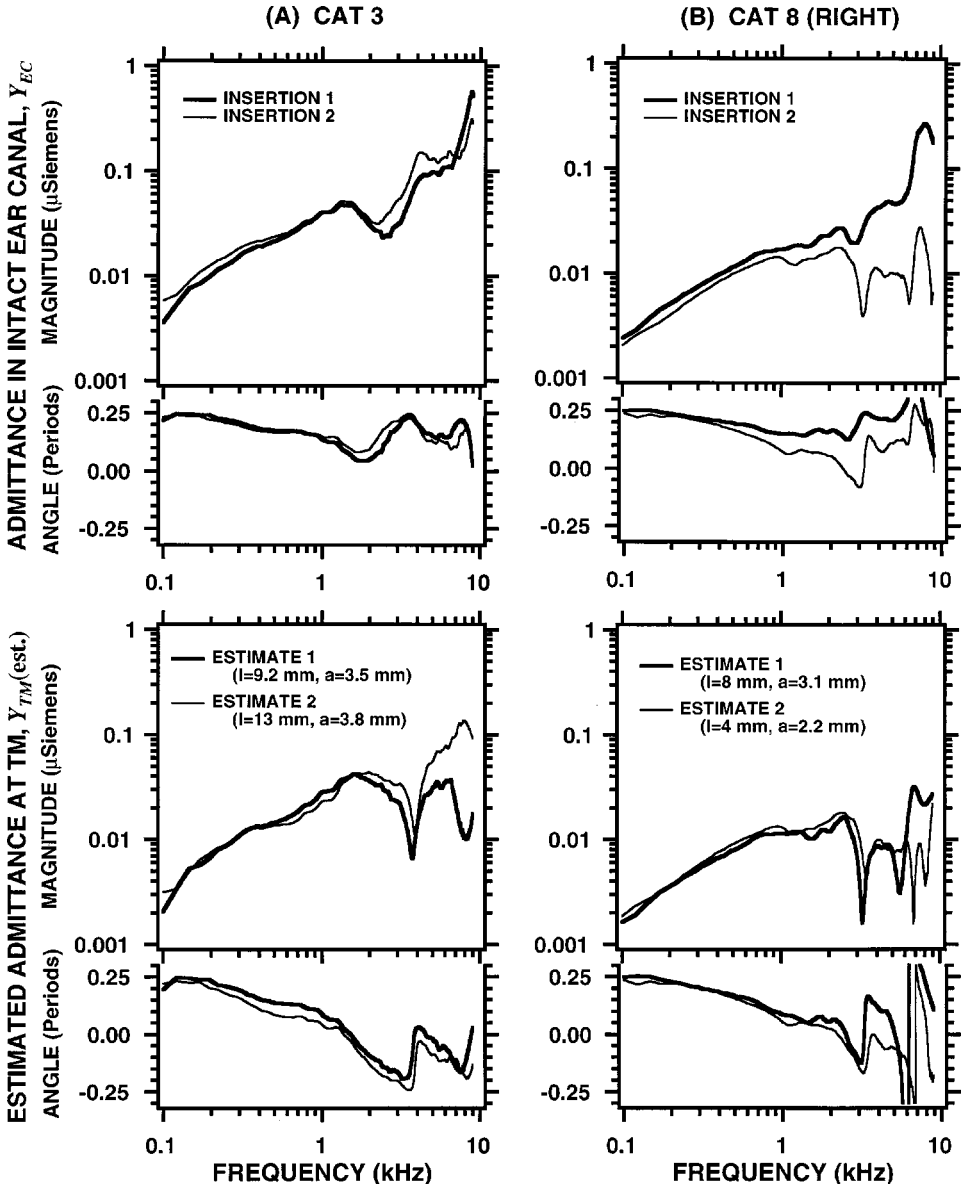


FIG. 14. Repeatability of Y_{TM}(est.) with two separate insertions of the acoustic system into each of two intact ears: (A) cat 3, and (B) cat 8 (Right). The UPPER plots show Y_{EC} curves for two intact ear canals. Data from these particular ears were selected to show (A) two quantitatively similar Y_{EC} measurements and (B) two quantitatively different Y_{EC} measurements. The LOWER plots show the Y_{TM}(est.) curves obtained by transforming each of the Y_{EC} curves in the upper plots to the TM via Eq. (5). For each ear, estimate 1 is based on insertion 1, and estimate 2 is based on insertion 2. Between insertions 1 and 2 (elapsed time 30–60 min), measurements were made with static pressures in the canal, and the acoustic system was removed from and replaced in the ear canal.

3. Design improvements

(a) In principle, the magnitude of nonuniform modes generated by the narrow earphone tube can be reduced with a high-impedance acoustic system with a larger earphone port (see Fig. 3). The earphone and microphone tubes should be housed in a flexible eartip that can be sealed into intact ear canals of different sizes (as in the ER-10C).

(b) A smaller acoustic assembly (or considerably longer eartips) would facilitate the placement and sealing of the acoustic system in the intact ear canals of cats and other animal species. The assembly's present size limits its mobility in the concha and makes difficult the insertion of both the eartip and the syringe for injection of sealant.

(c) Obtaining a stable and complete (i.e., space-filling) seal in the intact canal of cats is a continuing problem. The main reason is the deep, narrow concha and sharp bend at the canal-concha border. In this work, we had a rather low success rate per eartip insertion (roughly 50%). It might be possible to design a more effective method for sealing the acoustic system in the canal, e.g., through use of an inflatable cuff around the eartip that can be controlled from outside the canal and concha (Rabinowitz, 1981). Failing that approach, different sealant materials could be tested for their facility in reaching the eartip guide. For example, a more fluid material might flow more easily past the canal bend and fill in the space behind the guide more consistently.

4. Further work

(a) Measurements could be made to test the strength of nonuniform wave components, e.g., by varying the length of the microphone-tube extension. It is possible that use of a longer calibration load (>12 mm) would reduce the effects of nonuniform waves on the source calibration. Keefe *et al.* (1992) described ''evanescent mode interactions between the ends of the tubes,'' that degraded accuracy when the diameter-to-length ratio of the load was greater than 0.3. However, a longer calibration load would introduce sharp resonances in the frequency range of interest (<7 kHz); the source parameters would be sensitive to small errors in the theoretical length of the load near these resonant frequencies.

(b) Does the accuracy of admittance measurements depend on the number of calibration loads and/or the manner in which loads are coupled to the source? Neely and Gorga (1998, Fig. 3) recently reported accurate measurements made with an ER-10C probe—*without* a microphone-tube extension—for frequencies above 8 kHz. While this result is inconsistent with our finding that an extension is crucial for accuracy, there are important differences in their methods. They used a five-load calibration scheme (optimized by adjusting the theoretical load lengths) and a coupler that fixed the geometry of the source-load junction. Perhaps this approach incorporates the effects of nonuniform waves into the source parameters, for their test loads. Resolution of this issue hinges on understanding the implications of adjusting load lengths in the calibration procedure. Others with similar calibration schemes have used an extension (Keefe *et al.*, 1992; Voss and Allen, 1994). Thus it is difficult to reach a firm conclusion at this time.

B. Use of method in ears

1. Assessment of results

Measurements in domestic cats show that the method gives fairly repeatable estimates of Y_{TM} for frequencies up to 5 kHz (Fig. 14, lower). In this frequency range, the Y_{TM}(est.) data are quantitatively similar to the measurements of Margolis *et al.* (1978) and Lynch *et al.* (1994) [see Fig. 12(B)]. In particular, (1) the mean $|Y_{TM}(\text{est.})|$ of four ears is not statistically different from the mean of the pooled previous data at 0.22 kHz or at 0.66 kHz; (2) the cavity notches in Y_{TM}(est.) between 3 and 4 kHz are similar in depth, bandwidth, and center frequency to previous measurements in cats (see also Møller, 1965; Guinan and Peake, 1967; Peake *et al.*, 1992). Thus our noninvasive estimates of Y_{TM} have features that are ''catlike.'' Normal inter-ear variability could explain some differences between our results and Lynch's data, at least for frequencies below 5 kHz.

For frequencies above 5 kHz, Y_{TM}(est.) curves based on repeated eartip insertions have greater differences (Fig. 14, lower). In this frequency range, there are also greater differences between Y_{TM}(est.) and Lynch's data, particularly in the angles [Fig. 12(B)]. Although we cannot rule out changes in the ear with time ($<$60 min between insertions), the results are consistent with the idea that the total error in the method increases for higher frequencies.

The acoustically estimated canal dimensions for the admittance-location transformations (see Fig. 12 caption) are roughly consistent with structural measurements of cats' ear canals (Rosowski *et al.*, 1988).[16] That is, the canal-space area and volume estimates obtained from the measurement procedure are anatomically reasonable.

2. Sources of error

Above 5 kHz, there are three main sources of error. First, our admittance measurements in test loads have increasing errors above 6 kHz that are probably due to nonuniform waves. Second, the ear canal transformation becomes more sensitive to small errors in its parameters above 2 kHz (although the frequency of the cat middle-ear-admittance notch is relatively unaffected by such errors) (Fig. 13). Third, the irregular geometry of the canal might introduce greater errors for higher frequencies, for two reasons. (a) Consider the shape of the canal space at the location of the microphone-tube tip. This geometry (and its effect on the eartip guide—see Fig. 10) is part of the source, but the source is calibrated in smooth cylindrical loads. This structural discrepancy might be significant at higher frequencies. (b) The comparison of a tube model to a canal impedance measurement in Rosowski *et al.* (1988, Fig. 11) suggests that errors resulting from the approximation of the canal as a uniform tube increase above 6 kHz.

ACKNOWLEDGMENTS

We thank Denny Freeman and the late Tom McMahon for helpful discussions and comments on the manuscript; Chris Shera and Susan Voss for comments on the manuscript; Leslie Liberman for assistance with animal preparations; the staff of the Eaton-Peabody Laboratory; Steve Iceberg and Etymotic Research for the custom ER-10C probe; Andy Brughera and the Sensory Communication group at Research Laboratory of Electronics for the ear simulator; and Jont Allen, Doug Keefe, and Bill Rabinowitz for helpful discussions. Portions of this work were presented at the 21st Midwinter Meeting of the Association for Research in Otolaryngology (Huang *et al.*, 1998) and in a doctoral thesis (Huang, 1999). This work was supported by NSF and NIDCD.

[1]The phrase ''at the TM'' can be ambiguous, because the TM terminates the ear canal obliquely (e.g., Stinson and Khanna, 1994). However, for frequencies below 6 kHz in humans (10 kHz in cat), the sound pressure is essentially constant over the surface of the TM (Stinson, 1985; Chan and Geisler, 1990, p. 1237; Stinson and Khanna, 1994). For these frequencies, Y_{TM} represents a driving-point admittance at a point approximation of the TM.

[2]The volume of the residual canal was estimated by a different approach. Rabinowitz (1981) estimated a volume from admittance measurements made with static pressures in the canal, and then the volume was adjusted such that (1) the TM compliance was consistent with the subjects' perceived loudness changes with static pressures, and (2) Y_{TM} was consistent with a representative ear canal sound-pressure gain function. In contrast, Møller (1960) and Zwislocki (1970) estimated the volume by filling the canal with alcohol from a calibrated syringe. In all three studies, a value between 7.5 and 8.0 mm was assumed for the diameter of the human ear canal.

[3]Electric stimuli are generated and responses measured with an Ariel DSP-16+ board and SYSID (www.sysid-labs.com, Berkeley, CA) software (Puria and Allen, 1992).

[4]Another approach is to make calibration measurements in four or more loads to create an overdetermined system of equations for the two source parameters (Allen, 1986; Keefe *et al.*, 1992; Neely and Gorga, 1998; Puria and Allen, 1998). The theoretical load lengths and source parameters are chosen iteratively to give the best match between measured and theoretical admittances. The differences in accuracy between this scheme and ours are not clear (see Sec. III A 4).

[5]From a theoretical analysis, Brass and Locke (1997, p. 2164) conclude that ''The maximum measurement error [in sound-pressure magnitude] due to the presence of the evanescent wave in human ear canals below 10 kHz is predicted to be 3 dB in adults...when the loudspeaker and microphone ports are spaced more than 2 mm apart.''

[6]Repeated admittance measurements of these test loads (e.g., on different days) were quite repeatable with variations of less than 10% in magnitude and 0.03 periods in angle.

[7]This damper model is similar to that of Voss and Allen (1994, pp. 375–

376), except that we did not adjust the parameter values to improve the fit between measurement and theory.

[8]Measurements made in this particular simulator differ from the original specifications (Zwislocki, 1970) and from measurements made in the same product by Voss and Allen (1994). It is possible that this simulator was defective (e.g., through dust buildup). A defect would not be crucial here, as the load needs only to have a more earlike admittance than a closed tube of the same equivalent volume (see Fig. 6), a criterion which it does satisfy.

[9]Results were obtained with this transformation in Fig. 3, but the focus there was the effect of the microphone-tube extension.

[10]For the frequency range (0.1–10 kHz) and ear canal diameters of interest (5–12 mm), the inclusion of viscous and thermal losses in the transmission-line model yields no significant differences from the lossless model.

[11]For a description of measurements in these ears, see Huang (1999, Table 2.1, and Figs. 2.12 and 2.13).

[12]We also found that the size of these admittance errors increases with increasing load diameter, presumably because the sound-pressure signal level in a larger load is reduced whereas the artifact level is unchanged.

[13]To construct each guide, a thick coating of RTV silicone sealant is applied around the eartip near its distal end and shaped either (1) with a fine hand tool or (2) by inserting the eartip into an otoscope speculum for a smooth taper. The guide material cures in about a day.

[14]In some ears, the sealant went past the lateral end of the guide, between the guide and the ear canal wall. In some other ears, the sealant did not reach the guide.

[15]Use of this reduced bandwidth requires that the real part of the load impedance be roughly constant with frequency. This was generally the case in our ear canal measurements.

[16]Ear canal molds described by Rosowski *et al.* (1988, Table I, p. 1699) had minimum cross-sectional area at a point about midway between the canal-concha border and the TM. A tube model of the canal with this small radius (2.5 mm) represented radiation-impedance measurements of the external ear fairly well. Here, in contrast, larger radii were inferred from acoustic measurements at lateral ear-canal locations [e.g., 3.5 mm and 3.8 mm in Fig. 14(A)], which are consistent with reported dimensions of the canal-concha interface.

Allen, J. B. (**1986**). ''Measurement of eardrum acoustic impedance,'' in *Peripheral Auditory Mechanisms*, edited by J. L. Hall, J. B. Allen, A. Hubbard, S. T. Neely, and A. Tubis (Springer-Verlag, New York), pp. 44–51.

Beranek, L. L. (**1986**). *Acoustics* (American Institute of Physics, New York).

Blauert, J. (**1983**). *Spatial Hearing* (Massachusetts Institute of Technology, Cambridge, MA).

Brass, D., and Locke, A. (**1997**). ''The effect of the evanescent wave upon acoustic measurements in the human ear canal,'' J. Acoust. Soc. Am. **101**, 2164–2175.

Burns, E. M., Keefe, D. H., and Ling, R. (**1998**). ''Energy reflectance in the ear canal can exceed unity near spontaneous otoacoustic emission frequencies,'' J. Acoust. Soc. Am. **103**, 462–474.

Chan, J. C. K., and Geisler, C. D. (**1990**). ''Estimation of eardrum acoustic pressure and of ear canal length from remote points in the canal,'' J. Acoust. Soc. Am. **87**, 1237–1247.

Chan, J. C. K., Musicant, A. D., and Hind, J. E. (**1993**). ''An insert earphone system for delivery of spectrally shaped signals for physiological studies,'' J. Acoust. Soc. Am. **93**, 1496–1501.

Egolf, D. P. (**1977**). ''Mathematical modeling of a probe-tube microphone,'' J. Acoust. Soc. Am. **61**, 200–205.

Feeney, M. P., and Keefe, D. H. (**1999**). ''Acoustic reflex detection using wide-band acoustic reflectance, admittance, and power measurements,'' J. Speech Lang. Hear. Res. **42**, 1029–1041.

Flanders, P. B. (**1932**). ''A method of measuring acoustic impedance,'' Bell Syst. Tech. J. **11**, 402–410.

Guinan, J. J. (**1990**). ''Changes in stimulus frequency otoacoustic emissions produced by two-tone suppression and efferent stimulation in cats,'' in *Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, and C. R. Steele (Springer-Verlag, New York), pp. 170–177.

Guinan, Jr., J. J., and Peake, W. T. (**1967**). ''Middle-ear characteristics of anesthetized cats,'' J. Acoust. Soc. Am. **41**, 1237–1261.

Huang, G. T. (**1999**). ''Measurement of middle-ear acoustic function in intact ears: Application to size variations in the cat family,'' Ph.D. thesis, Massachusetts Institute of Technology.

Huang, G. T., Rosowski, J. J., Flandermeyer, D. T., Lynch III, T. J., and Peake, W. T. (**1997**). ''The middle ear of a lion: Comparison of structure and function to domestic cat,'' J. Acoust. Soc. Am. **101**, 1532–1549.

Huang, G. T., Rosowski, J. J., and Peake, W. T. (**2000a**). ''Relating middle-ear acoustic performance to body size in the cat family: Measurements and models,'' J. Comp. Physiol. A . **186**, 447–465.

Huang, G. T., Rosowski, J. J., Puria, S., and Peake, W. T. (**1998**). ''Noninvasive technique for estimating the acoustic impedence at the tympanic membrane in ear canals of different size,'' in Abstracts of the 21st Midwinter Meeting of the Association for Research in Otolaryngology, p. 111.

Huang, G. T., Rosowski, J. J., Puria, S., and Peake, W. T. (**2000b**). ''Tests of some common assumptions of ear-canal acoustics in cats,'' J. Acoust. Soc. Am. **108**, 1147–1161.

Ingard, K. U. (**1948**). ''On the radiation of sound into a circular tube, with an application to resonators,'' J. Acoust. Soc. Am. **20**, 665–682.

Jerger, J. F., Ed. (**1975**). *Handbook of Clinical Impedance Audiometry* (American Electromedics, New York).

Karal, F. C. (**1953**). ''The analogous acoustical impedance for discontinuities and constrictions of circular cross section,'' J. Acoust. Soc. Am. **25**, 327–334.

Keefe, D. (**1984**). ''Acoustical wave propagation in cylindrical ducts: Transmission line parameter approximations for isothermal and nonisothermal boundary conditions,'' J. Acoust. Soc. Am. **75**, 58–62.

Keefe, D. H., and Benade, A. H. (**1981**). ''Impedance measurement source and microphone proximity effects,'' J. Acoust. Soc. Am. **69**, 1489–1495.

Keefe, D. H., Bulen, J. C., Arehart, K. H., and Burns, E. M. (**1993**). ''Ear-canal impedance and reflection coefficient in human infants and adults,'' J. Acoust. Soc. Am. **94**, 2617–2638.

Keefe, D. H., and Levi, E. (**1996**). ''Maturation of the middle and external ears: Acoustic power-based responses and reflectance tympanometry,'' Ear Hear. **17**, 361–373.

Keefe, D. H., Ling, R., and Bulen, J. C. (**1992**). ''Method to measure acoustic impedance and reflection coefficient,'' J. Acoust. Soc. Am. **91**, 470–485.

Lynch III, T. J. (**1981**). ''Signal processing by the cat middle ear: Admittance and transmission, measurements and models,'' Ph.D. thesis, Massachusetts Institute of Technology.

Lynch III, T. J., Peake, W. T., and Rosowski, J. J. (**1994**). ''Measurements of the acoustic input-impedance of cat ears: 10 Hz to 20 kHz,'' J. Acoust. Soc. Am. **96**, 2184–2209.

Margolis, R. H., and Shanks, J. (**1985**). ''Tympanometry,'' in *Handbook of Clinical Audiology*, edited by J. Katz (Williams & Wilkens, Baltimore, MD), 438–475.

Margolis, R. H., Saly, G. L., and Keefe, D. H. (**1999**). ''Wideband reflectance tympanometry in normal adults,'' J. Acoust. Soc. Am. **106**, 265–280.

Margolis, R. H., Osguthorpe, J. D., and Popelka, G. R. (**1978**). ''The effects of experimentally produced middle ear lesions on tympanometry in cats,'' Acta Oto-Laryngol. **86**, 428–436.

Miles, J. A. (**1946**). ''The analysis of plane discontinuities in cylindrical tubes. Part I,'' J. Acoust. Soc. Am. **17**, 259–271.

Møller, A. R. (**1960**). ''Improved technique for detailed measurements of middle ear impedance,'' J. Acoust. Soc. Am. **32**, 250–257.

Møller, A. R. (**1965**). ''An experimental study of the acoustic impedance of the middle ear and its transmission properties,'' Acta Oto-Laryngol. **60**, 129–149.

Morse, P. M., and Ingard, K. U. (**1968**). *Theoretical Acoustics* (McGraw-Hill, New York), pp. 343–346.

Mundie, J. R. (**1963**). ''The impedance of the ear—A variable quantity,'' in U. S. Army Medical Research Lab. Report 576, edited by J. L. Fletcher, pp. 65–85.

Neely, S. T., and Gorga, M. P. (**1998**). ''Comparison between intensity and pressure as measures of sound level in the ear canal,'' J. Acoust. Soc. Am. **104**, 2925–2934.

Onchi, Y. (**1961**). ''Mechanism of the middle ear,'' J. Acoust. Soc. Am. **33**, 794–805.

Peake, W. T., Rosowski, J. J., and Lynch III, T. J. (**1992**). ''Middle-ear transmission: Acoustic vs. ossicular coupling in cat and human,'' Hear. Res. **57**, 245–268.

Puria, S. (**1991**). ''A theory of cochlear input impedance and middle ear parameter estimation,'' Ph.D. thesis, City University of New York.

Puria, S., and Allen, J. B. (**1992**). ''SYSid—Audio-band measurement and analysis system,'' J. Acoust. Soc. Am. **92**, 2469.
Puria, S., and Allen, J. B. (**1998**). ''Measurements and model of the cat middle ear: Evidence of tympanic membrane acoustic delay,'' J. Acoust. Soc. Am. **104**, 3463–3481.
Rabbitt, R. D. (**1988**). ''High-frequency plane waves in the ear canal: Application of a simple asymptotic theory,'' J. Acoust. Soc. Am. **84**, 2070–2080.
Rabbitt, R. D., and Friedrich, M. T. (**1991**). ''Ear canal cross-sectional pressure distributions: Mathematical analysis and computation,'' J. Acoust. Soc. Am. **89**, 2379–2390.
Rabinowitz, W. (**1981**). ''Measurement of the acoustic input immittance of the human ear,'' J. Acoust. Soc. Am. **70**, 1025–1035.
Ravicz, M. E., Rosowski, J. J., and Voigt, H. F. (**1992**). ''Sound power collection by the auditory periphery of the Mongolian gerbil *Meriones unguiculatus*: I. Middle-ear input impedance,'' J. Acoust. Soc. Am. **92**, 157–177.
Rosowski, J. J. (**1994**). ''Outer and middle ears,'' in *Springer Handbook of Auditory Research, Volume 4: Comparative Hearing: Mammals*, edited by R. R. Fay and A. N. Popper (Springer-Verlag, New York), pp. 172–247.
Rosowski, J. J., Carney, L. H., Lynch III, T. J., and Peake, W. T. (**1986**). ''The effectiveness of the external and middle ears in coupling acoustic power into the cochlea,'' in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer-Verlag, New York), pp. 3–12.
Rosowski, J. J., Carney, L. H., and Peake, W. T. (**1988**). ''The radiation impedance of the external ear of cat: Measurements and applications,'' J. Acoust. Soc. Am. **84**, 1695–1708.
Shanks, J. E., and Lilly, D. J. (**1981**). ''An evaluation of tympanometric estimates of ear canal volume,'' J. Speech Hear. Res. **24**, 557–566.
Shaw, E. A. G. (**1974**). ''The external ear,'' in *Handbook of Sensory Physiology: Volume V/1: Auditory System*, edited by W. D. Keidel and W. D. Neff (Springer-Verlag, New York), pp. 455–490.
Stevens, K. N., Berkovitz, R., Kidd, Jr., G., and Green, D. M. (**1987**). ''Calibration of ear canals for audiometry at high frequencies,'' J. Acoust. Soc. Am. **81**, 470–484.
Stinson, M. R. (**1985**). ''The spatial distribution of sound pressure within scaled replicas of the human ear canal,'' J. Acoust. Soc. Am. **78**, 1596–1602.
Stinson, M. R. (**1990**). ''Revision of estimates of acoustic energy reflectance at the human eardrum,'' J. Acoust. Soc. Am. **88**, 1773–1778.
Stinson, M. R., and Khanna, S. M. (**1994**). ''Spatial distribution of sound pressure and energy flow in the ear canals of cats,'' J. Acoust. Soc. Am. **96**, 170–180.
Stinson, M. R., Shaw, E. A. G., and Lawton, B. W. (**1982**). ''Estimation of acoustical energy reflectance at the eardrum from measurements of pressure distribution in the human ear canal,'' J. Acoust. Soc. Am. **72**, 766–773.
Voss, S. E., and Allen, J. B. (**1994**). ''Measurement of acoustic impedance and reflectance in the human ear canal,'' J. Acoust. Soc. Am. **95**, 372–384.
Wada, H., Kobayashi, T., Suetake, M., and Tachizaki, H. (**1989**). ''Dynamic behavior of the middle ear based on sweep frequency tympanometry,'' Audiology **28**, 127–134.
White, R. E. C., Studebaker, G. A., Levitt, H., and Mook, D. (**1980**). ''The application of modeling techniques to the study of hearing aid acoustic systems,'' in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. E. Studebaker and I. Hochberg (University Park Press, Baltimore, MD), pp. 267–296.
Wilber, L. A., and Feldman, A. S. (**1976**). ''The middle ear measurement battery,'' in *Acoustic Impedance and Admittance: The Measurement of Middle Ear Function*, edited by A. S. Feldman and L. A. Wilber (Williams and Wilkins, Baltimore, MD), pp. 345–377.
Zuercher, J. C., Carlson, E. V., and Killion, M. C. (**1988**). ''Small acoustic tubes: New approximations including isothermal and viscous effects,'' J. Acoust. Soc. Am. **83**, 1653–1660.
Zwislocki, J. (**1962**). ''Analysis of the middle-ear function. Part I: Input impedance,'' J. Acoust. Soc. Am. **34**, 1514–1523.
Zwislocki, J. J. (**1970**). ''An acoustic coupler for earphone calibration,'' Laboratory of Sensory Communication Special Report LSC-S-7, Syracuse University.

Tests of some common assumptions of ear-canal acoustics in cats

Gregory T. Huang,[a] John J. Rosowski, Sunil Puria,[b] and William T. Peake
Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114 and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 5 August 1999; revised 27 April 2000; accepted 18 May 2000)

The accuracy of ear-canal admittance and reflectance as measures of the ear's properties depends on the acoustic effects of the canal. Here, measurements of acoustic admittance at different canal locations in domestic cats are used to test three common assumptions. (1) Can a uniform-tube model of the canal represent spatial variations in admittance? Data from cats support this assumption for frequencies below 3 kHz, where the admittance inferred at the tympanic membrane (TM) based on a uniform-tube model differs by less than 3 dB in magnitude and 0.07 periods in angle from the admittance measured at the TM; for higher frequencies greater differences occur. (2) Do large static air pressures in the canal make the middle ear rigid without affecting the properties of the canal space? The measurements reported indicate that large negative static pressures reduce the low-frequency compliance of the cat middle ear to about 10% of the compliance of the canal air volume. Static displacements of the acoustic probe, TM, and canal walls with static pressure may affect estimates of the canal volume and middle-ear compliance by as much as 15% to 20%. (3) Is the acoustic-reflectance magnitude constant with position along the canal? Reflectance data from cat ear canals generally support this idea, except within a frequency region near 0.5 kHz for which there is evidence of energy loss. These results demonstrate that noninvasive measurements in the canal describe middle-ear acoustic properties to within tolerances that depend on the effects of the canal. [S0001-4966(00)00509-9]

PACS numbers: 43.64.Yp, 43.64.Ha, 43.64.Tk, 43.58.Bh [BLM]

INTRODUCTION

A. Motivation

Measurements of acoustic admittance and reflectance in the ear canal are useful measures of the mechanics of the ear (particularly the middle ear) for scientific and clinical purposes. The input admittance at the tympanic membrane (TM), Y_{TM}, describes the acoustic response of the middle ear (Onchi, 1961; Zwislocki, 1962; Mundie, 1963; Møller, 1965; Rabinowitz, 1981; Hudde, 1983; Lynch *et al.*, 1994; Huang *et al.*, 2000) and affects the acoustics of the external ear (Rosowski *et al.*, 1986; Stinson and Khanna, 1994).[1] The acoustic reflectance (or reflection coefficient) in the canal, R, relates the waves propagating into and out of the ear (Stinson *et al.*, 1982; Keefe *et al.*, 1993; Voss and Allen, 1994; Keefe and Levi, 1996; Burns *et al.*, 1998; Puria and Allen, 1998; Margolis *et al.*, 1999).[2] Recently there has been interest in determining the effects of the middle ear on otoacoustic emissions, in part to assess cochlear function more accurately (Shera and Zweig, 1992; Keefe, 1997; Puria and Rosowski, 1997; Avan *et al.*, 2000). The accuracy of all of these noninvasive measures depends on fundamental assumptions about the ear canal. In this article, we report tests of three assumptions that are prevalent in the literature, as motivated below.

(1) How are spatial variations of admittance along the ear canal represented? It is often convenient to approximate the canal as a smooth, rigid tube with uniform cross-sectional area. A common approach for estimating Y_{TM} is to measure the admittance at a lateral position in the canal and to apply a uniform-tube transformation to account for the space between the measurement point and the TM (Møller, 1965; Zwislocki, 1970; Rabinowitz, 1981; Lynch *et al.*, 1994; Margolis *et al.*, 1999; Huang *et al.*, 2000a; 2000b).[3] In reality, the canals of both humans and animals have changes in area and axial direction (Shaw, 1974; Khanna and Stinson, 1985; Stinson, 1985; Rosowski *et al.*, 1988). The errors in admittance that result from the uniform-tube approximation are generally assumed to be small for frequencies below 4 kHz and distances of roughly 10 mm along the canal (Rabinowitz, 1981, p. 1033; Huang *et al.*, 1997, p. 1536; Margolis *et al.*, 1999). However, direct tests of the accuracy of this approximation have not been reported.

(2) How do the properties of the ear canal and middle ear change in response to static air pressures? Measurements of admittance with controlled static pressures in the canal (''tympanometry'') are used clinically to assess middle-ear function (Jerger, 1975; Wilber and Feldman, 1976; Margolis and Shanks, 1985). In this approach, the acoustic effects of the canal are separated from those of the middle ear using admittances measured with large static pressures (Rabinowitz, 1981; Margolis *et al.*, 1999; Huang *et al.*, 2000b, Sec. II B 2 b). Common

[a]Electronic mail: gthuang@MIT.EDU
[b]Current address: Stanford University, 801 Welch Road, Palo Alto, CA 94304.

assumptions for low frequencies are that (a) large static pressures render the middle ear effectively rigid (as seen from the canal), and that (b) static pressures have no effect on the properties of the canal space. Previous data suggest that these assumptions may not be accurate. Contradictory to (a), Rabinowitz (1981) concluded, based on a combination of acoustic and perceptual data, that large static pressures do not make the middle ear functionally rigid. In disagreement with (b), Keefe *et al.* (1993, pp. 2634–2635) concluded from ear-canal acoustic measurements and a lumped-circuit model that canal-wall motion can be significant, especially in infants, and that ''nonzero static pressure modifies the elasticity of the compliant walls.'' Here, we test the effects of static pressure on the canal and middle ear with measurements in cats.

(3) How does reflectance vary with position along the ear canal? If the canal is approximated locally (near the measurement location) as a uniform tube, then the reflectance R at that point is related approximately to the (''load'') admittance Y_L at that point by

$$R=\frac{Y_0-Y_L}{Y_0+Y_L}, \tag{1}$$

where Y_0 is the characteristic admittance of the canal.[4] The quantity $|R|^2$—the squared magnitude of the pressure reflectance (''power reflectance'')—is the fraction of the incident sound power that is reflected. It is generally assumed that $|R|^2$ is constant with position along the canal (Stinson *et al.*, 1982, p. 766; Hudde, 1983, p. 246; Keefe *et al.*, 1993, p. 2617; Voss and Allen, 1994, pp. 372–373), which would be correct if the canal was a lossless tube. If this assumption is valid, one can estimate power flow to the ear without knowledge of the exact location of the acoustic probe.[5] However, measurements that test the idea of spatial independence of power reflectance have not been reported previously.

The goal of this article is to test the three assumptions discussed above: (1) that a uniform-tube approximation of the canal represents spatial variations in acoustic admittance for frequencies up to about 4 kHz; (2) that large static pressures in the canal make the middle ear effectively rigid (for low frequencies) without affecting the canal; and (3) that acoustic power reflectance is constant with position along the canal. We determine the accuracy of each assumption based on acoustic measurements in domestic cats. These tests are a step toward a more complete understanding of the effects of the ear canal on noninvasive measurements of the acoustic behavior of ears.

B. Approach and justification

We have reported a noninvasive method to estimate the middle-ear input admittance (Y_{TM}) based on admittance measurements in the lateral portion of the ear canal (Y_{EC}) (Huang, 1999; Huang *et al.*, 2000b). The method makes use of the assumptions stated above. In this article, we report measurements on cats that test the accuracy of these assumptions and of the method. The experimental approach is to compare acoustic measurements made at two locations in a given ear canal—in the lateral portion of the intact canal, and in the ''residual'' canal (within 3–5 mm of the TM) after removal of most of the cartilaginous canal (Fig. 1).

Tests of the ear-canal assumptions are conducted as follows. (1) To test the uniform-tube representation of spatial variations in admittance, a noninvasive estimate of Y_{TM} based on a uniform-tube transformation of Y_{EC} is compared to a more direct measurement of Y_{TM}. This comparison allows us to determine the accuracy of the uniform-tube approximation for specific frequency regions. (2) To assess the effects of applied static pressures on the canal and middle ear, the low-frequency compliances measured in the lateral portion of the canal are compared to those measured near the TM, as functions of static-pressure level. This approach allows us to estimate variations in the middle-ear compliance and in the volume of the space between the two measurement points, with changes in static pressure. (3) To test the spatial independence of reflectance magnitude, we compare reflectances computed from the admittances at the two canal locations [from Eq. (1)].

A potential complication in the measurements near the TM [Fig. 1(B)] is the effect of nonuniform waves generated by the interaction of incident sound with the complex geometry and mechanics of the TM (Khanna and Stinson, 1985; Rabbitt and Holmes, 1988). Calculations of admittance assume that the measured sound pressure consists of only the uniform plane-wave mode. Two lines of evidence support the idea that for frequencies below 8 kHz, nonuniform waves generated at the TM have a small effect on our pressure measurements made 3–5 mm away from the TM: (1) Khanna and Stinson (1985, Figs. 9 and 13) report that the sound pressure measured near (and roughly parallel to) the surface of the TM in cat ''remains essentially unchanged between curves up to 8 kHz as the probe tube is pulled back by 6 mm.'' (2) Theory of nonuniform waves in a cylindrical tube predicts that, based on the approximate dimensions of the residual-canal space (length 3 mm, radius 3 mm), the amplitude of the strongest evanescent (nonpropagating) mode sensed by our probe is less than 15% of its amplitude at the TM at 8 kHz and even smaller at lower frequencies (Morse and Ingard, 1968; Huang *et al.*, 2000b, Sec. I B 1 b). The analysis of Rabbitt and Holmes (1988, Fig. 12), which is based on a more detailed model of the TM and canal, is in agreement with this prediction. The conclusion is that below 8 kHz a uniform plane-wave propagation model is valid in our measurement region near the cat TM. We restrict our tests of ear-canal assumptions to frequencies below 8 kHz.[6]

I. METHODS

A. Scheme of experiments

Treatment of experimental animals conformed to the guidelines of the National Institutes of Health, and experimental protocols were approved by the Animal Care committees at the Massachusetts Institute of Technology and at the Massachusetts Eye and Ear Infirmary.

Measurements were made on Dial-anesthetized domestic cats with vented middle-ear cavities (Huang *et al.*, 2000b, Secs. II A 1–2). Acoustic admittances were measured at two

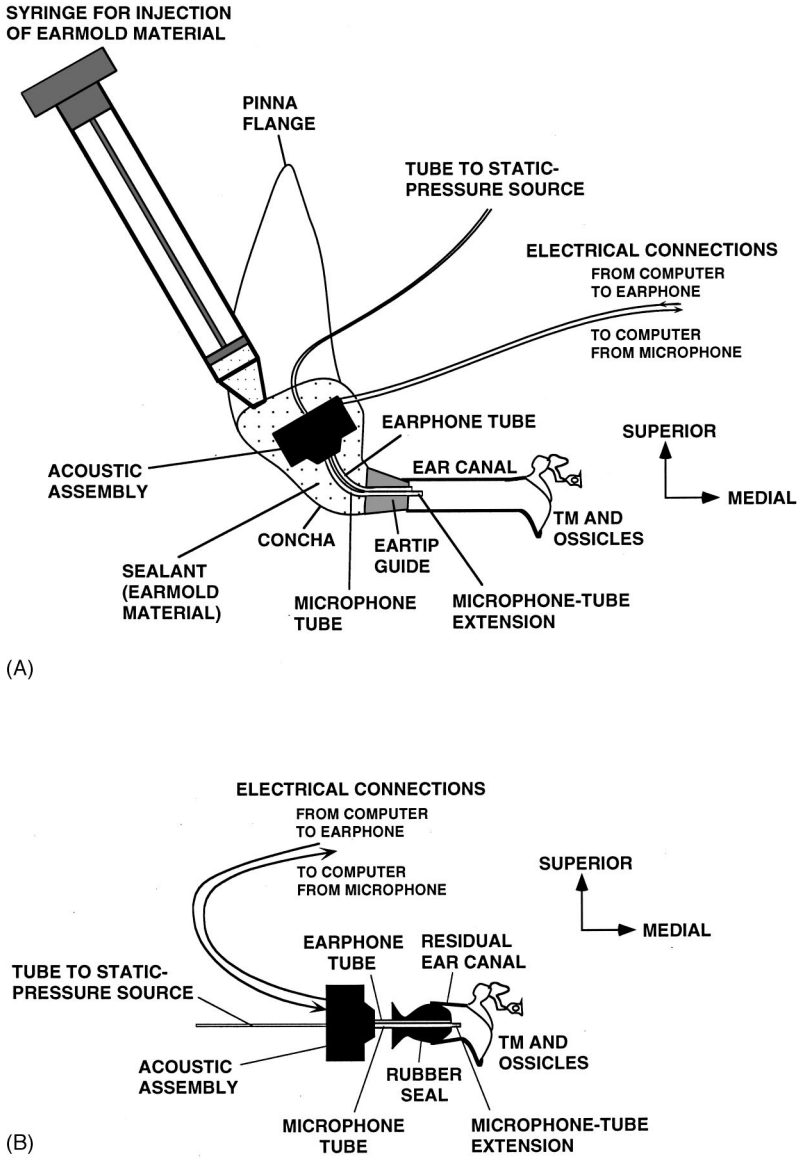


FIG. 1. Methods for measurement of acoustic admittance (and reflectance) in two configurations of a cat ear. The schematic diagrams do not accurately portray the geometry of the ear structures. The middle-ear cavities (not shown) were intact and vented. (A) Methods in the intact ear canal. A custom eartip connected to the acoustic assembly (Etymotic Research ER-10C) was inserted into the canal just medial to the sharp bend at the canal-concha border. Hearing-aid earmold-impression material (Westone Laboratories) was injected behind the eartip guide via a syringe, to form a static-pressure seal in the canal and to hold the eartip in place. The admittance in the canal, Y_{EC}, was measured with controlled static pressures in the canal. The dimensions of the canal space between the measurement point and the tympanic membrane (TM) were estimated from Y_{EC}. The middle-ear input admittance Y_{TM} was then inferred from these measurements. (B) Methods at the TM. After resection of most of the cartilaginous ear canal, a standard Etymotic eartip with a rubber seal was inserted into the residual (shortened) canal to within a few millimeters of the TM. Admittance measurements were made with controlled static pressures (as above), in part to obtain an independent estimate of Y_{TM}. For each canal configuration, reflectance was computed from the admittance and an acoustic estimate of the canal area at the measurement point [Eq. (1)].

locations (separated by 5–10 mm) in each ear canal: (1) in the lateral portion of the intact canal [Fig. 1(A)], and (2) near the TM after surgical removal of most of the cartilaginous canal [Fig. 1(B)]. In each configuration, measurements were made with controlled static pressures in the canal. From the admittances measured in the intact canal (Y_{EC}), we estimated the admittance at the TM (Y_{TM}) based on the uniform-tube model; we refer to this noninvasive estimate as ''Y_{TM} (est.)'' (Huang *et al.*, 2000b, Sec. II B). The admittances measured near the TM were used to obtain an independent estimate of Y_{TM}; this quantity, our most direct measurement of Y_{TM}, is referred to as ''Y_{TM}(meas.).''

B. Acoustic-measurement procedure

1. Hardware

Acoustic-admittance measurements were made with an Etymotic Research (Elk Grove Village, IL) ER-10C probe system, modified to include (1) a tube for introduction of static pressures into the ear canal and (2) a microphone-tube extension to reduce the effects of nonuniform waves generated by the sound source [Fig. 1(A); see also Huang *et al.*, 2000b, Sec. I].

2. Measurements in intact ear canal

The procedure for measuring admittance (Y_{EC}) in the intact ear canal [Fig. 1(A)] consisted of sealing the acoustic system into the lateral portion of the canal and measuring admittances with varying static-pressure levels (Huang *et al.*, 2000b, Sec. II A 4). Hearing-aid earmold impression material (Westone Laboratories, Colorado Springs, CO) was injected to achieve a stable seal. To infer the middle-ear input admittance [Y_{TM}(est.)] from Y_{EC}, a uniform-tube transformation was used to account for the canal space between the measurement point and the TM (Møller, 1965; Zwislocki, 1970; Rabinowitz, 1981; Huang *et al.*, 2000b, Sec. I C); the tube dimensions were determined acoustically (Keefe *et al.*, 1992; Huang *et al.*, 2000b, Sec. II B).

The low-frequency compliance at the measurement point was computed as the average of $\mathrm{Im}\{Y_{EC}\}/(2\pi f)$ over

ten frequency points f between 0.1–0.3 kHz, where Im{} denotes the imaginary part. The reflectance at the measurement point, R_{EC}, was computed from Eq. (1) with $Y_L = Y_{EC}$ and with Y_0 based on the acoustically estimated cross-sectional area of the lateral portion of the canal.

3. Measurements near the TM

The goal of this portion of the experiment was to obtain measurements of admittance near the TM. After measurements were completed in the intact ear, the pinna flange, concha, and cartilaginous ear canal were resected at a point 5–8 mm lateral to the bony ear canal, to allow insertion of the acoustic system into the residual cartilaginous canal [Fig. 1(B)]. A standard Etymotic eartip was used to couple the acoustic assembly to the residual canal and to provide a static-pressure seal. No earmold material was used; occasionally, a small amount of petroleum jelly was spread around the eartip to improve the seal to the canal. The microphone-tube tip was extended 3–4 mm beyond the earphone-tube tip via a piece of steel tubing. Acoustic admittances were measured within a few millimeters of the TM with controlled static pressures in the residual canal.

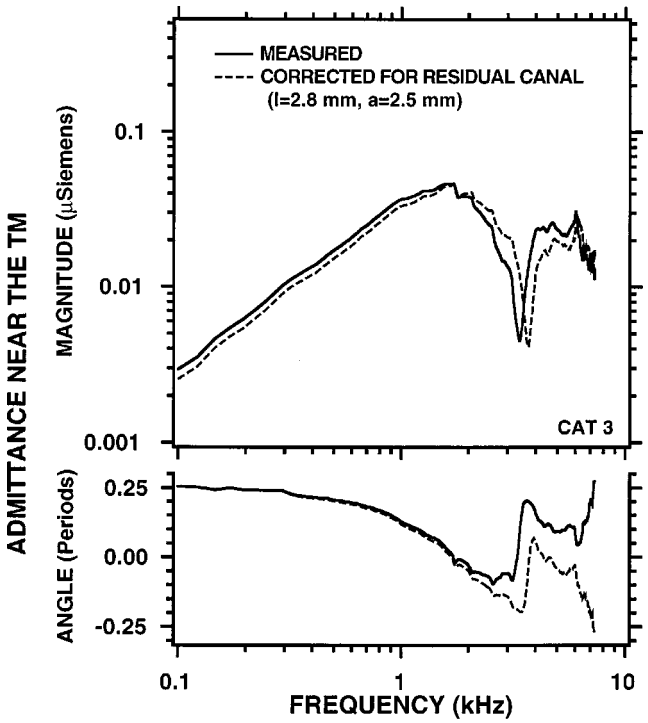
To obtain an independent estimate of middle-ear input admittance [Y_{TM}(meas.)], we accounted for the residual-canal space between the microphone-tube tip and the TM with a uniform-tube transformation (of the same form as above). An example of the effects of this ''correction'' is shown in Fig. 2 (upper). For frequencies below 1 kHz, the correction results in a decrease in admittance magnitude of approximately 10% and decreases in angle of less than 0.01 periods. These effects are consistent with the removal of the compliance of the residual-canal space (equivalent volume 0.06 cm^3). The correction also results in (1) a shift of the frequency of a prominent admittance notch from 3.4 to 3.7 kHz (an increase of 9%), and (2) a decrease in angle of 0.1 periods or greater for frequencies between 3 and 7 kHz.

The residual-canal dimensions (Fig. 2, lower) imply that in four ears the microphone-tube tip was 2–4 mm from the TM and the residual-canal volume was 0.05–0.1 cm^3. These acoustically estimated values are consistent with residual-canal dimensions reported by Lynch *et al.* (1994, p. 2193).[7] The estimated dimensions of the residual canal are smaller than those of the intact canal, and the residual-canal correction has a smaller effect than the intact-canal transformation (Huang *et al.*, 2000b, Fig. 12).

The low-frequency compliance near the TM was computed from the measured admittance as described above (Sec. I B 2). The reflectance near the TM, R_{TM}, was computed from Eq. (1) with $Y_L = Y_{TM}$(meas.) and with Y_0 based on the acoustically estimated cross-sectional area of the canal near the TM.

C. Data selection

Measurement series were accepted if they satisfied all of the following criteria: (1) The TM and middle-ear cavities were considered healthy based on visual monitoring throughout the experiment; (2) acoustic responses to variations in static pressure were repeatable with both configurations of the ear canal; and (3) the earmold-impression material used to seal the probe into the intact canal filled the space lateral to the probe [Fig. 1(A)]. Measurements made in eight ears failed to meet one or more of these criteria and were excluded from our full analyses (Huang, 1999, Table 2.1). Of these ears, five ears suffered from unstable preparations (perforated TM or blood in the middle-ear cavities), three ears showed unstable (not repeatable) changes in admittance with static pressure, and in three intact ears the seal did not completely fill the space lateral to the eartip. Our tests of ear-canal assumptions are based on measurement series in the four remaining ears. The implications of faulty seals in the intact canal are treated separately (Sec. II A 2).



CAT EAR	RESIDUAL CANAL CORRECTION
3	l = 2.8 mm, a = 2.5 mm, Vol. = 0.06 cm^3
7	l = 2.5 mm, a = 2.6 mm, Vol. = 0.05 cm^3
8L	l = 2.4 mm, a = 3.2 mm, Vol. = 0.08 cm^3
8R	l = 3.4 mm, a = 3.1 mm, Vol. = 0.11 cm^3

FIG. 2. Effect of the residual-canal correction on the admittance measured near the TM. Upper panel: Acoustic admittances from one ear (Cat 3). The solid curve was measured in the residual canal [Fig. 1(B)]. The dashed (''corrected'') curve was computed from the solid curve by employing a uniform-tube transformation (length $l = 2.8$ mm, radius $a = 2.5$ mm) to account for the small space between the measurement point and the TM. The dashed curve is taken to be the admittance ''measured at the TM,'' Y_{TM}(meas.). Admittance magnitude units: $1\ \mu S = 1/(\text{mks}M\Omega) = 10^{-6}$ $m^3/(Pa\ s)$. Phase angles are plotted in periods ($1\text{period} = 2\pi$ radians). Lower panel: Parameter values (l, a, and volume) of the residual-canal correction, for four ears used to test the accuracy of the noninvasive method. For each ear, a was estimated from the measured admittance in the residual canal [Keefe *et al.*, 1992; Huang *et al.*, 2000b, Eq. (6)], and l was inferred from a and a volume estimate based on the admittance measured with large negative static pressure in the residual canal (see Huang *et al.*, 2000b, Sec. II B 2).

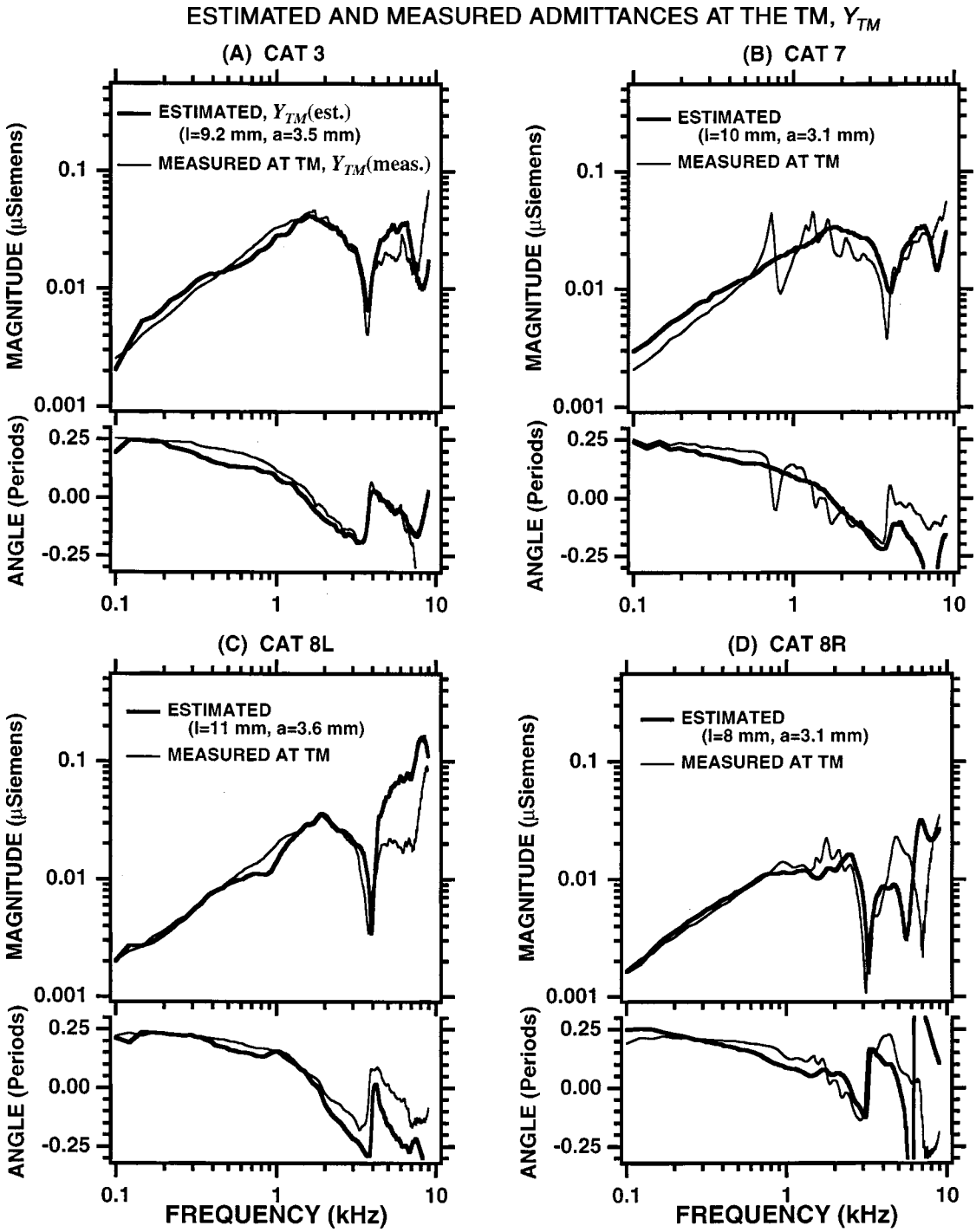


FIG. 3. Comparisons of Y_{TM}(est.) to Y_{TM}(meas.) in four cat ears. Y_{TM}(est.) was obtained from noninvasive measurements of Y_{EC} in intact ear canals [Fig. 1(A)], transformed to the location of the TM via a uniform-tube canal model (Huang *et al.*, 2000b, Fig. 12). Y_{TM}(meas.) was obtained with most of the ear canal removed and the microphone-tube tip near the TM [Fig. 1(B)]; for each ear, the measurement near the TM was corrected for the residual-canal space as described in Fig. 2.

II. RESULTS

A. Uniform-tube transformation of admittance

1. Illustrative cases

In this section, we compare estimates of Y_{TM} based on the Y_{EC} measurements and a uniform-tube canal model [Y_{TM}(est.)] to admittances measured at the TM [Y_{TM}(meas.)]. Results are reported from ears selected based on the criteria described in Sec. I C.

Figure 3 shows comparison of Y_{TM}(est.) and Y_{TM}(meas.) for four ears; in Fig. 4, ratios of Y_{TM}(est.) to Y_{TM}(meas.) are plotted for the same ears. In general, the estimated and measured admittances (Fig. 3) are similar in shape below 5 kHz, with some quantitative differences. In all ears, both sets of admittances are compliancelike for frequencies below 1 kHz and exhibit a notch in magnitude and a positive angle transition between 3–4 kHz (Fig. 3). This sharp feature is consistent with a ''cavity notch,'' an antiresonance that results from the two-compartment structure of

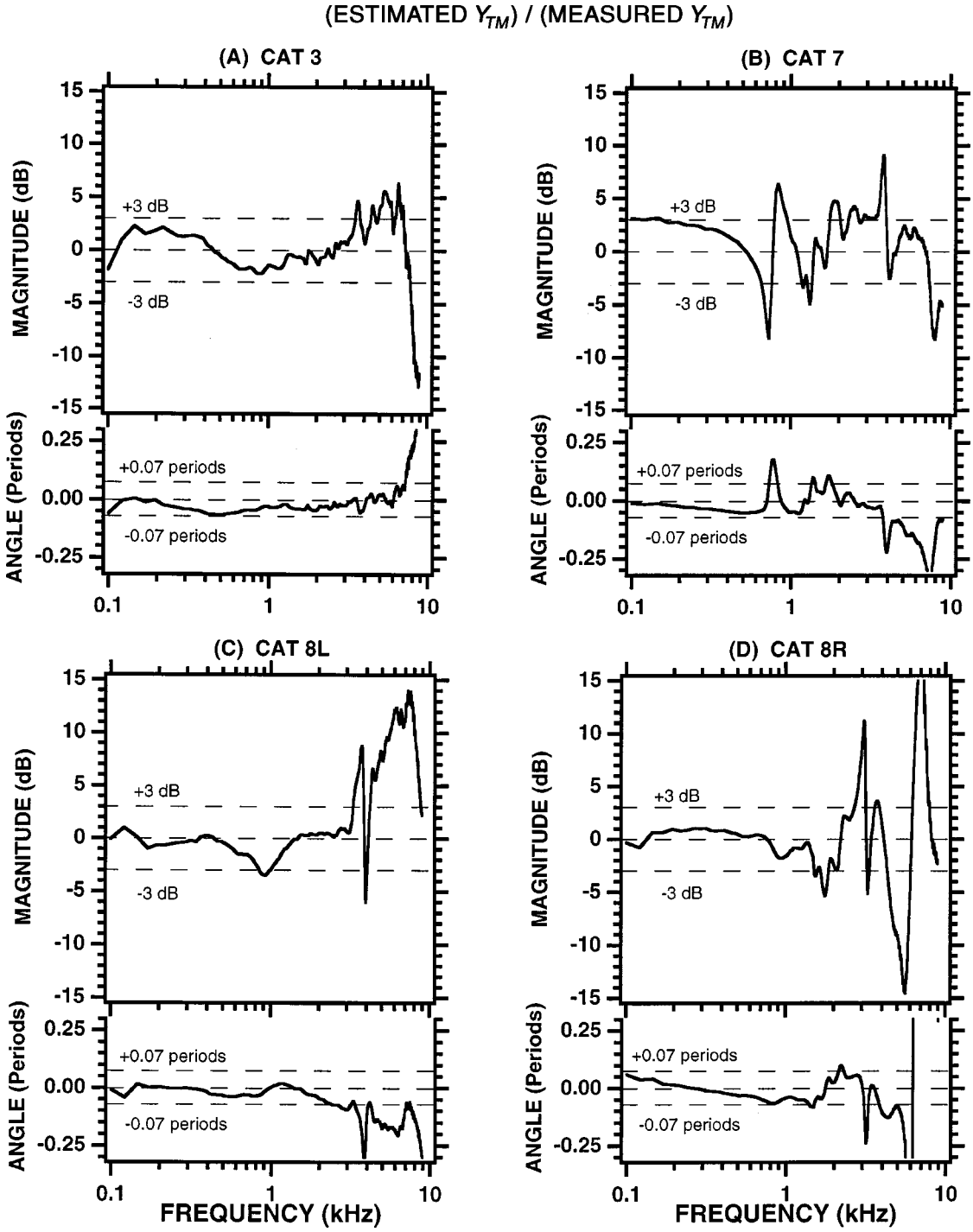


FIG. 4. Ratios Y_{TM}(est.)/Y_{TM}(meas.) for the four ears from Fig. 3. The dashed horizontal lines indicate 0 ± 3 dB (magnitude) and 0 ± 0.07 periods (angle).

the cat middle-ear cavity (Møller, 1965; Lynch *et al.*, 1994; Huang *et al.*, 1997). Below 3 kHz, the ratios are less than 5 dB in magnitude and 0.1 periods in angle (Fig. 4)—and to a large extent are less than 3 dB and 0.07 periods—except for cat 7 (see below). The magnitude ratios are 6 dB (factor of 2) or greater in the cavity-notch region for three ears (cats 7, 8L, and 8R), and $|Y_{TM}(\text{est.})|/|Y_{TM}(\text{meas.})|$ is greater than 2–2.5 dB (25%–35%) below 0.5 kHz for two ears (cats 3 and 7).

Above 5 kHz, the two admittances diverge and exhibit large variations with frequency (Fig. 3). At 8 kHz, all ears show ratios of at least 8 dB in magnitude and 0.15 periods in angle (Fig. 4). Our measurements suggest three main sources of error for these higher frequencies: (1) the measurements of Y_{EC} on which Y_{TM}(est.) is based have increased errors [Huang *et al.*, 2000b, Fig. 2(b)]; (2) Y_{TM}(est.) is sensitive to small errors in the ear-canal transformation parameters at higher frequencies (Huang *et al.*, 2000b, Fig. 13); and (3) at frequencies above 8 kHz, nonuniform sound waves generated at the TM may affect the pressure measured near the TM and thereby introduce errors in Y_{TM}(meas.) (see Introduction).

The Y_{TM}(meas.) curve of cat 7 [Fig. 3(B)] has some sharp features between 0.7 and 2 kHz that could be the result

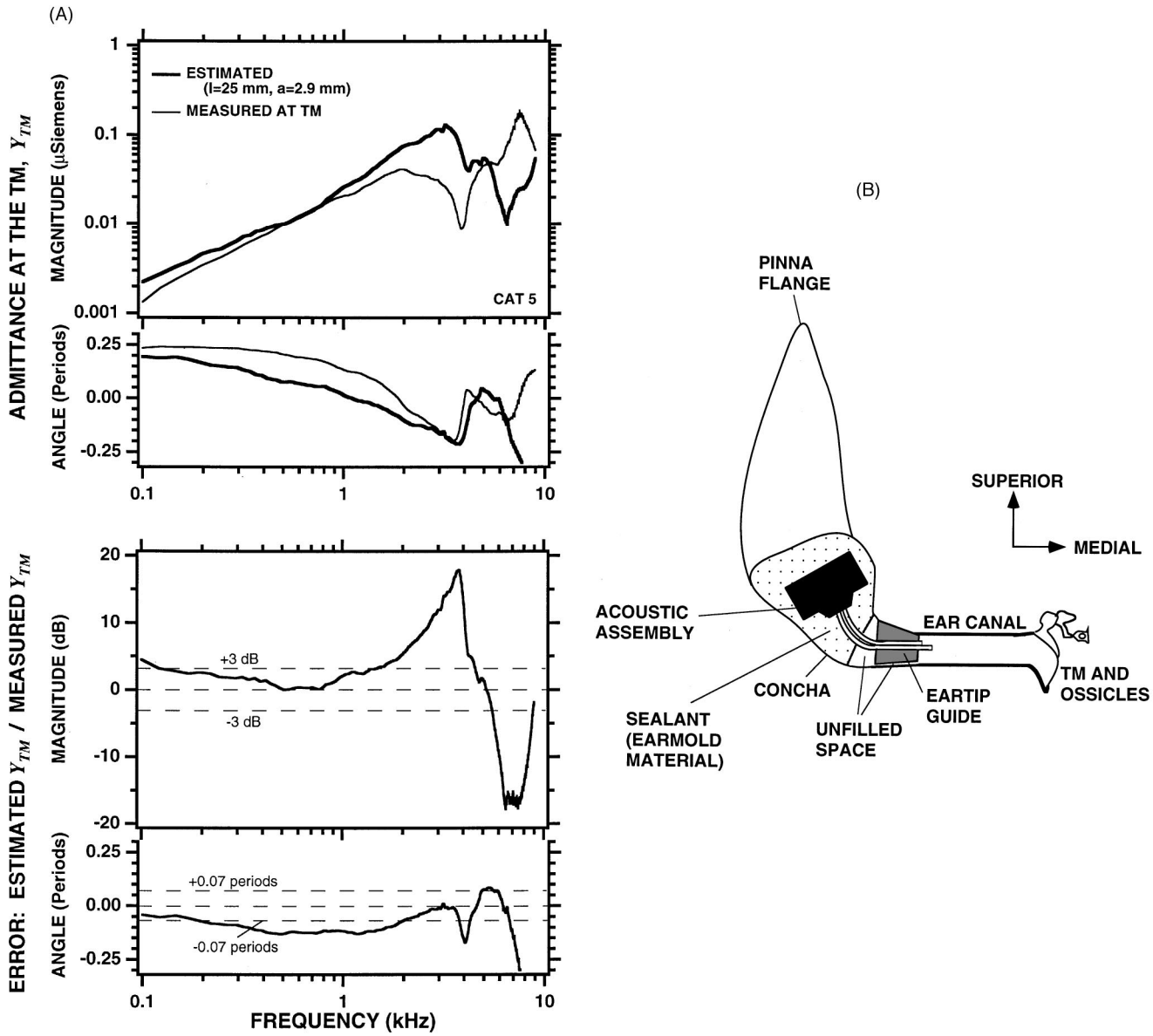


FIG. 5. (A) Estimated and measured admittances Y_{TM} for an ear (cat 5) in which the seal around the eartip in the intact canal was judged to be faulty (incomplete). The lower panel shows the error ratio of estimated admittance to measured admittance; the dashed horizontal lines indicate 0 ± 3 dB (magnitude) and 0 ± 0.07 periods (angle). For this insertion, the sealant (earmold material) injected behind the eartip guide did not completely fill the space in the canal and concha lateral to the guide. (B) Schematic representation of the incomplete filling of sealant in a cat's intact ear canal. An air space (''unfilled space'') is present at the canal-concha border, between the eartip guide and the medial extent of the sealant, and is coupled acoustically to the measurement point via a small gap between the guide and the canal wall. The drawing is based on (1) the appearance of the sealant mold after it was removed from the ear and (2) comparison with an earmold made before the acoustic-measurement system was inserted.

of a physical change in the TM. At the end of the intact-ear-canal measurement series, sharp features appeared in the sound-pressure response below 1 kHz. A small spot of blood was observed, via otoscope, on the anterodorsal quadrant of the TM. The blood was not present in earlier observations of the TM. After resection of the external ear, acoustic measurements made near the TM consistently showed the sharp features in Fig. 3(B). These observations suggest that a change in the TM occurred between the last few intact-canal measurements and the beginning of the resected-canal measurements. Still, Y_{TM}(meas.) from this ear has features that are within the ''normal'' range, as compared with Lynch *et al.* (1994): (1) a low-frequency compliance with an equivalent volume of 0.45 cm^3 and (2) sharp changes in magnitude and angle near 4 kHz that are consistent with a cavity notch. In addition, variations in measured compliance with static pressure in this ear were similar to those reported in Huang *et al.* [2000b, Fig. 11(c)].

2. Effects of an incomplete ear-canal seal

Here, similar comparisons of admittance are used to address a methodological problem in making noninvasive measurements in the ear canal. We found that measuring Y_{EC} with a faulty (incomplete) earmold-sealant injection can lead to large errors in Y_{TM}(est.), as shown in Fig. 5(A). For frequencies between 2 and 4 kHz, the estimated and measured admittance magnitudes differ by more than 5 dB. The cavity

notch near 4 kHz [in Y_{TM}(meas.)] is not captured in Y_{TM}(est.) and the magnitude error is nearly 20 dB (a factor of 10) at this frequency. For frequencies between 0.5 and 1.5 kHz, the error in angle is 0.12–0.14 periods. There was a clear difference in the configuration of the sealant with respect to the eartip in this ear compared to the ears of Fig. 3: The sealant did not completely fill the space lateral to the eartip guide [Fig. 5(B)]. Although static pressure in the canal was stable, it is likely that air volume lateral to the eartip was acoustically coupled to the canal (via small gaps between the guide and the canal wall and/or the sound tubes) and thus affected the Y_{EC} measurements. In particular, the canal-volume estimate, based on the low-frequency compliance measured with static pressure, was probably too large. This error would explain why the inferred canal length (chosen to match the canal volume given a radius estimate) was unrealistically large in this ear—25 mm; in our structural measurements on resected ears of domestic cats (e.g., Rosowski *et al.*, 1988), the canal length never exceeded 20 mm.

The result in Fig. 5 is instructive for the noninvasive measurement method in that it suggests a criterion by which to assess the quality of the ear-canal seal, independent of the acoustic response. Measurement series in two other ears (not reported here) yielded acoustic features of Y_{TM}(est.) qualitatively similar to those of Fig. 5(A) and shared the common physical features that (1) the sealant did not completely fill the space lateral to the eartip guide and (2) the inferred canal length was greater than 20 mm. We conclude that it is crucial for the sealant to fill the space in the canal lateral to the guide and around the sound tubes. Care must be taken to place the tip of the earmold syringe deeply into the concha—as close as possible to the eartip guide—before injecting the sealant. Upon removal of the acoustic system, the configuration of the sealant around the eartip can be examined. If the seal does not fill the space behind the eartip, the data from that measurement series should, in general, be rejected.

B. Compliance versus static pressure at two canal locations

In this section, we report measurements that test the idea that large static pressures in the ear canal render the middle ear effectively rigid but do not affect the acoustic properties of the canal, for low frequencies. Figure 6 compares the compliances measured in the lateral portion of the intact canal to the compliances measured near the TM, as functions of static-pressure level. The shapes of the compliance tympanograms are qualitatively similar among ears. For both measurement locations in all four ears, the maximum compliance occurs with zero static pressure in the ear canal, and the compliances are generally lower for negative canal pressures than for positive canal pressures of the same magnitude. The latter result is consistent with asymmetries reported previously in cat (Møller, 1965; Margolis *et al.*, 1978; Lynch, 1981, p. 250) which suggest that negative canal pressures (relative to the pressure in the middle ear) stiffen the middle ear more than positive pressures. For this reason, the compliance measured in the lateral portion of the canal at the most negative static-pressure level would seem to give a more accurate estimate of the canal volume than the measurement at the most positive static-pressure level.

To test whether static pressures make the middle ear effectively rigid, compliances measured near the TM can be compared to compliances measured in the lateral portion of the canal. Lateral-canal compliances at the most negative static pressure yield equivalent air volumes of 0.28–0.47 cm^3 among ears (Fig. 6, filled circles). Compliances measured near the TM at the most negative static pressure vary from 0.05 to 0.11 cm^3 among ears (Fig. 6, open squares); these values include the residual-canal volume as well as the stiffened middle-ear compliance. The ratios of near-TM compliance to lateral-canal compliance at the most negative static pressure in the four ears are 14% (0.05/0.37), 13% (0.05/0.38), 17% (0.08/0.47), and 35% (0.11/0.31), for a mean of 20%. At zero static pressure, these compliance ratios are 66% (0.73/1.1), 58% (0.50/0.86), 55% (0.45/0.82), and 47% (0.26/0.55), for a mean of 57%. That is, on average, large negative static pressure reduces the compliance near the TM from 57% to 20% of the compliance measured in the lateral canal (with the same static pressure). Because some residual canal is included in the near-TM measurements, these ratios underestimate the extent of middle-ear stiffening seen at the TM.

The extent to which the middle-ear compliance is reduced can also be quantified by computing the ratio of the near-TM compliance at the most negative static pressure to the near-TM compliance at zero static pressure (Fig. 6, open squares). The four ears yield 7% (0.05/0.73), 10% (0.05/0.50), 18% (0.08/0.45), and 42% (0.11/0.26), for a mean of 19%. That is, on average, the most negative static pressure reduces the compliance near the TM to less than 20% of its ambient value. As above, these computations underestimate the extent of middle-ear stiffening that would be seen at the TM, as all of the compliances include the contributions of residual-canal volumes.

To provide a basis for testing whether static pressure affects the properties of the canal, we compute the difference between each pair of compliance curves in Fig. 6. This difference (ΔC) is equivalent to the canal air volume between the two measurement points. If static pressure affects the middle ear *only*, then ΔC should be independent of static-pressure level. To a first approximation, this relationship holds. However, for three of the four ears (cats 3, 7, and 8R), ΔC varies by $\pm$10%–20% with static-pressure level; the largest variations occur in cat 3, where ΔC ranges between 0.32 and 0.50 cm^3 [Fig. 6(A)]. These variations are not monotonic with static pressure, and there are similarities in shape among all ears. As static-pressure level becomes positive, ΔC has a local maximum at negative or zero pressure, decreases to a local minimum at positive or zero pressure, and increases again. We consider the implications of these data in terms of static-pressure effects on the canal and middle ear in Sec. III B 1.

C. Reflectance in the ear canal

Admittance measurements in the lateral portion of the ear canal and at the TM can be used to test the idea that the power reflectance $|R|^2$ is independent of position along the

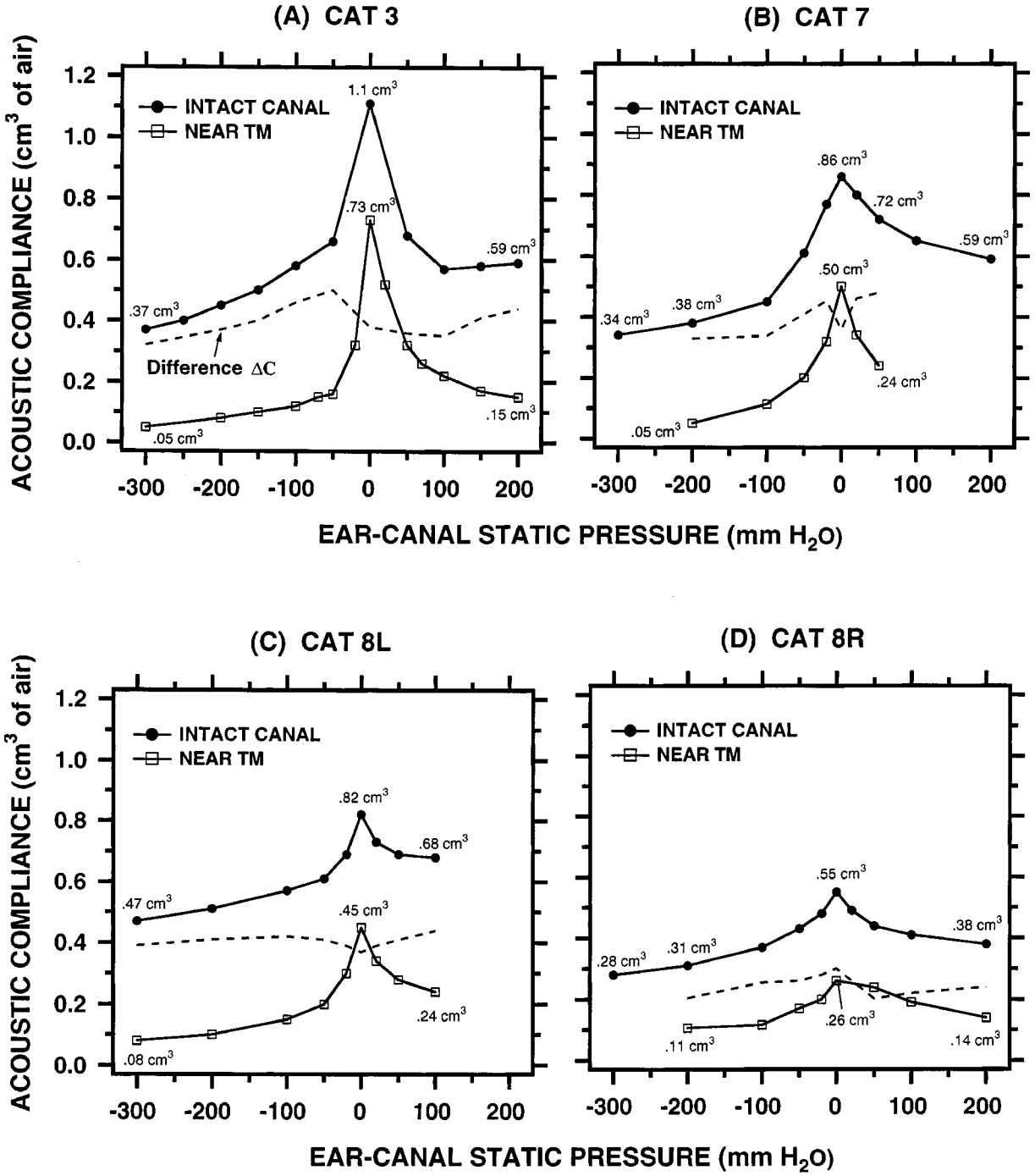


FIG. 6. Low-frequency acoustic compliances measured in the lateral portion of the ear canal and near the TM with varying static pressures. Data are from the four cat ears of Fig. 3. ''Intact canal'' data (filled circles) are from Y_{EC} measurements in the configuration of Fig. 1(A) (e.g., Huang *et al.*, 2000b, Fig. 11). ''Near TM'' data (open squares) are from admittances measured near the TM with most of the ear canal removed [Fig. 1(B)]. The data points, taken at discrete static-pressure levels, are connected by straight line segments. The dashed curves (''Difference ΔC'') are the differences between each pair of compliances, computed at static-pressure levels for which compliance data were taken at both canal locations. The axes are the same in all four plots. Acoustic-compliance values C are computed from admittances Y by averaging $\mathrm{Im}\{Y\}/(2\pi f)$ over ten frequency points f between 0.1–0.3 kHz, where Im{ } denotes the imaginary part. The compliances are plotted as equivalent volumes of air, computed as $(\rho_0 c^2)C$, where ρ_0 is the density of air and c is the speed of sound in air.

canal. Figure 7 shows power reflectances computed at the two locations in each ear canal. For each ear, $|R_{TM}|^2$ and $|R_{EC}|^2$ are qualitatively similar below 5 kHz, with the exception of the sharp features of cat 7 between 0.7 and 2 kHz (see Sec. II A 1). All of the reflectances have a broad minimum near 2 kHz and a sharper peak between 3 and 4 kHz. At the frequency of the minimum, the middle-ear input admittance is most closely matched to the characteristic admittance of the canal, and the greatest fraction of incident sound power in the canal is transmitted to the middle ear (i.e., $1-|R|^2 = 70\%-80\%$). The peak between 3 and 4 kHz corresponds in frequency to the sharp admittance notch (cavity notch) in the data of Fig. 3 (see also Lynch *et al.*, 1994, Figs. 22 and 23). The similarity of these spectral features within each pair of reflectance curves is consistent with the assumption of spatial independence of $|R|^2$.

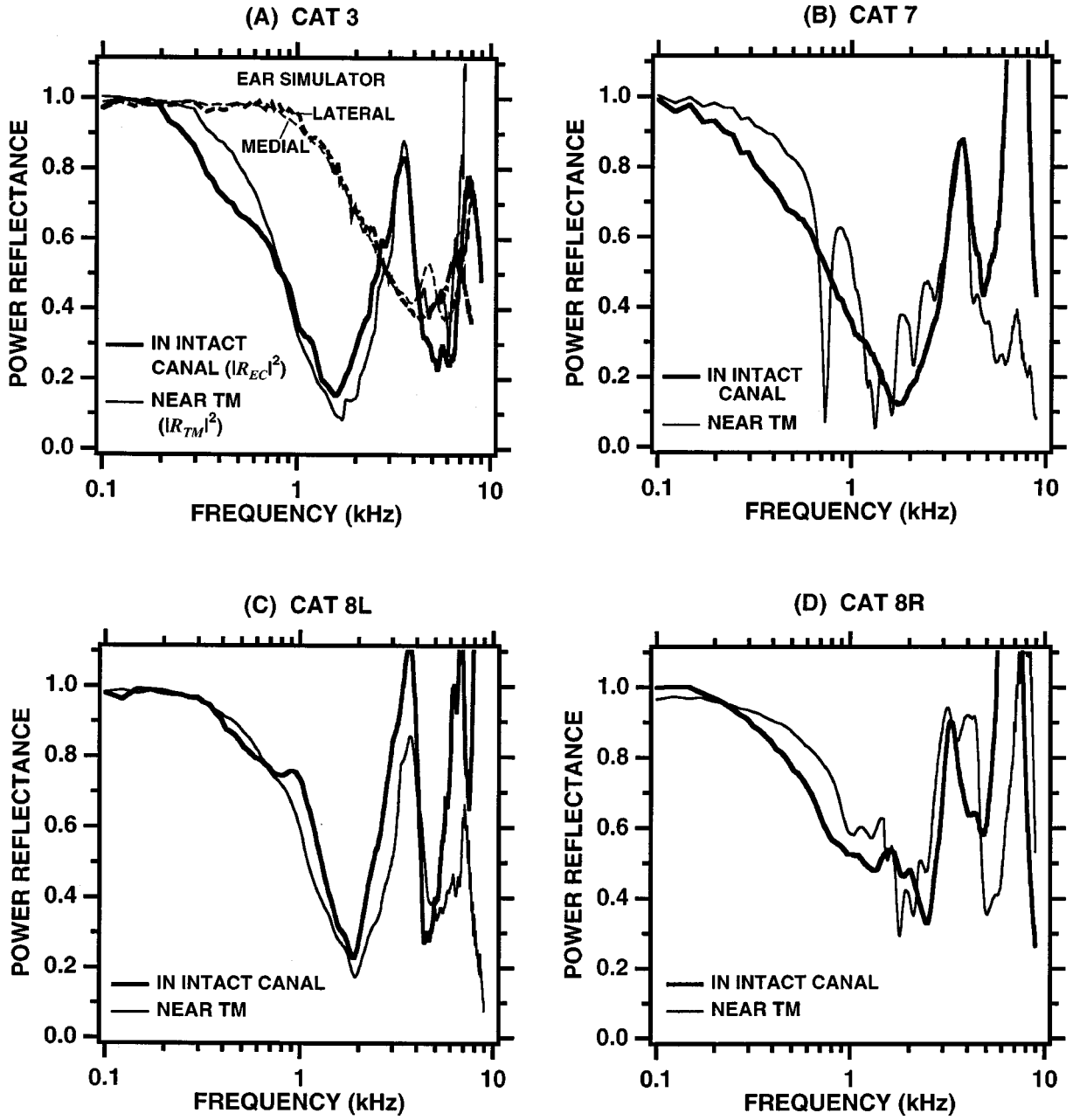


FIG. 7. Comparisons of power reflectance in the lateral portion of the ear canal ($|R_{EC}|^2$) to power reflectance near the TM ($|R_{TM}|^2$). Data are from the four cat ears of Fig. 3. Power reflectance, the squared magnitude of the pressure reflection coefficient, was computed from Eq. (1) with the measured admittances (Y_{EC} and Y_{TM}) and the acoustically estimated canal diameters at the respective locations in these ear canals [Keefe *et al.*, 1992; Huang *et al.*, 2000b, Eq. (6)]. For canal-diameter values, see Fig. 2 (lower) and Fig. 3 legends. In panel (A), the dashed curves (''lateral'' and ''medial'') are power reflectances computed from admittances measured 10 mm apart in a Zwislocki-type DB-100 (Knowles Electronics) ear simulator (after Huang *et al.*, 2000b, Fig. 8). Reflectance values greater than 1, which are due to errors in the measured admittances, are included to show the curves' shapes.

However, there are also differences between $|R_{EC}|^2$ and $|R_{TM}|^2$ (Fig. 7). For frequencies below 1 kHz, $|R_{EC}|^2$ is generally less than $|R_{TM}|^2$. Between 2 and 4 kHz, the pairs of reflectances from cats 8L and 8R have sizable differences. For frequencies above 5 kHz, the differences are larger still (and of varying sign) in all four ears; an explanation is that for higher frequencies, errors in the measured admittances are greater. We consider the implications of these reflectance comparisons in terms of energy loss in the ear canal in Sec. III C.

III. DISCUSSION

A. Accuracy of uniform-tube admittance transformation

1. General comments

The tests in cat ears show that a noninvasive method for estimating Y_{TM} that approximates the canal as a uniform tube is generally accurate—i.e., errors of less than 3 dB in magnitude and 0.07 periods in angle—for frequencies up to 3 kHz (Figs. 3 and 4). Although greater errors occur for higher frequencies, the method captures qualitative features of Y_{TM}

TABLE I. Percent differences between the estimated Y_{TM} and the measured Y_{TM} curves for 4 ears (see Figs. 3 and 4): % Difference=(Estimated−Measured)/(Measured)×100%. Ears: ''L'' and ''R'' denote left and right ears; ''(a)'' and ''(b)'' denote different insertions of the probe in an ear.[a] ''Absolute mean'' is the mean of the absolute values of the six differences in each row. The quantity C_{ME} is the total middle-ear acoustic compliance, computed as the average of $\mathrm{Im}\{Y_{TM}\}/(2\pi f)$ over ten frequency points f from 0.1 to 0.3 kHz, where Im{} denotes the imaginary part. $|Y_{TM}(0.22\ \mathrm{kHz})|$ is the admittance magnitude at 0.22 kHz. $f_{Y\mathrm{notch}}$ is the frequency of the local minimum in $|Y_{TM}|$ near 4 kHz. $|Y_{TM}(\text{measured } f_{Y\mathrm{notch}})|$ denotes comparison of admittance magnitudes at the frequency $f_{Y\mathrm{notch}}$ of the measured curve. $|Y_{TM}(\text{respective } f_{Y\mathrm{notch}})|$ denotes comparison of the estimated $|Y_{TM}|$ at the estimated curve's $f_{Y\mathrm{notch}}$ to the measured $|Y_{TM}|$ at the measured curve's $f_{Y\mathrm{notch}}$. The quantity $f_{45°}$ is the lowest frequency for which the angle of Y_{TM} is less than 1/8 of a period (45 degrees), except for cat 7 for which a sharp dip in angle near 0.8 kHz was discounted [see Fig. 3(B)].

% differences	Cat 3(a)	Cat 3(b)	Cat 7	Cat 8L	Cat 8R(a)	Cat 8R(b)	Absolute mean
C_{ME}	16%	19%	27%	−5%	9%	10%	14%
$\lvert Y_{TM}(0.22\ \mathrm{kHz})\rvert$	27%	34%	34%	−8%	10%	12%	21%
$\lvert Y_{TM}(\text{measured } f_{Y\mathrm{notch}})\rvert$	63%	400%	160%	160%	250%	800%	306%
$\lvert Y_{TM}(\text{respective } f_{Y\mathrm{notch}})\rvert$	58%	144%	140%	−2%	49%	302%	116%
$f_{Y\mathrm{notch}}$	1%	5%	6%	5%	4%	9%	5%
$f_{45°}$	−25%	−59%	−26%	6%	−28%	−36%	30%

[a]In each of two ears (cats 3 and 8R), two measurement series were accepted. Data from cats 3(b) and 8R(b) are not shown in the figures, to simplify the presentation.

up to 5 kHz. These data support the common assumption that a uniform-tube approximation of the canal can represent spatial variations in admittance for frequencies up to about 4 kHz.

In conducting these tests, we made an important methodological observation. The noninvasive measurement technique requires a stable and complete seal of the acoustic system into the ear canal. The injected sealant must completely fill the space lateral to the eartip guide to allow (1) accurate measurements of ear-canal admittance and (2) accurate estimation of the dimensions of the canal space between the measurement point and the TM. Measurement series that did not satisfy this criterion (Fig. 5) were excluded from the assessment of accuracy described above.

2. Features of middle-ear input admittance

Here, we quantify the differences between noninvasive estimates of middle-ear input admittance (which use a uniform-tube model of the intact canal) and measurements made near the TM, in terms of specific features of Y_{TM}. Table I summarizes the size of the percent differences, for the ears in Fig. 3, for six middle-ear input admittance features. The mean difference in total middle-ear acoustic compliance (C_{ME}) for low frequencies is 14%. C_{ME} quantifies the volume displacement of the TM per Pascal of sound pressure at the TM. The mean difference in $|Y_{TM}|$ at 0.22 kHz is 21%, and the difference in the frequency of the cavity notch near 4 kHz ($f_{Y\mathrm{notch}}$) is 5%. In the frequency region of the notch (3–5 kHz), differences in admittance magnitude are large because of the sharp changes in magnitude with frequency.[8]

For a given net compliance at the TM, the lowest frequency for which the angle is $\frac{1}{8}$ of a period or 45 degrees ($f_{45°}$) is a measure of the resistance of the middle ear and cochlea. The difference in values of $f_{45°}$ based on measurements at the two canal locations is 30%; the frequency tends to be lower—i.e., the resistance is higher—when estimated from measurements in the lateral canal. A possible explanation for this difference is that ear-canal losses are significant for these low frequencies (see Sec. III C). In summary, measurements in the lateral canal (transformed via a uniform tube) and measurements near the TM yield similar estimates of C_{ME} and $f_{Y\mathrm{notch}}$, and more divergent estimates of $|Y_{TM}|$ for frequencies near $f_{Y\mathrm{notch}}$ and the resistance of the ear.

3. Limiting factors

The estimated canal dimensions (length and radius) determine the uniform-tube transformation between Y_{EC} and Y_{TM}(est.). The effect of small changes in these parameters on Y_{TM}(est.) is well understood (Huang *et al.*, 2000b, Fig. 13). In an attempt to improve the agreement between Y_{TM}(est.) and Y_{TM}(meas.), we investigated heuristic schemes to adjust the canal parameters. We did not find a scheme that improved the agreement over a broad range of frequencies. For example, increasing the canal dimensions for cat 3 improves $|Y_{TM}(\text{est.})|$ for low frequencies but worsens the estimate of $f_{Y\mathrm{notch}}$ [see Fig. 3(A) and Huang *et al.*, 2000b, Fig. 13]. These observations lead us to conclude that the uniform-tube approximation of the canal—not the manner in which the canal parameters are chosen—is a limiting factor in the accuracy of the noninvasive method, for frequencies above 3 kHz. The ''baseline'' accuracy of our admittance measurements is also a limiting factor, for frequencies above 6–8 kHz (Huang *et al.*, 2000b).

B. Low-frequency effects of static pressure

1. Effects on the middle ear and ear canal

Does static pressure make the cat middle ear effectively rigid? The residual-canal volumes included in the near-TM compliances (Fig. 6) lead us to underestimate the extent of middle-ear stiffening with static pressure (Sec. II B). If we had independent estimates of the residual-canal volumes, we could subtract them from the near-TM compliances to estimate the middle-ear compliances; we could then deduce the intact-canal air volumes as well. Suppose we assume that in each ear the residual-canal volume is 0.04 cm^3, which is a conservative estimate based on volumetric saline-filling data

from similar cat preparations (Lynch *et al.*, 1994, p. 2193). Based on the near-TM data of Fig. 6, the compliance of the four stiffened middle ears would be 0.01, 0.01, 0.04, and 0.07 cm^3. Subtraction of these middle-ear compliances from the lateral-canal compliances (Fig. 6) would give intact-canal air volumes of 0.36, 0.37, 0.43, and 0.24 cm^3. Thus, the ratios of middle-ear compliance to intact-canal volume at the most negative static pressure would be 3% (0.01/0.36), 3% (0.01/0.37), 9% (0.04/0.43), and 29% (0.07/0.24), for a mean of 11%. The ratios of stiffened middle-ear compliance to ambient middle-ear compliance would be 1% (0.01/0.69), 2% (0.01/0.46), 10% (0.04/0.41), and 32% (0.07/0.22), for a mean of 11%. Therefore, based on a reasonable estimate of the residual-canal volumes, we conclude that large negative static pressures reduce the compliance of the cat middle ear to (1) about 10% of its ambient value and (2) about 10% of the compliance of the canal air volume.

Does static pressure affect the configuration of the canal space? In Table I, the derived compliances and low-frequency magnitudes of Y_{TM}(est.) are greater than those of Y_{TM}(meas.) by 14% and 21%, on average. A possible cause of this bias is that the canal-volume estimates from the intact-ear measurements at -300 mm H_2O (Fig. 6) are too *small*, by 14%–21%. (If the canal volume is underestimated, then the fraction of the measured compliance attributed to the middle ear will be too large.) This error is in the opposite direction from that which we would expect if large static pressure were not making the middle ear rigid. Mechanisms that would decrease the canal volume with applied negative static pressure include (1) static displacement of the TM in the lateral direction, (2) displacement of the acoustic probe in the medial direction, and (3) constriction of the canal walls. Here, we evaluate these possible explanations.

(1) With a one-dimensional piston-in-a-cylinder approximation, the volume displacement of the TM can be estimated by the product of the ear's static compliance and the static pressure acting on the TM. Assumption of the values 7.1×10^{-6} cm^3/Pa for the TM compliance (the acoustic compliance of 1 cm^3 of air) and 2940 Pa for the pressure acting on the TM (-300 mm H_2O) gives a TM volume displacement of 0.02 cm^3, which is 4%–7% of the estimated canal volumes (0.28–0.47 cm^3). However, data reviewed by Rabinowitz (1981, Fig. 12) suggest that in human ears the compliance at the TM increases from 0.8 cm^3 at 0.1 kHz to about 2 cm^3 in the quasi-static limit. *If* such a compliance change occurred in cat ears, a TM volume displacement of 0.04 cm^3 would affect the canal-volume estimate by 9%–14% [0.04/0.28 is the ''worst'' case, in Fig. 6(D)].

(2) With a canal radius of 3.1–3.6 mm at the measurement point (see Fig. 3 legend), displacement of the acoustic probe by 1 mm (say) toward the TM would result in a volume displacement of 0.03–0.04 cm^3. Thus, a 1-mm eartip displacement would affect the canal-volume estimate by 6%–14%.

(3) For a given ear, there are nonmonotonic variations of 10%–20% in ΔC (the difference between each pair of curves in Fig. 6) versus static-pressure level. These data do not suggest a simple relationship between static-pressure level and canal volume over the extent of static-pressure levels tested.[9] However, ΔC at the most negative static pressure is smaller than that at zero pressure in three out of four ears. If the variations in ΔC versus static pressure result from flexing of the canal wall (or movement of the probe, as described above), the displacements are small. With average canal dimensions—volume 0.37 cm^3, length 10 mm, and radius 3.4 mm—the canal radius would need to decrease by 0.1–0.3 mm to achieve a 7%–14% decrease in canal volume.

The sizes of the hypothetical displacements above seem physically plausible. We conclude that the combined effects of these mechanisms could be a substantial contribution to the observed low-frequency differences between Y_{TM}(est.) and Y_{TM}(meas.), and to the observed asymmetries with respect to positive and negative pressures. In summary, the data support the idea that large negative static pressures make the cat middle ear effectively rigid and can also affect the configuration of the canal space.

2. Comparisons to previous data

The relative changes in acoustic compliance with static pressure measured near the TM (Fig. 6, open squares) are consistent with those previously reported in cat ears (Møller, 1965, Fig. 11; Margolis *et al.*, 1978, Fig. 4; Lynch, 1981, p. 250). These data indicate that compliances measured with negative canal pressures are smaller than those measured with positive canal pressures of the same magnitude, where the sign of the pressure is relative to the pressure in the middle-ear cavity. The measurements of Lynch (1981) were made with static pressures introduced into the middle-ear cavity, not the ear canal. The presence of asymmetries with respect to the sign of static pressure in Lynch's data indicates that these asymmetries cannot be explained solely by static displacements of the canal wall and/or the acoustic probe (not likely to occur with Lynch's preparation). The suggestion is that static displacements of the TM (see above) may contribute to the asymmetric acoustic response of the ear to static pressures.

A combination of acoustic and psychophysical data suggests that large static pressure in human ear canals reduces the compliance at the TM to about 28% of its ambient value (Rabinowitz, 1981, pp. 1029–1030, $n=4$). The acoustic data reported here suggest that large static pressure in cat ear canals reduces the compliance at the TM (the middle-ear compliance) to about 10% of its ambient value ($n=4$). It is unclear whether the disparity between the human and cat data reflects differences among species or differences in methodology, as the sample sizes are small.

C. Evidence of ear-canal losses

Is there any evidence here that disagrees with a lossless-canal assumption? For frequencies between 1 and 5 kHz, many of the differences between $|R_{TM}|^2$ and $|R_{EC}|^2$ (Fig. 7) can be accounted for by variations of $\pm10\%$ in the two ear-canal diameters, which is within our tolerance;[10] the excep-

tions are (1) the extrema of $|R_{TM}|^2$ in cat 7 between 0.7 and 2 kHz, (2) $|R_{TM}|^2<|R_{EC}|^2$ in cat 8L between 2 and 4 kHz, and (3) $|R_{TM}|^2>|R_{EC}|^2$ in cat 8R between 2.3 and 4.5 kHz. That is, between 1 and 5 kHz, most of the differences between the reflectances at the two locations are not significant with respect to our measurement error. However, for $f=0.5\pm0.1$ kHz, all four ears show $|R_{TM}|^2>|R_{EC}|^2$, and this result *cannot* be accounted for by variations of $\pm10\%$ in the diameters. This frequency region is the only region for which the difference between the reflectances at the two locations is significant *and* has the same sign for all four ears. At 0.5 kHz, the values of $|R_{TM}|^2-|R_{EC}|^2$ for Figs. 6(A)–6(D) are 0.13, 0.16, 0.05, and 0.11, respectively, for a mean value of 0.11. For comparison, we computed reflectances from the admittances measured at different locations in an ear simulator [Fig. 7(A)]. The difference between power reflectances measured 10 mm apart in the smooth, rigid, simulated canal is less than 0.02 for frequencies up to 1 kHz.

With assumption of a uniform-tube canal, these results imply that roughly 10% of the incident power is dissipated in the cat ear canal for frequencies near 0.5 kHz. At this frequency (where the uniform approximation is reasonable) the dissipation in air from viscous and thermal losses (<1%) cannot account for a 10% loss, based on the canal lengths, radii, properties of air, and a lossy transmission-line model. Therefore, although the data do not provide a compelling measure of ear-canal losses over most of the frequency range, they are consistent with the idea that (1) in general losses in the canal are small, and (2) a loss mechanism other than dissipation in air (e.g., nonrigid behavior of the canal walls, or complexities in the canal's shape) is present for frequencies near 0.5 kHz.[11]

D. Clinical implications

In ''multifrequency tympanometry,'' the acoustic response of the ear is measured as a function of frequency and static-pressure level in the canal (Vanhuyse *et al.*, 1975; Colletti, 1976; Margolis *et al.*, 1999). This procedure may be performed when a patient exhibits a conductive hearing loss and a normal low-frequency tympanogram, which may occur with ossicular pathologies such as disarticulation or otosclerosis. The diagnostic value of multifrequency tympanometry might be enhanced by more accurate removal of the acoustic effects of the ear canal for higher frequencies, so that the results are more sensitive to middle-ear abnormalities. The results of this article demonstrate that the acoustic effects of the cat ear canal can be removed accurately for frequencies up to 3 kHz with a uniform-tube transformation. Based on ear-canal sizes, one might expect the frequency range of validity to be similar for human infants, and somewhat lower for human adults because of the greater canal dimensions in adults. For low frequencies, our results support the idea that large static pressures render the middle ear rigid compared to the canal air volume (at least in cats). We cannot rule out the possibility that asymmetries in tympanograms result in part from changes in the configuration of the canal space, in particular, static displacement of the TM. Still, the results of this article define bounds on the effects of static pressure on the canal and middle ear; these bounds might be used to estimate the size of errors in clinical measurements.

Recently there has been interest in developing diagnostic measures based on ear-canal reflectance (Keefe *et al.*, 1993; Keefe and Levi, 1996; Piskorski *et al.*, 1999; Margolis *et al.*, 1999; Feeney and Keefe, 1999). The clinical utility of reflectance depends on the assumption that the reflectance magnitude is insensitive to measurement position in the canal. The results of this article indicate that this assumption is generally valid in cat ears below 5 kHz, except for frequencies near 0.5 kHz (±0.1 kHz) for which we have reported evidence of propagation losses in the canal. As frequencies between 0.5 and 2 kHz are thought to be important for speech perception, this result warrants further investigation.

IV. CONCLUSIONS

(1) For representing spatial variations in acoustic admittance, a uniform-tube approximation of the cat ear canal is accurate to within 3 dB in magnitude and 0.07 periods in angle for frequencies up to 3 kHz, with greater errors for higher frequencies. This result demonstrates the accuracy of a noninvasive admittance-measurement method (Huang *et al.*, 2000b).

(2) For low frequencies (<0.5 kHz), large negative static pressures in the canal reduce the compliance of the cat middle ear to about 10% of the canal compliance. Data suggest that static displacements of the acoustic probe, TM, and/or canal wall may also occur; the effect on estimates of canal volume and middle-ear compliance may be as large as 15% to 20%.

(3) Reflectance magnitude (and sound power transmission) is relatively insensitive to position along the canal for most frequencies below 5 kHz. For frequencies near 0.5 kHz, however, there is evidence of losses incurred via propagation through the canal.

ACKNOWLEDGMENTS

We thank Denny Freeman and the late Tom McMahon for helpful discussions and for comments on earlier versions of this manuscript; Chris Shera and Susan Voss for comments on the manuscript; Leslie Liberman for assistance with animal preparations; and the staff of the Eaton–Peabody Laboratory. Portions of this work were presented at the 21st Midwinter Meeting of the Association for Research in Otolaryngology (Huang *et al.*, 1998) and in a doctoral thesis (Huang, 1999). This work was supported by NSF and NIDCD.

[1] Y_{TM} is defined as the ratio of the complex amplitudes of the volume velocity of the TM and the sound pressure just lateral to the TM.

[2] R is defined as the ratio of the complex amplitudes of the laterally traveling pressure wave and the medially traveling pressure wave in the ear canal.

[3] Uniform-tube models have also been used to represent (1) the admittance at the tympanic ring in the direction of the external ear (Rosowski *et al.*, 1988) and (2) the sound-pressure distribution along the canal (Stinson *et al.*, 1982).

[4] $Y_0=A/(\rho_0 c)$, where A is the cross-sectional area of the canal at the measurement point, ρ_0 is the density of air, and c is the speed of sound in air.

[5] The power reflectance describes features of power flow in the ear canal, but

it does not characterize the mechanics of the middle ear as seen from the canal. Knowledge of the reflectance angle and the distance between the measurement point and the TM (or equivalently, Y_{TM}), are necessary for a more complete characterization. For a review of reflectance and impedance, see Puria and Allen (1998).

[6]Above 8 kHz, errors in our baseline admittance measurements become significant as well (Huang *et al.*, 2000b).

[7]Lynch *et al.* (1994) estimated the canal dimensions with a different approach from that described in the caption of Fig. 2. They set the canal radius equal to the radius of their acoustic probe (1.95 mm), and measured the canal volume by filling the residual space with saline up to a mark in the canal wall made by the probe. In general, such a mark was not made by the eartips used in this article. However, the acoustically estimated volumes in two ears (Cats 8L and 8R) are roughly consistent with (and smaller than) volumes measured by saline-filling up to the entrances of these residual canals (0.10–0.13 cm^3). Also, the acoustically estimated diameters are consistent with caliper estimates of diameters at the canal entrances in all four ears (4–6 mm).

[8]The admittance-notch feature is considered here for two reasons. (1) The notch occurs near the high-frequency limit of the spatial-transformation method. (2) We are interested in the possible functional significance of the notch in species of the cat family (Huang *et al.*, 2000a); hence we wish to quantify the errors in measuring it.

[9]For example, the simple hypothesis that negative pressures constrict the canal walls and positive pressures expand the canal walls is not supported. A potential complication is that the response of the ear to static pressures could depend on the time course of the measurement series and/or the direction of static-pressure changes. (For information on these experiments, see Huang *et al.*, 2000b, Sec. II A 4 and caption of Fig. 11.)

[10]For each ear, we computed ratios of $|R_{EC}|^2/|R_{TM}|^2$ with canal diameters altered by $\pm 10\%$ to define a range of values versus frequency. In frequency regions for which the range included 0 dB, differences between $|R_{EC}|^2$ and $|R_{TM}|^2$ were considered accountable by errors in canal diameter.

[11]The acoustic power in the canal is given by $|P|^2\,\mathrm{Re}\{Y\}$, where P is sound pressure, Y is admittance, and Re{ } denotes the real part. For low frequencies (i.e., 0.5 kHz) P is approximately constant with position in the canal, so that power dissipation may be assessed by comparing $\mathrm{Re}\{Y_{EC}\}$ to $\mathrm{Re}\{Y_{TM}\}$. The only frequency region in which $\mathrm{Re}\{Y_{EC}\} > \mathrm{Re}\{Y_{TM}\}$ (signifying canal losses) for all four ears is 0.5 ± 0.1 kHz. Unfortunately, Re{Y} is sensitive to small measurement errors for low frequencies, so that it is difficult to quantify losses with this measure.

Avan, P. M., Büki, B., Maat, B., Dordain, M., and Wit, H. P. (**2000**). ''Middle ear influence on otoacoustic emissions. I:Noninvasive investigation of the human transmission apparatus and comparison with model results,'' Hear. Res. **140**, 189–201.

Burns, E. M., Keefe, D. H., and Ling, R. (**1998**). ''Energy reflectance in the ear canal can exceed unity near spontaneous otoacoustic emission frequencies,'' J. Acoust. Soc. Am. **103**, 462–474.

Colletti, V. (**1976**). ''Tympanometry from 200 to 2000 Hz probe tone,'' Audiology **15**, 106–119.

Feeney, M. P., and Keefe, D. H. (**1999**). ''Acoustic reflex detection using wide-band acoustic reflectance, admittance, and power measurements,'' J. Speech Lang. Hear. Res. **42**, 1029–1041.

Huang, G. T. (**1999**). ''Measurement of middle-ear acoustic function in intact ears: Application to size variations in the cat family,'' Ph.D. thesis, Massachusetts Institute of Technology.

Huang, G. T., Rosowski, J. J., and Peake, W. T. (**2000a**). ''Relating middle-ear acoustic performance to body size in the cat family: Measurements and models,'' J. Comp. Physiol. A **186**, 447–465.

Huang, G. T., Rosowski, J. J., Puria, S., and Peake, W. T. (**1998**). ''Noninvasive technique for estimating the acoustic impedance at the tympanic membrane in ear canals of different size,'' in Abstracts of the 21st Midwinter Research Meeting of the Association for Research in Otolaryngology, p. 111.

Huang, G. T., Rosowski, J. J., Puria, S., and Peake, W. T. (**2000b**). ''A noninvasive method for estimating acoustic admittance at the tympanic membrane,'' J. Acoust. Soc. Am. **108**, 1128–1146.

Huang, G. T., Rosowski, J. J., Flandermeyer, D. T., Lynch III, T. J., and Peake, W. T. (**1997**). ''The middle ear of a lion: Comparison of structure and function to domestic cat,'' J. Acoust. Soc. Am. **101**, 1532–1549.

Hudde, H. (**1983**). ''Measurement of the eardrum impedance of human ears,'' J. Acoust. Soc. Am. **73**, 242–247.

Jerger, J. F., ed. (**1975**). *Handbook of Clinical Impedance Audiometry* (American Electromedics, New York).

Keefe, D. H. (**1997**). ''Otoreflectance of the cochlea and middle ear,'' J. Acoust. Soc. Am. **102**, 2849–2859.

Keefe, D. H., and Levi, E. (**1996**). ''Maturation of the middle and external ears: Acoustic power-based responses and reflectance tympanometry,'' Ear Hear. **17**, 361–373.

Keefe, D. H., Ling, R., and Bulen, J. C. (**1992**). ''Method to measure acoustic impedance and reflection coefficient,'' J. Acoust. Soc. Am. **91**, 470–485.

Keefe, D. H., Bulen, J. C., Arehart, K. H., and Burns, E. M. (**1993**). ''Ear-canal impedance and reflection coefficient in human infants and adults,'' J. Acoust. Soc. Am. **94**, 2617–2638.

Khanna, S. M., and Stinson, M. R. (**1985**). ''Specification of the acoustical input to the ear at high frequencies,'' J. Acoust. Soc. Am. **77**, 577–589.

Lynch III, T. J. (**1981**). ''Signal processing by the cat middle ear: Admittance and transmission, measurements and models,'' Ph.D. thesis, Massachusetts Institute of Technology.

Lynch III, T. J., Peake, W. T., and Rosowski, J. J. (**1994**). ''Measurements of the acoustic input-impedance of cat ears: 10 Hz to 20 kHz,'' J. Acoust. Soc. Am. **96**, 2184–2209.

Margolis, R. H., and Shanks, J. (**1985**). ''Tympanometry,'' in *Handbook of Clinical Audiology*, edited by J. Katz (Williams and Wilkens, Baltimore, MD), pp. 438–475.

Margolis, R. H., Osguthorpe, J. D., and Popelka, G. R. (**1978**). ''The effects of experimentally-produced middle ear lesions on tympanometry in cats,'' Acta Oto-Laryngol. **86**, 428–436.

Margolis, R. H., Saly, G. L., and Keefe, D. H. (**1999**). ''Wideband reflectance tympanometry in normal adults,'' J. Acoust. Soc. Am. **106**, 265–280.

Møller, A. R. (**1965**). ''An experimental study of the acoustic impedance of the middle ear and its transmission properties,'' Acta Oto-Laryngol. **60**, 129–149.

Morse, P. M., and Ingard, K. U. (**1968**). *Theoretical Acoustics* (McGraw-Hill, New York), pp. 343–346.

Mundie, J. R. (**1963**). ''The impedance of the ear—a variable quantity,'' in U.S. Army Medical Research Lab. Report 576, edited by J. L. Fletcher, pp. 65–85.

Onchi, Y. (**1961**). ''Mechanism of the middle ear,'' J. Acoust. Soc. Am. **33**, 794–805.

Piskorski, P., Keefe, D. H., Simmons, J. L., and Gorga, M. P. (**1999**). ''Prediction of conductive hearing loss based on acoustic ear-canal response using a multivariate clinical decision theory,'' J. Acoust. Soc. Am. **105**, 1749–1764.

Puria, S., and Allen, J. B. (**1998**). ''Measurements and model of the cat middle ear: Evidence of tympanic membrane acoustic delay,'' J. Acoust. Soc. Am. **104**, 3463–3481.

Puria, S., and Rosowski, J. J. (**1997**). ''Measurement of reverse transmission in the human middle ear: Preliminary results,'' in *Diversity in Auditory Mechanics*, edited by E. R. Lewis, G. R. Long, R. F. Lyon, P. M. Narins, C. R. Steele, and E. Hecht-Poinar (World Scientific, Singapore), pp. 151–157.

Rabbitt, R. D., and Holmes, M. H. (**1988**). ''Three-dimensional acoustic waves in the ear canal and their interaction with the tympanic membrane,'' J. Acoust. Soc. Am. **83**, 1064–1080.

Rabinowitz, W. (**1981**). ''Measurement of the acoustic input immittance of the human ear,'' J. Acoust. Soc. Am. **70**, 1025–1035.

Rosowski, J. J., Carney, L. H., and Peake, W. T. (**1988**). ''The radiation impedance of the external ear of cat: Measurements and applications,'' J. Acoust. Soc. Am. **84**, 1695–1708.

Rosowski, J. J., Carney, L. H., Lynch III, T. J., and Peake, W. T. (**1986**). ''The effectiveness of the external and middle ears in coupling acoustic power into the cochlea,'' in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer-Verlag, New York), pp. 3–12.

Shaw, E. A. G. (**1974**). ''The external ear,'' in *Handbook of Sensory Physiology: Volume V/1: Auditory System*, edited by W. D. Keidel and W. D. Neff (Springer-Verlag, New York), pp. 455–490.

Shera, C. A., and Zweig, G. (**1992**). ''Analyzing reverse middle-ear transmission: Noninvasive Gedankenexperiments,'' J. Acoust. Soc. Am. **92**, 1371–1381.

Stinson, M. R. (**1985**). ''The spatial distribution of sound pressure within scaled replicas of the human ear canal,'' J. Acoust. Soc. Am. **78**, 1596–1602.

Stinson, M. R., and Khanna, S. M. (**1994**). ''Spatial distribution of sound pressure and energy flow in the ear canals of cats,'' J. Acoust. Soc. Am. **96**, 170–180.

Stinson, M. R., Shaw, E. A. G., and Lawton, B. W. (**1982**). ''Estimation of acoustical energy reflectance at the eardrum from measurements of pressure distribution in the human ear canal,'' J. Acoust. Soc. Am. **72**, 766–773.

Vanhuyse, V. J., Creten, W. L., and Van Camp, K. J. (**1975**). ''On the W-notching of tympanograms,'' Scand. Audiol. **4**, 45–50.

Voss, S. E., and Allen, J. B. (**1994**). ''Measurement of acoustic impedance and reflectance in the human ear canal,'' J. Acoust. Soc. Am. **95**, 372–384.

Wilber, L. A., and Feldman, A. S. (**1976**). ''The middle ear measurement battery,'' in *Acoustic Impedance and Admittance: The Measurement of Middle Ear Function*, edited by A. S. Feldman and L. A. Wilber (Williams and Wilkins, Baltimore, MD), pp. 345–377.

Zwislocki, J. (**1962**). ''Analysis of the middle-ear function. Part I: Input impedance,'' J. Acoust. Soc. Am. **34**, 1514–1523.

Zwislocki, J. J. (**1970**). ''An acoustic coupler for earphone calibration,'' Laboratory of Sensory Communication Special Report LSC-S-7, Syracuse University.

Perceived continuity and pitch perception

Christopher J. Plack
Department of Psychology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, England

Louise J. White
Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, England

(Received 25 October 1999; revised 19 January 2000; accepted 4 May 2000)

Three experiments investigated the importance of perceived stimulus continuity for the perception of the fundamental frequency ($F0$) of an unresolved complex tone. The $F0$ of the complex was 250 Hz and the harmonics were bandpass filtered between 5500 and 7500 Hz. In the first experiment, $F0$ discrimination was measured for single-burst tones with durations of 20, 40, and 80 ms, and for stimuli containing two 20- or 40-ms tone bursts separated by an 8- or 16-ms gap. For the single-burst conditions, there was a large decrease in threshold as the duration was increased from 20 to 40 ms. However, performance in the gapped conditions was much worse than that for the single-burst condition with the same cumulative duration (e.g., two 20-ms bursts separated by 8 ms produced higher thresholds than one 40-ms burst). Adding a bandpass noise (with the same spectral envelope as the tone) in the gap between the two tone bursts improved performance to the level of the single-burst condition. When the noise was added, the two discrete tone bursts were perceived as one single tone burst interrupted by the noise, and this seemed to facilitate discrimination. In a second experiment, the effects on pitch of an envelope delay (phase shift) of 0.75 periods between two tone bursts separated by an 8-ms gap were investigated. If the gap was silent, the pitch of the pair was unaffected by the phase shift. However, if the gap contained the bandpass noise, the phase shift between the bursts did produce a significant downward shift in the pitch of the pair. Finally, the third experiment showed that presenting a noise before a single 20-ms burst may improve discrimination performance in some listeners, but not sufficiently to account for the results of the first experiment purely in terms of an improvement in the discriminability of the second tone burst in the pair. The experiments suggest that a level decrease between two tone bursts may disrupt or reset a long integration mechanism, decreasing performance. When there is no level decrease between the bursts, the auditory system may assume that the two bursts belong to the same single tone and analyze them together in order to derive $F0$.
[S0001-4966(00)00209-5]

PACS numbers: 43.66.Ba, 43.66.Hg, 43.66.Mk, 43.66.Nm [DWG]

INTRODUCTION

White and Plack (1998) investigated pitch integration mechanisms for complex tones consisting only of harmonics unresolved by the peripheral auditory system. Fundamental frequency ($F0$) discrimination performance was measured in terms of the detectability index, d', for a complex tone with an $F0$ of 250 Hz and with harmonics bandpass filtered between 5500 and 7500 Hz. White and Plack found that there was about a threefold increase in d' when the duration of the complex was increased from 20 to 40 ms. The improvement in detectability was much greater than that expected on the basis of multiple independent sampling using a brief time window (Green and Swets, 1966; Viemeister and Plack, 1993; Viemeister and Wakefield, 1991). Instead, the results suggest that the auditory system uses some sort of long-term analysis mechanism for determining $F0$. In a second experiment, however, White and Plack found that when the 40-ms burst was split into two 20-ms bursts separated by a gap of only 5 ms, performance was almost identical to the predictions of the multiple looks model (which predicts a $\sqrt{n}$ improvement in d' for every n-fold increase in duration). In the case of the gapped stimuli, the data suggest that the auditory system was sampling the pitch of the two tone bursts independently.

White and Plack argued that the long integration window for $F0$ analysis may be reset in response to discontinuities in the stimulus, producing separate pitch estimates for the tone bursts on either side of the gap. In the past, the resetting hypothesis has been discussed in relation to the pitch of pure tones (Bregman *et al.*, 1994a, b; Nabelek *et al.*, 1973; Nabelek, 1996) and in relation to binaural processing (Hafter and Buell, 1990). To test their hypothesis, Plack and White (2000) introduced an instantaneous phase change to the envelope of the complex, such that the interpulse interval (IPI) between the waveform peaks of the two 20-ms bursts was not equal to an integer number of $F0$ periods. The manipulation produced a pitch shift when the bursts were contiguous, but had little effect on pitch when there was a gap between the bursts. The presence of the pitch shift is consistent with a long-term, phase-sensitive, analysis mechanism for uninterrupted stimuli. The absence of the pitch shift for the gapped stimuli suggests that a long-term mechanism was not operating in these conditions.

If the pitch mechanism is being reset, then the question arises as to what is causing the resetting. More generally it can be asked: What is it about the gapped stimuli that causes the performance deterioration compared to the uninterrupted stimuli? After all, the actual information content of the two

types of stimuli is the same. Two possibilities spring to mind:

(i) The level discontinuity produces the reduction in performance.
(ii) The loss of synchrony (i.e., regular pitch pulses) during the gap produces the reduction in performance.

There is evidence that the auditory system is sensitive to the timing of pitch pulses in an unresolved complex tone, even when the mean rate of pulses is held constant (Carlyon, 1997). Many models of pitch perception depend on this synchrony to derive $F0$ (Licklider, 1951; Meddis and Hewitt, 1991).

Either of these physical changes may disrupt the pitch mechanism. In terms of the resetting hypothesis, these events may be used by the auditory system as cues that a separate auditory object, requiring independent analysis, has been presented (White and Plack, 1998). The experiments presented here test these two hypotheses and provide a further test of the resetting hypothesis.

I. EXPERIMENT 1: EFFECT OF LEVEL CONTINUITY ON *F0* DIFFERENCE LIMENS

A. Rationale

It has been observed that an interrupted signal is perceived as being continuous if the period of the interruption is filled with a masker that would have obscured the signal had the signal actually been continuous (Bregman, 1990; Houtgast, 1972; Warren, 1999; Warren *et al.*, 1972). In the case of interrupted connected speech, presenting noise in the gaps can lead to an improvement in intelligibility (Cherry and Wiley, 1967; Dirks and Bower, 1970). In the White and Plack (1998) design, if the detriment in performance in the gapped conditions is the result of a level discontinuity, then introducing a noise into the gap such that the tone is perceived as being uninterrupted may actually improve discrimination performance. (In terms of the resetting hypothesis, the noise may prevent resetting, allowing the integration of information from the tone bursts on either side of the gap.) On the other hand, if the performance detriment is due to a discontinuity in temporal synchrony, then adding the noise should not improve performance.

In addition to the noise presented in the gap, a second, low-pass noise was added to all the stimuli. The low-pass noise was presented to eliminate the possibility of combination tones (that may fall within the region of resolvability for a 250-Hz $F0$) being used to perform the discrimination. The spectrum level of 15 dB was chosen to be sufficient to mask any expected distortion (Goldstein, 1967).

B. Stimuli

Stimuli consisted of complex tone bursts, and two distinct noise bursts. The complex-tone bursts were composed of the harmonics of a 250-Hz $F0$, bandpass filtered between 5500 and 7500 Hz (3-dB downpoints, 90-dB/octave slope). The harmonics were always added in sine phase and each harmonic component had a level of 50 dB SPL.

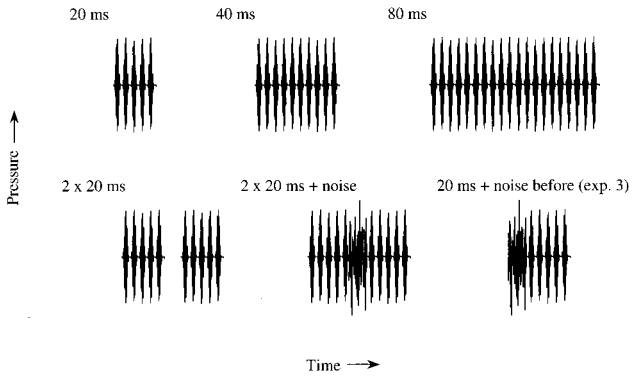


FIG. 1. An illustration of the temporal characteristics of some of the stimuli used in the experiments.

Figure 1 illustrates the temporal characteristics of the stimuli. For the single-burst conditions, the tone burst was 20, 40, or 80 ms in duration. For the gapped conditions, either two 20-ms bursts (2×20) or two 40-ms bursts (2×40) were presented, separated by a gap of either 8 or 16 ms. The starting envelope phase of the complex was random and the envelope phase was preserved across the gap (see Plack and White, 2000). The gap was either silent or filled with a bandpass noise, passed through the same filter as the complex tone, with a spectrum level of 26 dB. The level was chosen so that the rms level of the noise and the tone would be the same. In other words, there should have been no change in overall level during the course of the stimuli in the noise conditions. It was found that this was sufficient to produce an illusion of continuity. Higher noise levels were not used to avoid masking effects. The second noise, which was presented with all the stimuli, was low-pass filtered at 2500 Hz, and had a spectrum level of 15 dB. The noise was gated on 50 ms before the onset of the first tone burst and off 50 ms after the offset of the second tone burst. There were no onset or offset ramps for either the tone bursts or the noise, although spectral splatter was avoided by filtering the stimuli after gating. The filtering had the effect of introducing very brief (around 1 ms) ramps to the stimuli. Stimuli were presented to the right ear via Sennheiser HD580 headphones.

C. Procedure

A two-interval, two-alternative forced-choice (2AFC) procedure was used throughout and feedback was provided after every trial. The interstimulus interval was 500 ms. A two-down, one-up adaptive tracking procedure (Levitt, 1971) was used to estimate the $F0$ discrimination thresholds. The $F0$ of the complex in the standard interval was fixed at 250 Hz, and the $F0$ in the comparison interval had a higher $F0$, which was varied. The percentage $F0$ difference between the two observation intervals was increased and decreased by a factor of 2 for the first four turnpoints and by a factor of 1.41 thereafter. Sixteen turnpoints were measured and the threshold estimate was taken as the geometric mean of the last 12. The seven conditions were presented randomly until each listener had completed at least five such runs/threshold estimates for each condition. The experiment took about 2 h for each listener.

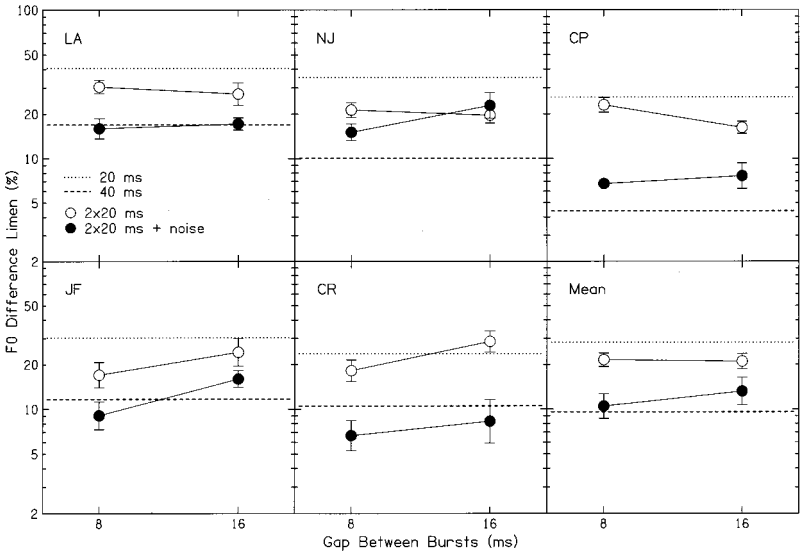


FIG. 2. The individual and mean results of experiment 1 showing $F0$ discrimination performance as a function of the gap between two 20-ms unresolved complex tone bursts. Open symbols show thresholds for conditions in which the gap was silent, and filled symbols for conditions in which the gap was filled with a bandpass noise. The upper and lower horizontal lines show performance for 20- and 40-ms single-burst tones, respectively. Error bars show standard errors.

Listeners made their responses using the numeric keypad on the computer keyboard. ''Lights'' were presented in a graphical display on the computer monitor to delineate the observation intervals and provide feedback.

D. Listeners

Five normally hearing listeners were employed, aged between 21 and 33 years. All of the listeners had prior experience of psychophysical tasks. Listeners were given at least 4 h training on the task before data collection began. There were no noticeable improvements in performance during the data collection. Listeners were tested individually in an IAC single-walled sound-attenuating booth.

E. Results and discussion

The individual and (geometric) mean results are presented in Figs. 2 and 3. Figure 2 shows the results for 20- and 40-ms single-burst conditions (horizontal lines) and the 2×20-ms gap conditions (circles). Open symbols show conditions with a silent gap and filled symbols show conditions in which a bandpass noise was presented in the gap so that the tone was perceived as being a single burst (this was indeed reported by listeners). Consider first the single-burst conditions. For all listeners, the $F0$ difference limen for the 40-ms stimulus was at least half that for the 20-ms stimulus, consistent with previous results (Plack and Carlyon, 1995; Plack and White, 2000; White and Plack, 1998). Also consistent with the earlier data are the results for the 2×20 gap conditions without the noise (open circles). Performance is generally superior to that for the 20-ms burst alone but substantially inferior to that for the single-burst 40-ms condition. The surprising new finding is the effect of the noise. For four out of the five listeners (LA, JF, CR, and CP), adding the noise in the gap produces a substantial improvement in performance, so that thresholds for the 2×20 gap conditions with noise are similar to the thresholds for the single-burst 40-ms condition. Listener NJ shows a slight effect of the noise for the 8-ms gap, but no effect for the 16-ms gap.

Figure 3 shows the results for the 40- and 80-ms single-burst conditions and the 2×40-ms gap conditions. For four of the listeners (LA, NJ, JF, and CP) the doubling of duration

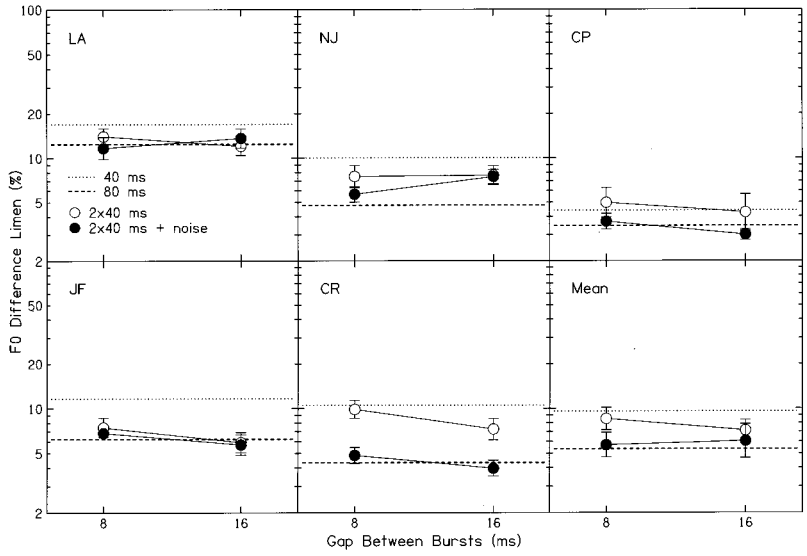


FIG. 3. As Fig. 2, except showing the results for two 40-ms bursts separated by a gap. The upper and lower horizontal lines show performance for 40- and 80-ms single-burst tones, respectively.

from 40 to 80 ms produces a much smaller effect than the doubling from 20 to 40 ms, perhaps suggesting incomplete integration of the 80-ms burst. The suspicion is reinforced by the similarity between the thresholds for the 2×40-ms gapped conditions and the 80-ms single-burst condition for listeners LA and JF. If it is assumed that the gap always produces resetting (and therefore multiple looks), then the result suggests that the improvement with duration between 40 and 80 ms (for the single-burst complex) is dominated by a multiple-looks-type mechanism for these listeners, with sampling windows less than 80 ms long. Listener CR may be regarded as the exception who proves the rule: The (large) performance improvement from 40 to 80 ms is similar to that from 20 to 40 ms; there is a deficit for the gapped 2×40-ms conditions compared to the 80-ms single-burst condition; and there is a large performance improvement on addition of the noise in the 2×40-ms conditions. For this listener, perhaps, long integration *is* being used for the 80-ms tone burst.

Despite the variability between listeners, it seems fair to summarize the data by stating that for each gapped condition the addition of noise produces performance similar to that for the single-burst condition with the same cumulative steady-state duration. A three-way repeated-measures analysis of variance (ANOVA) (stimulus duration$\times$noise$\times$gap) conducted on all the gapped conditions revealed a significant effect of duration [i.e., the 2×20 thresholds were higher than the 2×40 thresholds: $F(1,4)=74.3$, $p=0.001$] and a significant effect of adding the noise in the gap [$F(1,4)=10.4$, $p=0.032$]. There was no effect of gap duration. The effect of the noise is as predicted by the first hypothesis outlined in Sec. I A. The results suggest that in the absence of a level discontinuity, the auditory system may be able to integrate information across the gap. This might occur despite the fact that the temporal synchrony of the complex tone is lost during this period.

II. EXPERIMENT 2: EFFECT OF LEVEL CONTINUITY ON PITCH SHIFTS

A. Rationale

As described in the Introduction, Plack and White (2000) showed that applying a phase shift to the envelope of a 40-ms single-burst unresolved complex could produce a substantial shift in pitch. The largest effect was observed for an envelope delay of 0.75 periods (equivalent to a phase advance of $\pi/2$). In this case, the gap between successive envelope peaks across the phase discontinuity (the IPI) was 7 ms, as compared to the 4-ms period of the 250-Hz $F0$ complex. The pitch shift for this condition was about -8%. The presence of the pitch shift proves that the auditory system was integrating information across the phase discontinuity. The absence of the pitch shift when a gap of only 4 ms was introduced before the phase shift, separating the stimulus into two 20-ms bursts, was taken as evidence that the auditory system did not integrate information across the gap, consistent with the resetting hypothesis.

Now, if presenting a noise in the gap between the two tone bursts causes the bursts to be integrated together, then it is conceivable that an effect of the phase shift may be observed across the gap, even if the tone itself is interrupted. This finding would add further support to the idea that in the absence of a level discontinuity, the pitch mechanism conducts phase-sensitive long integration.

B. Stimuli

The stimuli were broadly similar to those in experiment 1, including the use of a low-pass noise to mask combination tones. The gap between the two 20-ms bursts was either 0 ms (single burst) or 8 ms. The (nominal) waveforms for the pregap and postgap bursts (before filtering) are given by

$$a(t)=A\sum_{n=1}^{\infty}[\sin(n(2\pi Ft+\theta))] \quad (0\leqslant t<0.02), \tag{1}$$

$$a(t)=A\sum_{n=1}^{\infty}[\sin(n(2\pi Ft+\theta+\varphi))] \quad ((G+0.02)\leqslant t<(G+0.04)), \tag{2}$$

where t is time in seconds, $a(t)$ is the waveform of the complex, A is the peak amplitude of each harmonic, n is the harmonic number, F is $F0$ in Hz, θ is the starting phase of the $F0$ component ($n=1$) of the first tone burst, and φ is the starting phase of the $F0$ component of the second tone burst relative to the first (the phase shift). G is the gap duration, in seconds.

Two phase-shift conditions were tested: ''no shift'' and ''$\pi/2$.'' In the no-shift reference condition, the phase of the complex was preserved across the gap. In this case, the starting phase of the first tone burst (θ) was π,[1] and the phase shift (φ) was 0. This condition was included to control for any bias in the matching procedure. In the $\pi/2$ shift condition, the starting phase of the first tone burst was $7\pi/4$. The phase shift of the second tone burst (φ) was $\pi/2$. This combination of phases produces a 7-ms IPI across the discontinuity in the 0-ms gap condition. (Because the phase shift occurred just before a pitch pulse, the effect of the phase advance was to effectively ''miss out'' one of the waveform peaks, producing a long IPI. This effect depends on the value of θ. For the no-shift condition, the value of θ was chosen so that the gap started after a pitch pulse and ended before a pitch pulse, rather than bisecting a pitch pulse in two.) In short, the effect of the phase shift is almost the same as applying a 3-ms delay (0.75 periods) to the second tone burst. The characteristics of the stimuli are described in more detail in Plack and White (2000).

For the 8-ms gap condition, stimuli were tested both with and without a bandpass noise presented in the gap. The characteristics of the bandpass noise were the same as in experiment 1.

C. Procedure

The pitch-matching procedure was the same as that used by Plack and White (2000), based on the procedure suggested by Jesteadt (1980). In each trial, two observation intervals (separated by 500 ms), each containing two tone bursts, were presented. The ''standard'' interval contained

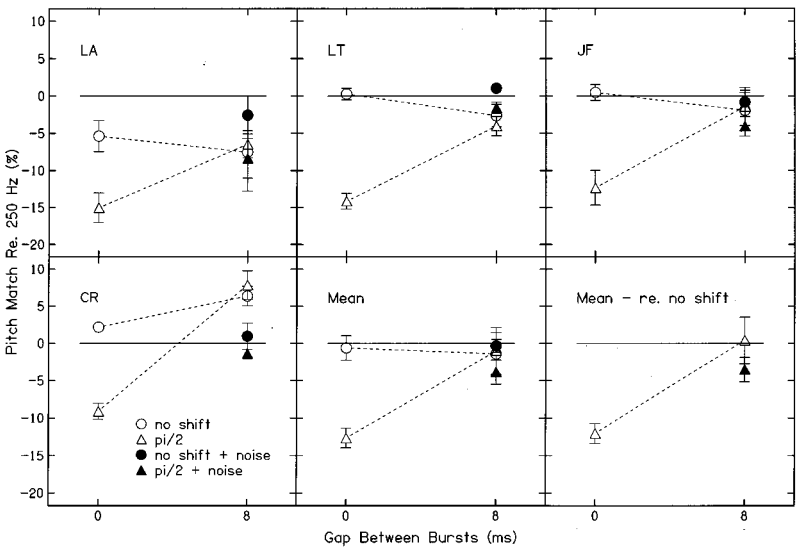


FIG. 4. The individual and mean pitch matches obtained in experiment 2 as a function of the gap between the tone bursts with phase shift as a parameter. The bottom right-hand panel shows mean pitch matches plotted relative to the ''no-shift'' reference conditions. Error bars show standard errors.

the phase-shifted condition with the $F0$ fixed at 250 Hz. The ''comparison'' interval contained a stimulus identical to that in the standard interval except that there was no phase shift between the bursts and the $F0$ was varied adaptively to obtain the pitch match. The standard and comparison intervals were presented in a random order on each trial. Two interleaved 2AFC adaptive tracks were used to bracket the pitch match. In one track the $F0$ of the preserved-phase condition started 20% below 250 Hz and in the other the $F0$ started 20% above. Each trial in a block was selected randomly from the two adaptive tracks. A ''two-up, one-down'' rule was applied to the lower track and a ''two-down, one-up'' rule to the upper track (Levitt, 1971). For each track, the step size was 4% for the first four reversals and 2% thereafter. Testing continued until 16 turnpoints had been completed for each track and the mean of the last 12 turnpoints from each track was calculated.

On each trial the listener was required to select which interval had the higher pitch. For the upper track, if the listener chose the comparison interval two times in a row, the $F0$ of the comparison interval was decreased (by the step size) for the next trial. Each time the listener chose the standard interval, the $F0$ of the comparison interval was increased (by the step size) for the next trial. Similarly for the lower track, if the listener chose the standard interval two times in a row, the $F0$ of the comparison interval was increased for the next trial. Each time the listener chose the comparison interval, the $F0$ of the comparison interval was decreased for the next trial.

The mean of the thresholds from the upper and lower tracks is an estimate of the pitch match. Half the difference between the thresholds from the upper and lower tracks is an estimate of the $F0$ discrimination threshold. The upper track converges, effectively, on the 70.7%-point on the psychometric function for detecting an increase in $F0$, and the lower track converges on the 29.3% point. The conditions were tested in a random order, within each experimental session. Each condition was repeated at least three times and the (arithmetic) mean of the repetitions was used to calculate the pitch matches.

Listeners made their responses using the numeric keypad on the computer keyboard. ''Lights'' were presented in a graphical display on the computer monitor to delineate the observation intervals. No feedback was given.

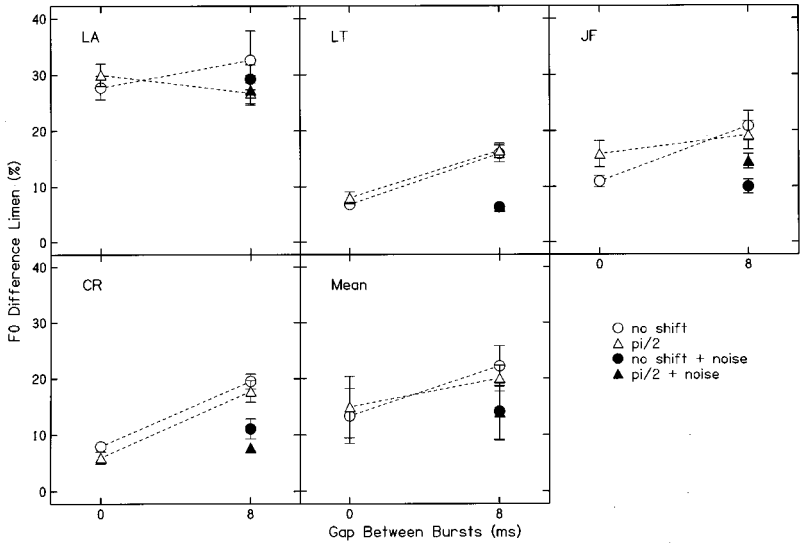


FIG. 5. The individual and mean $F0$ difference limens, derived from the tracking procedure in experiment 2. Error bars show standard errors.

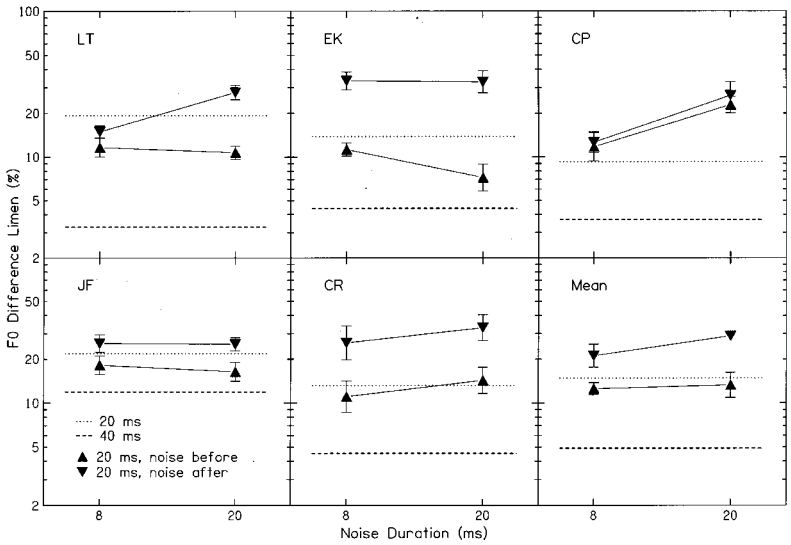


FIG. 6. The individual and mean results of experiment 3 showing discrimination performance for a 20-ms tone burst immediately preceded or followed by a bandpass noise, as a function of the duration of the noise. The upper and lower horizontal lines show performance for 20- and 40-ms single-burst tones, respectively. Error bars show standard errors.

D. Listeners

Four normally hearing listeners were employed. Listeners were given at least 4 h training on the task before data collection began.

E. Results and discussion

The individual and (arithmetic) mean results are presented in Figs. 4 and 5. Figure 4 shows the pitch matches and Fig. 5 shows the $F0$ discrimination thresholds. Consider the pitch-match data first. Circles show the matches for the no-shift reference conditions. For these conditions, the stimuli in the standard and the comparison intervals were essentially identical, although in each case the $F0$ of the comparison interval was allowed to vary to obtain the pitch match. The deviations of these values from zero indicate the bias of the listener and/or the procedure. As is evident from the mean data, there was little consistent bias across the listeners, although two of the listeners (LA and CR) did show some bias, albeit in opposite directions. The pattern of results for the $\pi/2$ phase-shift conditions is consistent across the four listeners, with a large pitch shift for the 0-ms gap, but very little effect for the 8-ms gap, at least in the conditions without the noise. The finding is consistent, qualitatively, with the results of Plack and White, although the magnitude of the downward pitch shift was larger in the present study. Moving to the conditions with the bandpass noise in the gap (filled symbols), there seems to be some effect of the phase shift across all listeners. In other words, the filled circles tend to be above the filled triangles. The effect can be seen in the bottom right-hand panel, where the mean pitch matches for the $\pi/2$ phase-shift conditions are plotted relative to the pitch matches for the no-shift conditions. For the 8-ms gap, the effect of the $\pi/2$ phase shift is negligible for the no-noise condition, but a downward shift of around -3.5% is seen for the noise condition. Although the effect of the phase shift for the 8-ms gap is much smaller than that for the 0-ms gap, the difference between the pitch match for the no shift reference and the pitch match for the $\pi/2$ phase shift is significantly greater when a noise is in the gap than when the gap is silent [$t=3.416$, $df=3$, $p=0.042$]. Adding the noise in the gap seems to make the auditory system sensitive to the relative phase across the gap.

Figure 5 shows the $F0$ difference limens derived from the matching procedure. With the exception of Listener LA, performance for the 0-ms gap (40-ms single-burst) condition is much better than that for the 8-ms gap condition. Also consistent with experiment 1 is the effect of adding the bandpass noise. In the mean data, performance for the 8-ms gap condition with the noise is similar to that for the 0-ms gap condition.

III. EXPERIMENT 3: EFFECT OF NOISE BEFORE AND AFTER TONE ON *F0* DIFFERENCE LIMENS

A. Rationale

One interpretation of the results of experiment 1 is that the addition of the bandpass noise in the gap removes the level discontinuity and allows the pitch analysis mechanism to integrate information from the two tone bursts. However, it is also conceivable that the noise may improve the discriminability of one (or both) of the tone bursts individually and thereby improve overall discrimination performance for the two-burst pair. For example, there is some physiological evidence that neural firing is more variable at stimulus onset (Relkin and Pelli, 1987; Teich and Khanna, 1985; Young and Barta, 1986). $F0$ discrimination performance may be poor for a 20-ms burst compared to a 40-ms burst because the neural representation of the first 20 ms of an unresolved complex tone is variable. Similarly, in the 2×20-ms condition, the gap might be sufficient to give each tone burst a separate onset response, leading to poor performance compared to the 40-ms single-burst condition (Plack and White, 2000). When the noise is added in the gap, the onset response for the second tone burst may be eliminated, so that the detectability of the last 20-ms of the stimulus is enhanced.

To control for this specific possibility, and to control for other possible benefits of the noise with respect to performance for a single tone burst, the effects of adding bandpass

noise both before and after a 20-ms burst were investigated. Noise durations of 8 and 20 ms were chosen, the former for comparison with the 8-ms gap conditions with noise in experiment 1, the latter to be equivalent to increasing tone duration from 20 to 40 ms. If, for example, the performance improvement from a 20- to a 40-ms uninterrupted tone is a result of the greater detectability of the part of the stimulus 20 ms after the onset response, then adding the noise before the 20-ms burst should produce a similar effect.

B. Stimuli and procedure

The basic stimuli were the same as in experiment 1 (see also Fig. 1). Single-burst tone durations of 20- and 40-ms were tested. In addition, conditions were tested in which the bandpass noise was presented immediately before or immediately after a 20-ms burst. The duration of the bandpass noise was either 8 or 20 ms.

The procedure was the same as that used in experiment 1. At least eight threshold estimates were collected for each condition.

C. Listeners

Five normally hearing listeners were employed. Listeners were given at least 4 h training on the task before data collection began.

D. Results and discussion

The individual and mean results are presented in Fig. 6. Horizontal lines show the results for 20- and 40-ms single-burst conditions. Upward-pointing triangles show results for the conditions in which the noise was presented immediately before the 20-ms tone, and downward-pointing arrows show results for the conditions in which the noise was presented immediately after the tone. The results are somewhat variable. As expected, all listeners show a substantial decrease in threshold as the duration of the complex is increased from 20 to 40 ms. For the noise presented before the 20-ms burst, three of the five listeners (LT, EK, and JF) show some performance improvement compared to the 20-ms burst alone, whereas CR and CP show no performance improvement on addition of the noise. However, as the mean data show, there is little consistent effect. There is also no consistent difference between the 8-ms and 20-ms noise durations.

For the noise presented after the tone burst, performance is generally worse than that in the absence of the noise. The results provide evidence of some effect of backward masking on $F0$ discrimination. Backward recognition masking (Massaro, 1975) has been demonstrated previously for pure-tone frequency discrimination (Massaro, 1975; Turner *et al.*, 1992). In suprathreshold tasks such as frequency discrimination and intensity discrimination, backward masking has been shown to be more effective than forward masking (Plack and Viemeister, 1992; Turner *et al.*, 1992).

Overall, while there is no convincing evidence that the effect of the noise on discrimination for a single tone burst can have accounted for the results of experiment 1, for some listeners presenting a noise before the 20-ms tone burst did seem to aid discrimination. It is possible, therefore, that the neural information in the first 8 or so ms of a tone is degraded by the onset response.

IV. GENERAL DISCUSSION

A. The importance of continuity

The results of the experiments support the hypothesis that the performance deficit produced by introducing a brief gap between two unresolved complex-tone bursts is a result of the level discontinuity produced by the gap. When the level discontinuity is eliminated by presenting a bandpass noise in the gap, performance is similar to that for the single-burst condition with the same cumulative duration. The results are inconsistent with the hypothesis that the performance deficit for the gap conditions is the result of a disruption of temporal synchrony, since the noise band does not contain any regular peaks and dips.

If the noise in the gap is being interpreted by the auditory system as an interruption that is added to an uninterrupted tone, as suggested by the auditory occlusion experiments described by Bregman (1990) and Warren (1999), then it seems reasonable for the auditory system to regard the two tone bursts as belonging to a single auditory object, and to treat them accordingly (i.e., long integration). In this case, a sudden drop in level may be the crucial evidence that one sound has terminated, and a sudden rise in level evidence for the onset of another sound. Any change that did not result in a decrease in level in the frequency region occupied by the tone might be interpreted by the system as an interruption by a separate sound source, through which the original tone may be assumed to have continued unaffected.

B. Integration and resetting

Although performance for the gapped conditions improves when a bandpass noise is presented in the gap, this result by itself does not prove that listeners were integrating $F0$ information across the gap. However, experiments 2 and 3 go some way towards demonstrating this claim. The results of experiment 2 suggest that in the absence of the noise the auditory system is insensitive to the relative envelope phase across the gap, but that when the noise is added there is some phase sensitivity across the gap (albeit less than that for a phase change that is not preceded by a period of silence). Taken at face value, the finding implies some sort of phase-sensitive long-integration mechanism.

The results of experiment 3 are more equivocal. Although there was no consistent performance improvement when the noise was added before a single 20-ms burst, some of the individual listeners did show a moderate performance advantage in these situations. It is possible that some of the performance improvement in experiment 1 caused by the addition of the noise burst in the gap could have arisen from the improved detectability of the second tone burst, perhaps because of the elimination of the neural onset response to this burst. However, the two listeners (CP and CR) who showed the largest effect of presenting the noise in the gap in experiment 1 showed little or no performance improvement when the noise was added before a single 20-ms burst in

experiment 3. It seems fairly safe to assume, therefore, that the effects of the noise in experiment 1 were not due entirely to an improvement in the detectability of the second tone burst alone.

If it is assumed that performance is better for the gap conditions with the noise than for those without because the noise allows integration between the two tone bursts, then this implies that the poor performance for the gap conditions without the noise is the result of the absence of, or reduction in, the amount of integration. In other words, the results may provide further support for the resetting hypothesis of White and Plack (1998).

V. CONCLUSIONS

(i) $F0$ discrimination performance for two 20-ms unresolved complex tone bursts separated by a short gap is much worse than performance for an uninterrupted 40- ms tone burst.

(ii) Presenting a noise, with the same overall spectral characteristics as the complex tone, in the gap between the bursts improves performance significantly.

(iii) The auditory system is insensitive to the relative envelope phase of two 20-ms bursts separated by a silent gap of 8 ms. However, if the bandpass noise is presented in the gap, some phase sensitivity is observed.

(iv) Discrimination performance for a single 20-ms burst may be improved in some listeners by the addition of a bandpass noise before the tone burst. However, the effect is insufficient to account for the results for the gapped noise conditions.

(v) The results are consistent with a pitch mechanism using a long integration time (possibly of the order of 40–80 ms) that is reset in response to level discontinuities.

ACKNOWLEDGMENTS

The research was supported by a grant from the Medical Research Council (UK). The first author was supported by a Royal Society University Research Fellowship. The authors thank Dr. Hedwig Gockel and Dr. Richard Warren for constructive comments on an earlier version of the article.

[1]Phase in this context is effectively the phase of the fundamental frequency component, since all the harmonics were added in sine phase relative to this.

Bregman, A. S. (**1990**). *Auditory Scene Analysis* (M.I.T. Press, Cambridge, MA).

Bregman, A. S., Ahad, P., Kim, J., and Melnerich, L. (**1994a**). ''Resetting the pitch-analysis system. I. Effects of rise times of tones in noise backgrounds or of harmonics in a complex tone,'' Percept. Psychophys. **56**, 155–162.

Bregman, A. S., Ahad, P. A., and Kim, J. (**1994b**). ''Resetting the pitch-analysis system. II. Role of sudden onsets and offsets in the perception of individual components in a cluster of overlapping tones,'' J. Acoust. Soc. Am. **96**, 2694–2703.

Carlyon, R. P. (**1997**). ''The effects of two temporal cues on pitch judgements,'' J. Acoust. Soc. Am. **102**, 1097–1105.

Cherry, C., and Wiley, R. (**1967**). ''Speech communications in very noisy environments,'' Nature (London) **214**, 1164.

Dirks, D., and Bower, D. (**1970**). ''Effects of forward and backward masking on speech intelligibility,'' J. Acoust. Soc. Am. **47**, 1003–1008.

Goldstein, J. L. (**1967**). ''Auditory nonlinearity,'' J. Acoust. Soc. Am. **41**, 676–689.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Hafter, E. R., and Buell, T. N. (**1990**). ''Restarting the adapted binaural system,'' J. Acoust. Soc. Am. **88**, 806–812.

Houtgast, T. (**1972**). ''Psychophysical evidence for lateral inhibition in hearing,'' J. Acoust. Soc. Am. **51**, 1885–1894.

Jesteadt, W. (**1980**). ''An adaptive procedure for subjective judgements,'' Percept. Psychophys. **28**, 85–88.

Levitt, H. (**1971**). ''Transformed up–down methods in psychophysics,'' J. Acoust. Soc. Am. **49**, 467–477.

Licklider, J. C. R. (**1951**). ''A duplex theory of pitch perception,'' Experientia **7**, 128–133.

Massaro, D. W. (**1975**). ''Backward recognition masking,'' J. Acoust. Soc. Am. **58**, 1059–1066.

Meddis, R., and Hewitt, M. (**1991**). ''Virtual pitch and phase sensitivity studied using a computer model of the auditory periphery: Pitch identification,'' J. Acoust. Soc. Am. **89**, 2866–2882.

Nabelek, I. V., Nabelek, A. K., and Hirsh, I. J. (**1973**). ''Pitch of sound bursts with continuous or discontinuous changes in frequency,'' J. Acoust. Soc. Am. **53**, 1305–1312.

Nabelek, I. V. (**1996**). ''Pitch of a sequence of two short tones and the critical pause duration,'' Acustica **82**, 531–539.

Plack, C. J., and Carlyon, R. P. (**1995**). ''Differences in fundamental frequency discrimination and frequency modulation detection between complex tones consisting of resolved and unresolved harmonics,'' J. Acoust. Soc. Am. **98**, 1355–1364.

Plack, C. J., and Viemeister, N. F. (**1992**). ''Intensity discrimination under backward masking,'' J. Acoust. Soc. Am. **92**, 3087–3101.

Plack, C. J., and White, L. J. (**2000**). ''Pitch matches between unresolved complex tones difering by a single inerpulse interval,'' J. Acoust. Soc. Am. **108**, 696–705.

Relkin, E. M., and Pelli, D. G. (**1987**). ''Probe tone thresholds in the auditory nerve measured by two-interval forced choice procedures,'' J. Acoust. Soc. Am. **82**, 1679–1691.

Teich, M. C., and Khanna, S. M. (**1985**). ''Pulse-number distribution for the neural spike train in the cat's auditory nerve,'' J. Acoust. Soc. Am. **77**, 1110–1128.

Turner, C. W., Zeng, F.-G., Relkin, E. M., and Horwitz, A. R. (**1992**). ''Frequency discrimination in forward and backward masking,'' J. Acoust. Soc. Am. **92**, 3102–3108.

Viemeister, N. F., and Plack, C. J. (**1993**). ''Time analysis,'' in *Human Psychophysics*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer, New York).

Viemeister, N. F., and Wakefield, G. H. (**1991**). ''Temporal integration and multiple looks,'' J. Acoust. Soc. Am. **90**, 858–865.

Warren, R. M. (**1999**). Auditory Perception: A New Analysis and Synthesis (Cambridge University Press, Cambridge, UK).

Warren, R. M., Obusek, C. J., and Ackroff, J. M. (**1972**). ''Auditory induction: Perceptual synthesis of absent sounds,'' Science **176**, 1149–1151.

White, L. J., and Plack, C. J. (**1998**). ''Temporal processing of the pitch of complex tones,'' J. Acoust. Soc. Am. **103**, 2051–2063.

Young, E. D., and Barta, P. E. (**1986**). ''Rate responses of auditory-nerve fibers to tones in noise near masked threshold,'' J. Acoust. Soc. Am. **79**, 426–442.

The lower limit of pitch as determined by rate discrimination

Katrin Krumbholz,[a] Roy D. Patterson, and Daniel Pressnitzer
Centre for the Neural Basis of Hearing, Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, United Kingdom

(Received 21 December 1999; revised 12 May 2000; accepted 6 June 2000)

This paper is concerned with the lower limit of pitch for complex, harmonic sounds, like the notes produced by low-pitched musical instruments. The lower limit of pitch is investigated by measuring rate discrimination thresholds for harmonic tones filtered into 1.2-kHz-wide bands with a lower cutoff frequency, F_c, ranging from 0.2 to 6.4 kHz. When F_c is below 1 kHz and the harmonics are in cosine phase, rate discrimination threshold exhibits a rapid, tenfold decrease as the repetition rate is increased from 16 to 64 Hz, and over this range, the perceptual quality of the stimuli changes from flutter to pitch. When F_c is increased above 1 kHz, the slope of the transition from high to low thresholds becomes shallower and occurs at progressively higher rates. A quantitative comparison of the cosine-phase thresholds with subjective estimates of the existence region of pitch from the literature shows that the transition in rate discrimination occurs at approximately the same rate as the lower limit of pitch. The rate discrimination experiment was then repeated with alternating-phase harmonic tones whose envelopes repeat at twice the repetition rate of the waveform. In this case, when F_c is below 1 kHz, the transition in rate discrimination is shifted downward by almost an octave relative to the transition in the cosine-phase thresholds. The results support the hypothesis that in the low-frequency region, the pitch limit is determined by a temporal mechanism, which analyzes time intervals between peaks in the neural activity pattern. It seems that temporal processing of pitch is limited to time intervals less than 33 ms, corresponding to a pitch limit of about 30 Hz. © *2000 Acoustical Society of America.* [S0001-4966(00)01809-9]

PACS numbers: 43.66.Ba, 43.66.Hg, 43.66.Nm [DWG]

INTRODUCTION

The topic of this paper is the lower limit of pitch for sounds composed of multiple, harmonic frequency components, like the voiced portions of speech and the tonal notes in music. In hearing research, the pitch associated with these complex sounds has been referred to as ''residue pitch'' (Schouten, 1938; Schouten *et al.*, 1962), ''repetition pitch'' (Thurlow and Small, 1955), or ''virtual pitch'' (Terhardt, 1974). In this paper, we will refer to it simply as ''pitch'' when there is no ambiguity, and residue pitch when it needs to be distinguished from the ''spectral pitch'' associated with an isolated, resolved harmonic. Ritsma (1962) reported that the lowest perceptible pitch for amplitude-modulated (AM) tones was in the region of 40 to 50 Hz. He used a subjective criterion for the existence of pitch, asking the listeners whether or not they heard a pitch at the repetition rate. Warren and Bashford (1981) measured the lower limit of pitch for broadband periodic sounds using a subjective criterion similar to Ritsma's, and reported a value of about 20 Hz, which is more than an octave below Ritsma's value. This discrepancy may be due to the fact that the residue pitch produced by AM tones, which consist of only three harmonic components, is weaker than the pitch of broadband periodic sounds consisting of many harmonics (Patterson, 1973; Preisler, 1993). The bottom note on the piano has a nominal pitch of 27.5 Hz, which lies between the two psychophysical estimates of the lower limit of pitch.

Musical instruments that produce low-pitched notes have the majority of their energy in low-frequency harmonics and harmonic tones with low-frequency components produce a strong, coherent pitch. In contrast, the pitch elicited by high-frequency harmonics is often weak and unmusical (Houtsma and Smurzynski, 1990), indicating that the lower limit of pitch depends on the frequency content of the stimuli. The purpose of the present paper was to measure the lower limit of pitch for multiharmonic sounds and to investigate its dependence on frequency region using a rate discrimination procedure. In a companion study by Pressnitzer *et al.* (1999), we measured the lower limit of pitch with a melody task.

A. The lower limit of pitch as a function of frequency region

Ritsma (1962) showed that the lower limit of pitch for AM tones increases markedly with carrier frequency in the region above about 1 kHz. Figure 1(a) shows the average value of the lowest repetition rate, R_{rep}, for which Ritsma's three listeners reported perceiving a residue pitch as a function of the carrier frequency, F_c, of the AM tones. He expressed his data in terms of the harmonic number, $N = F_c / R_{\text{rep}}$, associated with the carrier frequency as shown by the solid line in Fig. 1(b). Ritsma (1962) originally interpreted his data in terms of a temporal pitch mechanism similar to Licklider's (1951) neural delay-and-coincidence network. In temporal pitch models, it is assumed that the auditory system measures the time intervals between peaks in the temporal responses of the cochlear filters and that the pitch corresponds to the most common time interval (Lyon, 1984; Meddis and Hewitt, 1991; Patterson *et al.*, 1995; Moore, 1997). The lower limit of pitch is determined by the

[a]Author to whom correspondence should be addressed. Electronic mail: katrin.krumbholz@mrc-cbu.cam.ac.uk

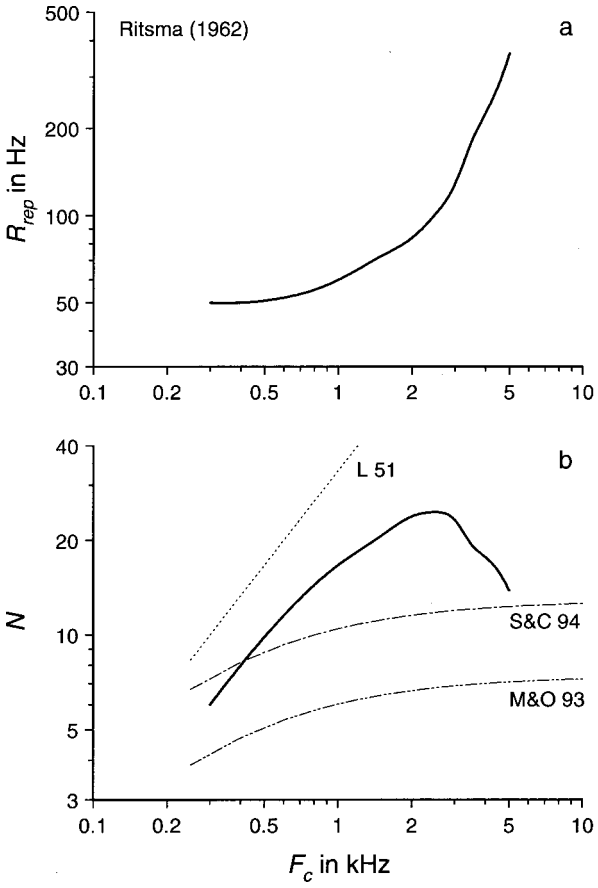


FIG. 1. Lower limit of pitch for AM tones, replotted from Ritsma (1962). (a) the average lower limit of pitch, R_{rep}, for Ritsma's three listeners as a function of the carrier frequency, F_c, of the AM tones; (b) the same data expressed in terms of the harmonic number, N, associated with F_c. The dotted line in panel (b) shows the harmonic number corresponding to a constant repetition rate of 30 Hz. The dash-dotted and the dash-dot-dotted lines show two definitions of the limit of spectral resolution proposed by Shackleton and Carlyon (1994) and by Moore and Ohgushi (1993), respectively.

longest time interval that is accurately represented in the system, which Licklider suggested is around 33 ms (corresponding to 30 Hz). The pitch limit is expected to be approximately constant, at least in the low-frequency region, where the auditory nerve encodes the temporal structure of the stimulus with high precision. The dotted line in Fig. 1(b) (labeled ''L 51'') represents the harmonic number corresponding to a constant repetition rate of 30 Hz (a time interval of 33 ms). For F_c below 1 kHz, the line is approximately parallel to Ritsma's data; however, in the region between 1 and 2 kHz, the prediction and the data diverge, and above 3 kHz, the prediction and the data are completely irreconcilable.

Subsequent to Ritsma's (1962) experiments on the lower limit of pitch, studies of spectral dominance in pitch perception (Ritsma, 1967; Plomp, 1967; Patterson and Wightman, 1976) showed that the low-frequency harmonics are most important for mediating pitch. Their results led to the formulation of several spectral models of pitch (Goldstein, 1973; Wightman, 1973; Terhardt, 1979; Cohen *et al.*, 1995), which assume that residue pitch is mediated by resolved peaks in the distribution of activity along the cochlear partition, and that the lower limit of pitch is determined by a limit on the spectral resolution imposed by cochlear filtering. The width of the cochlear filter is roughly proportional to filter center frequency in the region above about 1 kHz, and so the lower limit of pitch would be expected to occur at a roughly constant value of N in this region. Plomp (1964) and Plomp and Mimpen (1968) reported that listeners can hear out the first six components of a harmonic tone in the region of 1 kHz, and that the same is true for inharmonically shifted complex tones. Moore and Ohgushi (1993) measured the audibility of individual components in complex tones whose component frequencies were equally spaced on an equivalent rectangular bandwidth (ERB)[1] scale and concluded that individual components can be heard out when they are separated by 1.25 ERBs, which is consistent with the results from the two previous studies (Moore, 1997). The dash-dot-dotted line in Fig. 1(b) (labeled ''M&O 93'') shows the number of the highest harmonic in a complex tone which can be resolved at a given frequency, F_c, according to this definition. Harmonics are presumed to be resolved in the region below this line, and unresolved above it. Clearly, Ritsma's (1962) lower limit of pitch values all fall above the line, in the region where components would be expected to be unresolved. Thus, if resolved harmonics are required for the perception of a pitch, then this definition of resolution is at odds with Ritsma's data.

Shackleton and Carlyon (1994) compared data from pitch-matching experiments, rate discrimination experiments, and frequency-modulation experiments, and concluded that Moore and Ohgushi's criterion was somewhat conservative. Shackleton and Carlyon argued that, in the region of 1 kHz, we have to assume that harmonics up to the eleventh can mediate enough spectral information to create a pitch percept even if they are not separately audible. This more lenient definition of the spectral resolution of harmonics is presented by the dash-dotted line in Fig. 1(b) (labeled ''S&C 94''). The majority of Ritsma's data are still clearly in the ''unresolved'' region above this line, and, in a purely spectral model, it is not clear what the basis of residue pitch could be in this region.

B. Rate discrimination thresholds

Ritsma and Hoekstra (1974) proposed to determine the lower limit of pitch by means of rate discrimination measurements (see also Hoekstra, 1979). They measured rate discrimination threshold, RDT, for filtered click trains with 1/3-octave bandwidths using filter frequencies, F_c, between 1 and 8 kHz. Figure 2(a) shows their data as a function of the repetition rate, R_{rep}, of the click trains; different filter frequencies are denoted by different symbols [see the legend in Fig. 2(b)] Ritsma and Hoekstra described the discrimination functions as consisting of two plateaus with a frequency-dependent transition between them. For the upper plateau, discrimination performance was roughly constant at about 2% of R_{rep}, whereas for the lower plateau, discrimination was consistent with frequency discrimination thresholds for pure tones in the respective filter passband. Ritsma and Hoekstra's RDT data are plotted as a function of the harmonic number, N, associated with the filter frequency, F_c, in

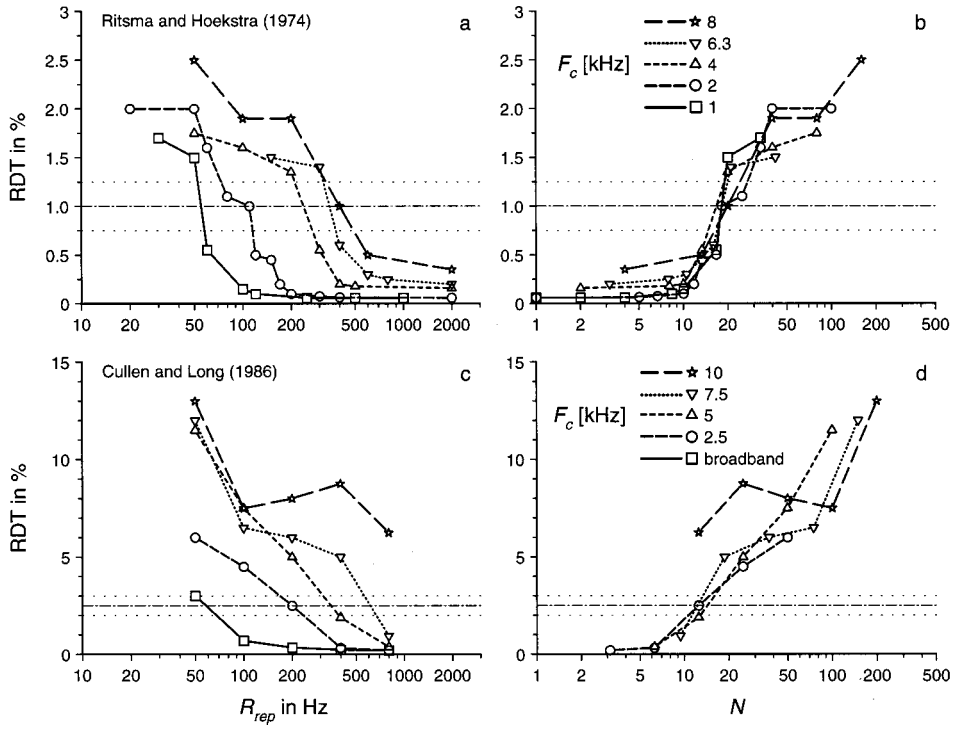


FIG. 2. Rate discrimination threshold, RDT, for bandpass-filtered click trains as a function of filter cutoff frequency, F_c; upper panels, replotted from Ritsma and Hoekstra (1974); lower panels replotted from Cullen and Long (1986). RDT is expressed as a percentage of repetition rate, R_{rep}, and is plotted as a function of R_{rep} (a) and (c), and as a function of the harmonic number, N, associated with the filter cutoff frequency (b) and (d). The parameter is filter cutoff frequency, F_c, [see the legend in panels (b) and (d)]. The dotted and dash-dotted lines highlight the transition range around a criterion level of performance [0.5% range around a 1% criterion in panels (a) and (b), 1% range around a 2.5% criterion in panels (c) and (d)].

Fig. 2(b), and they show that the transition in the discrimination functions occurs at a roughly constant value of N for all filter frequencies. Ritsma and Hoekstra concluded that the transition reflects the limit of spectral resolution in the cochlea. They pointed out that the rate at which the transition occurs in the RDT functions for filter frequencies of 1, 2, and 4 kHz is approximately the same as Ritsma's (1962) lower limit of pitch in this frequency region. In particular, Ritsma's lower limit of pitch values corresponds to a performance level of about 1% in Ritsma and Hoekstra's rate discrimination functions for these three filter frequencies. This led Ritsma and Hoekstra to postulate that ''the possibility of frequency analysis is a necessary condition for tonality of a residue constituted of a bandpass-filtered click train.'' Note, however, that the correspondence between the transition in the RDT functions and the lower limit of pitch only holds for the frequency region between 1 and 4 kHz, where Ritsma's values exhibit a broad maximum when expressed in terms of N [Fig. 1(b)]. Ritsma reported that residue pitch disappears completely above this region; in contrast, Ritsma and Hoekstra's RDT functions continue to exhibit a transition at the same value of N up to 8 kHz [see Fig. 2(b)]. Below 1 kHz, Ritsma's lower limit of pitch (expressed in terms of N) decreases sharply with decreasing carrier frequency [Fig. 1(b)] which is incompatible with Ritsma and Hoekstra's conclusion. Unfortunately, Ritsma and Hoekstra did not include filter frequencies below 1 kHz.

Cullen and Long (1986) extended Ritsma and Hoekstra's results, but only to higher filter frequencies. They used highpass-filtered click trains with cutoff frequencies, F_c, between 2.5 and 10 kHz. The lower panels of Fig. 2 [2(c) and 2(d)] show their data plotted in the same format as Ritsma and Hoekstra's in Figs. 2(a) and 2(b). Cullen and Long's data exhibit a strikingly similar pattern to Ritsma and Hoekstra's; when plotted as a function of harmonic number N [Fig. 2(d)] the transition in the discrimination functions occurs at a roughly constant value of N. However, there is a large discrepancy between the overall level of performance in the two studies which is difficult to explain. Note that the scale on the ordinate for Cullen and Long's data (lower panels) is a factor of 5 greater than that for Ritsma and Hoekstra's data (upper panels). Similar to Ritsma and Hoekstra, Cullen and Long associated large RDT values with temporal processing, and low RDT values with spectral processing. Shackleton and Carlyon (1994) provided additional support for this interpretation by comparing measures of rate discrimination with measures of pitch perception obtained in pitch-matching experiments with harmonic tones that produce conflicting temporal and spectral pitch cues. They showed that in the frequency region between 1.4 and 1.9 kHz (their MID region), the transition in the rate discrimination function coincides with a change in the octave of the pitch of these stimuli.

Cullen and Long did include a broadband condition which covers the frequency region below 1 kHz.[2] However, their lowest repetition rate in this condition (50 Hz) was above the lower limit of pitch for broadband stimuli (Warren and Bashford, 1981). Similarly, Shackleton and Carlyon (1994) examined three frequency regions (a LOW, MID, and HIGH region), but, in the low region (which was below 1 kHz) they only employed repetition rates above the lower limit of pitch. Houtsma and Smurzynski (1990) established a direct relation between rate discrimination thresholds and pitch salience; they used only one repetition rate, 200 Hz, for which the pitch limit occurs in the region of 2.5 kHz.

The current study investigates rate discrimination threshold for bandpass-filtered harmonic tones with repetition rates in and below the pitch region. Our aim was to specify the lower limit of pitch in the frequency region occupied by low-pitched musical instruments (below 1 kHz)

and to determine whether it supports temporal or spectral models of pitch. The pitch of AM tones is relatively weak and when the lower component is at the fourth harmonic or lower, the spectral pitch of the components obscures residue pitch. Patterson (1973) showed that residue pitch can be strengthened by adding high harmonics. Provided there are six or more adjacent components, residue pitch can be measured for all harmonic numbers, so in the current study we use multiharmonic tones. In experiment 1, RDT functions were measured for cosine-phase (CPH) harmonic tones whose temporal waveforms resemble those of the filtered click trains used by Ritsma and Hoekstra (1974) and Cullen and Long (1986). Experiment 2 measured RDT functions for alternating-phase (APH) harmonic tones whose temporal waveform and frequency spectrum produce conflicting pitch cues, which enables us to discriminate between temporal and spectral models of pitch (Patterson, 1987; Shackleton and Carlyon, 1994). The transition points in the RDT functions are compared with values of the lower limit of melodic pitch obtained with a melody task (Pressnitzer *et al.*, 1999) using the same bandpass-filtered harmonic tones as the present study.

I. EXPERIMENT 1: RATE DISCRIMINATION WITH CPH TONES

Rate discrimination thresholds were measured for harmonic tones whose components were added in cosine phase which produces a waveform with one sharp envelope peak per period. The solid line in the upper panel of Fig. 3 presents an example of the temporal waveform of a CPH tone used in this experiment; the dashed line shows its envelope.[3]

A. Method

1. Stimuli

Harmonic tones were filtered into frequency bands with six different lower-cutoff frequencies (F_c=0.2, 0.4, 0.8, 1.6, 3.2, or 6.4 kHz). All six bands had an equivalent rectangular bandwidth of 1.2 kHz; their edges were rounded with a quarter cycle of a cosine function (see the dashed line in lower panel of Fig. 3). The lower and the upper spectral ramps (dF_1 and dF_2 in Fig. 3) had widths of 0.2 and 1.0 kHz, respectively. The flat portion, bw, was 0.6 kHz wide. These values were chosen in order to avoid the perception of edge tones (see Kohlrausch and Houtsma, 1992) which can become particularly distracting in a rate discrimination task when the pitch associated with the repetition rate is weak, as in the high-frequency conditions. Rate discrimination thresholds were measured for different subsets of a total of seven repetition rates (R_{rep}=16, 32, 64, 128, 192, 256, and 320 Hz). The duration of the stimuli was selected at random from values between 800 and 1000 ms so that neither the duration nor the number of clicks could be used as a discrimination cue. The stimuli were gated on and off with 35-ms cosine-squared ramps. They were presented at an overall level of 55 dB SPL.

Before each run, a programmable filter (TDT PF1) was loaded with the filter weights corresponding to the current filter condition. In each two-interval trial, broadband harmonic tones with the appropriate repetition rates were computed digitally with a sampling rate of 25 kHz (TDT AP2), low-pass filtered at 10 kHz (TDT FT6-2), and bandpass filtered with the programmable filter. The filtered stimuli were attenuated (TDT PA4) and fed to an analog signal mixer (TDT SM3) where they were combined with a continuous low-pass noise masker designed to mask distortion products with frequencies below the stimulus passband. The masker was low-pass filtered (Kemo VBF 8) from a pink noise which had been recorded on DAT tape. It was attenuated (TDT PA4) and fed to the signal mixer (TDT SM3). The

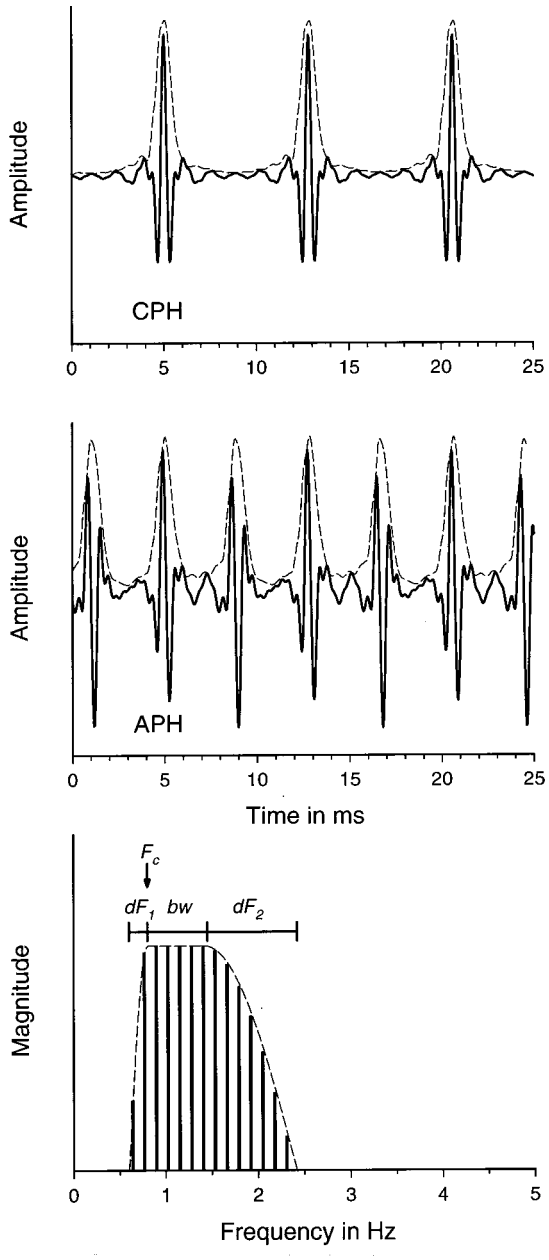


FIG. 3. Upper and middle panels: temporal waveforms (solid lines) and envelopes (see footnote 3) (dashed lines) of a typical pair of CPH and APH harmonic tones from the current experiments. In this example, the repetition rate of the stimuli is 128 Hz and the lower cutoff frequency, F_c, is 0.8 kHz (filter condition 3). The stimuli were filtered into frequency bands (dashed line in lower panel) with a lower cutoff frequency, F_c, between 0.2 and 6.4 kHz. The lower spectral ramp, dF_1, was 0.2 kHz wide, the flat portion, bw, was 0.6 kHz wide, and the upper ramp, dF_2, was 1 kHz wide. The solid lines show the magnitudes of the Fourier components of both the CPH and APH tones in arbitrary, linear units.

unfiltered pink noise had a level of 32 dB SPL in a 1/3-octave band around 1 kHz. On average, the sensation level of the unfiltered noise amounted to about 30 dB. The low-pass filter had a slope of 96 dB/octave and a cutoff frequency which was half an octave below the lower-cutoff frequency of the current filter condition (F_c). The stimuli were finally fed to a headphone buffer (TDT HB6) and presented diotically via headphones (AKG K 240 DF) to the listener, who was seated in a double-walled, sound-attenuated booth. The headphones had been calibrated using a B&K $\frac{1}{2}$-in. microphone (type 4134), artificial ear (type 4153), and measuring amplifier (type 2610).

2. Procedure

Rate discrimination thresholds were measured using an adaptive two-alternative, forced-choice procedure. On each trial, two harmonic tones were presented with a silent gap of 700 ms. The repetition rates of the stimuli differed by a percentage of their geometric mean, ΔR_{rep}. The mean rate was randomly selected from a uniform probability distribution with a width of 10% around the nominal mean rate. Either the first or the second stimulus had the higher repetition rate. The listener had to choose the stimulus with the higher rate by pressing one of two buttons on a response box (TDT PI2). Feedback was presented at the end of each trial. After three consecutive correct responses, ΔR_{rep} was reduced by a factor ν and after each incorrect answer, ΔR_{rep} was increased by the same factor, tracking the ΔR_{rep} that yields 79%-correct responses (Levitt, 1971). Up to the first reversal, ν was equal to 1.5; it was reduced to 1.3 for the following two reversals and then set to 1.19 for the rest of the ten reversals that made up a threshold run. Each threshold estimate is the mean of the ΔR_{rep}'s of the last six reversals. The data points presented in the figures are the mean of at least three thresholds and the error bars show their standard error.

3. Listeners

A total of four listeners participated in experiments 1 and 2, author KK and three other listeners who were paid for their services at an hourly rate. RH and KK were listeners in both experiments; RT and FH participated only in experiment 1 and 2, respectively. All listeners had normal hearing over the whole audible frequency range and were between 23 and 35 years old.

B. Results

The three panels of Fig. 4 show the RDT functions of three listeners for CPH tones in the six different filter conditions. RDT is expressed as a percentage of the repetition rate, R_{rep} (abscissa). Thresholds for different filter conditions are denoted by different open symbols (see the legend in the upper panel). The pattern of the discrimination functions in Fig. 4 is similar for the three listeners. In the three lower filter conditions ($F_c = 0.2$, 0.4, and 0.8 kHz), discrimination threshold decreases sharply from a value between 5% and 10% (depending on the listener) at a repetition rate of 16 Hz to about 1% at 64 Hz. This transition from high to low discrimination threshold corresponds to a change in the perceptual quality of the stimuli. Listeners reported that at 16 Hz, periodicity was perceived as a fast tapping rhythm; at 64 Hz, a clear pitch was perceived; at 32 Hz, the sensation was a mixture of pitch and flutter and the proportion was different for different listeners. In the higher filter conditions ($F_c = 1.6$, 3.2, and 6.4 Hz), the discrimination functions are shallower and the transition from high to low discrimination threshold occurs at progressively higher rates. At 16 Hz, listeners still perceived a fast tapping. As repetition rate was increased, the perception developed a pitch which was weaker and less coherent than in the lower filter conditions. The pitch produced by high-frequency harmonics has been described as ''rattle'' pitch (Plomp, 1976, p. 129; Kaernbach and Demany, 1998). In the highest filter condition ($F_c = 6.4$ kHz), the discrimination functions of RH and KK start near their asymptotic values and do not exhibit any signifi-

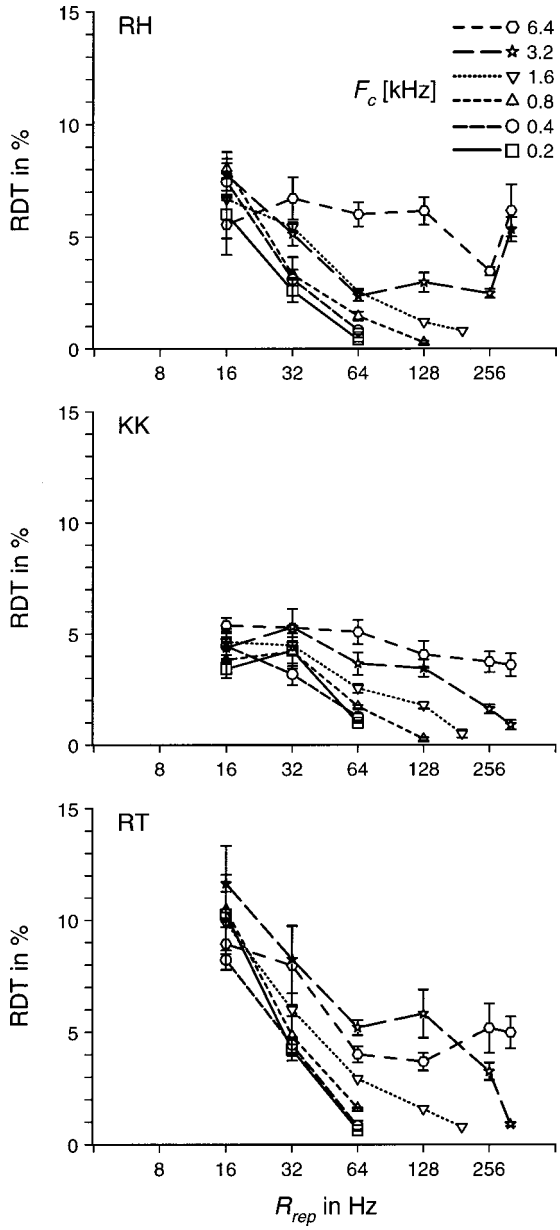


FIG. 4. Individual RDT functions for CPH tones. Discrimination threshold is expressed as a percentage of repetition rate, R_{rep}. The parameter is filter cutoff frequency, F_c (see the legend in the upper panel).

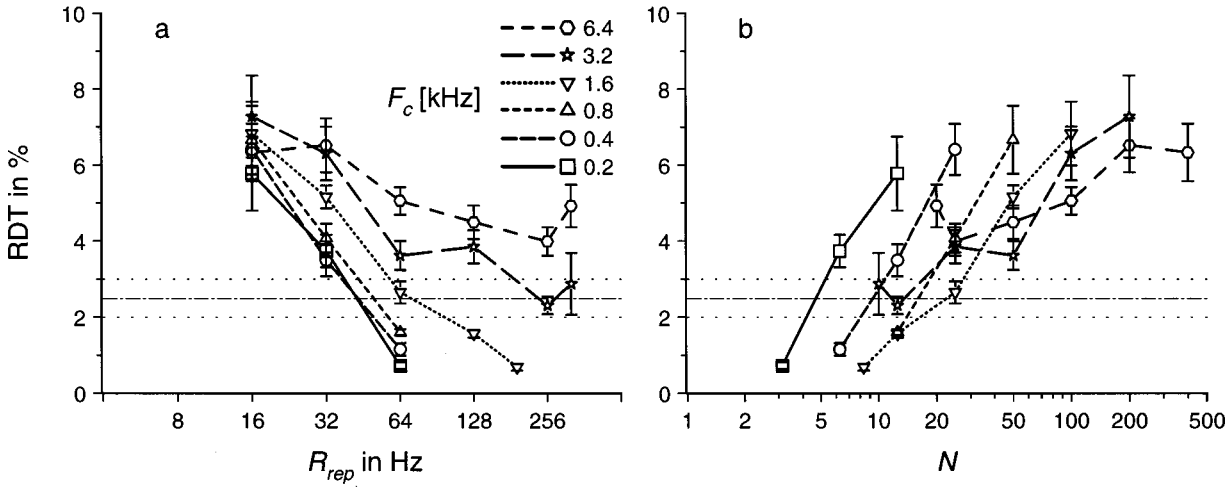


FIG. 5. Average RDT functions with CPH stimuli for listeners RH, KK, and RT, plotted as a function of (a) repetition rate, R_{rep}, and (b) harmonic number, N, associated with the cutoff frequency, F_c. The dotted and dash-dotted lines highlight the 1% range around the 2.5% transition criterion.

cant decrease within the range of repetition rates tested; RT's discrimination function has a very high starting value and decreases markedly between 16 and 64 Hz. In all filter conditions, RT had the most difficulty discriminating at low repetition rates.

The range of repetition rates where threshold could be measured was limited by stimulus bandwidth. Stimuli with a high repetition rate contain only a few harmonics whose amplitudes can vary strongly when the rate is changed and components move along the filter skirts. In this case, the dominant pitch was sometimes not the residue pitch but the pitch associated with an individual harmonic. The large amplitude changes meant that the dominant harmonic could change between the two intervals in one trial with the result that the dominant pitch could move in the opposite direction to the residue pitch. This pitch ambiguity affected RH's discrimination thresholds at the highest repetition rate (320 Hz); KK and RT were still able to attend to the residue pitch at this rate.

C. Discussion

Figure 5(a) shows the average discrimination thresholds for the three listeners [note that the range of the ordinate in Fig. 5(a) is smaller than in Fig. 4]. In Fig. 5(b), the average thresholds are plotted as a function of the harmonic number, N, associated with the cutoff frequency, F_c. This transformation aligned the discrimination functions of Ritsma and Hoekstra (1974) and of Cullen and Long (1986), which were obtained with filter frequencies above 1 kHz [Figs. 2(b) and 2(d)]. Figure 5(b) shows that the discrimination functions from the present study are not aligned by this transformation, particularly when F_c is below 1 kHz. There is a progressive shift between the discrimination functions for the first four filter conditions ($F_c = 0.2$, 0.4, 0.8, and 1.6 kHz) indicating that the transitions do not occur at a constant harmonic number, N, in this region. Indeed, the transition occurs at a roughly constant value of repetition rate. If the transition point is designated as the repetition rate at which the discrimination function crosses the 2.5% level of performance [horizontal dashed-dotted line in Figs. 5(a) and 5(b)], then the transition occurs at about 45, 46, and 52 Hz in the first three filter conditions ($F_c = 0.2$, 0.4, and 0.8 kHz), respectively. In the next two conditions ($F_c = 1.6$ and 3.2 kHz), the transition point rises to 73 and 240 Hz, respectively. In the sixth filter condition ($F_c = 6.4$ kHz), none of the individual discrimination functions crosses the 2.5% criterion within the tested range. The filled triangles in Fig. 6 show the transition rates for CPH tones as a function of filter cutoff frequency, F_c. The vertical bars delimit the range of repetition rates within a 1% range around the 2.5% criterion [horizontal dotted lines in Figs. 5(a) and 5(b)]; they reflect the steepness of the discrimination functions in the transition region. The filled squares are the transition rates for the APH tones which will be discussed in Sec. II B.

In Fig. 7, the transition rates for CPH tones are expressed in terms of the harmonic number, N, associated with the filter cutoff frequency, F_c (filled triangles). The open symbols show the transition rates from previous studies of rate discrimination. For all studies except Ritsma and Hoekstra's (open diamonds), the criterion and the range were set

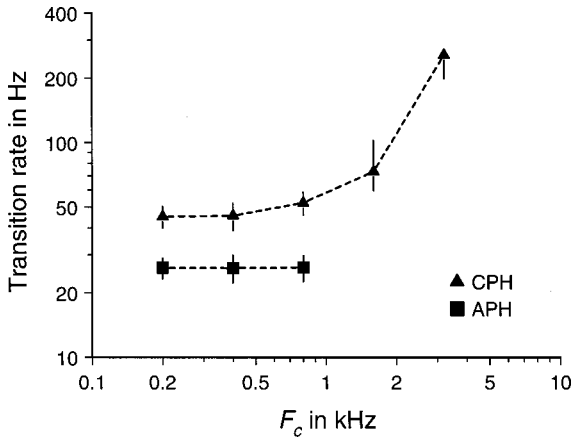


FIG. 6. The repetition rate at which the average RDT functions for CPH (filled triangles) and APH tones (filled squares) reach the 2.5% transition criterion. The vertical bars delimit the range of repetition rates within a 1% range around the 2.5% criterion. The transition rates are plotted as a function of filter cutoff frequency, F_c.

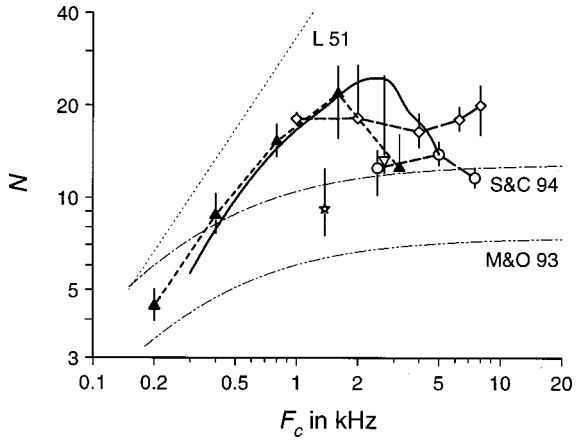


FIG. 7. Transition rates for the RDT functions from the current and previous studies expressed in terms of the harmonic number, *N*, associated with the frequency, F_c, which specifies the filter condition. The results are from Ritsma and Hoekstra (1974, open diamonds), Cullen and Long (1986, open circles), Houtsma and Smurzynski (1990, open, inverted triangle), Shackleton and Carlyon (1994, open star) and the CPH data from the current study (filled triangles). The criterion and the range were 2.5% and 1% for all but Ritsma and Hoekstra's data, for which they were 1% and 0.5%, respectively. The solid line shows Ritsma's (1962) lower limit of pitch for AM tones [reproduced from Fig. 1(b)]. The dotted line shows the harmonic number corresponding to a constant repetition rate of 30 Hz. The dash-dotted and the dash-dot-dotted lines show the two definitions of the limit of spectral resolution proposed by Shackleton and Carlyon (1994) and by Moore and Ohgushi (1993), respectively.

to 2.5% and 1%, respectively, as for the data from the current study. The upper plateau in Ritsma and Hoekstra's discrimination functions is below 2.5% for most filter frequencies [see Fig. 2(a)], and so we set the criterion to 1% with a range of 0.5%, which is approximately in the middle of the transition region in their discrimination functions [see the dash-dotted and dotted lines in Figs. 2(a) and 2(b)]. It is not clear why Ritsma and Hoekstra's discrimination thresholds are so much lower than in the other studies. The solid line in Fig. 7 shows the lower limit of pitch from Ritsma (1962) [reproduced from Fig. 1(b)]. The data from the current study (filled triangles) show that, in the frequency region below 1 kHz the transition rate, expressed in terms of *N*, increases dramatically as F_c increases. In this frequency region, the function relating transition rate to cutoff frequency closely resembles Ritsma's lower limit of pitch (solid line). Like Ritsma's pitch limit, the transition rates correspond to a constant repetition rate (dotted line, labeled ''L 51'') rather than a constant fraction of cochlear filter bandwidth (dash-dotted and dash-dot-dotted lines, labeled ''S&C 94'' and ''M&O 93'') for filter frequencies, F_c, below 1 kHz. This indicates that the transitions in the discrimination functions for the lower filter conditions are more likely determined by a limitation in a temporal pitch mechanism than by a limit in the spectral resolution of harmonics.

In the frequency region above 1 kHz, Cullen and Long's data (open circles) and those of Houtsma and Smurzynski (inverted, open triangle) show that the rate at which the transition occurs corresponds to a roughly constant value of *N* close to 14. The transition rates from Ritsma and Hoekstra's data (open diamonds) also correspond to a roughly constant harmonic number, but the average value in this case is about 19, and it is consistently above the values from the two other studies. In the frequency region above 1 kHz, the transition rates from the current study are in reasonable agreement with the values from the previous studies. The transition in the fourth filter condition ($F_c = 1.6$ kHz) occurs at a somewhat higher *N* than in the third and fifth filter conditions ($F_c = 0.8$ and 3.2 kHz), suggesting that *N* attains a maximum around 1.6 kHz. However, the discrimination functions for the higher filter conditions are shallow, so the absolute value of the transition rate is highly dependent on the precise value of the transition criterion. The transition in Shackleton and Carlyon's discrimination function for their MID frequency region (open star) occurs at $N \approx 9$, which is well below the values from the other studies. This discrepancy might be because they measured the discrimination function with only three repetition rates, or it may be due to the slow frequency modulation which was applied to their stimuli.[4]

In each of the previous studies reporting rate discrimination data in the frequency region above 1 kHz, the authors conclude that the transition in the RDT functions occurs at a roughly constant harmonic number, and it is hypothesized that this represents the limit of the spectral resolution of harmonics. In each case, the conclusion and hypothesis seem reasonable within the context of that specific study. It is important to note, however, that the different studies produce values of the ''constant,'' *N*, which range from 9 to 20 for reasons that are not explained. It is also the case that more than half of the transition rates in Fig. 7 fall above the dash-dotted line (labeled ''S&C 94'') which means that they are in the region where the harmonics would be unresolved, according to Shackleton and Carlyon's (1994) definition of the limit of spectral resolution. With the more traditional definition of the limit of spectral resolution based on the results of Plomp (1964), Plomp and Mimpen (1968), and Moore and Ohgushi (1993) (dash-dot-dotted line labeled ''M&O 93''), all of the transition rates in Fig. 7 would be in the unresolved region.

II. EXPERIMENT 2: RATE DISCRIMINATION WITH APH TONES

Experiment 2 measures rate discrimination threshold for APH tones as a function of repetition rate and filter condition. The results are then compared with the respective discrimination thresholds for CPH tones. The middle panel in Fig. 3 shows the waveform of an APH tone with the same frequency spectrum (lower panel of Fig. 3) as the CPH tone shown in the top panel of Fig. 3. The dashed lines in the upper and middle panels of Fig. 3 show the envelopes of the CPH and APH waveforms. Whereas the CPH envelope has only one peak per repetition period, the APH stimulus has another peak in the middle of each period. The envelope of the APH tone actually repeats at exactly twice the rate of the waveform, but the fine structure of the peaks in the middle of the period is phase shifted about 90° relative to the fine structure of the other peaks. Patterson (1987) measured the phase shift required in APH tones to discriminate them from CPH tones and explained the discrimination with a temporal model of pitch in terms of the extra peak in the APH waves. Similarly, Meddis and O'Mard (1997) showed that simulated neural activity patterns for spectrally unresolved APH tones

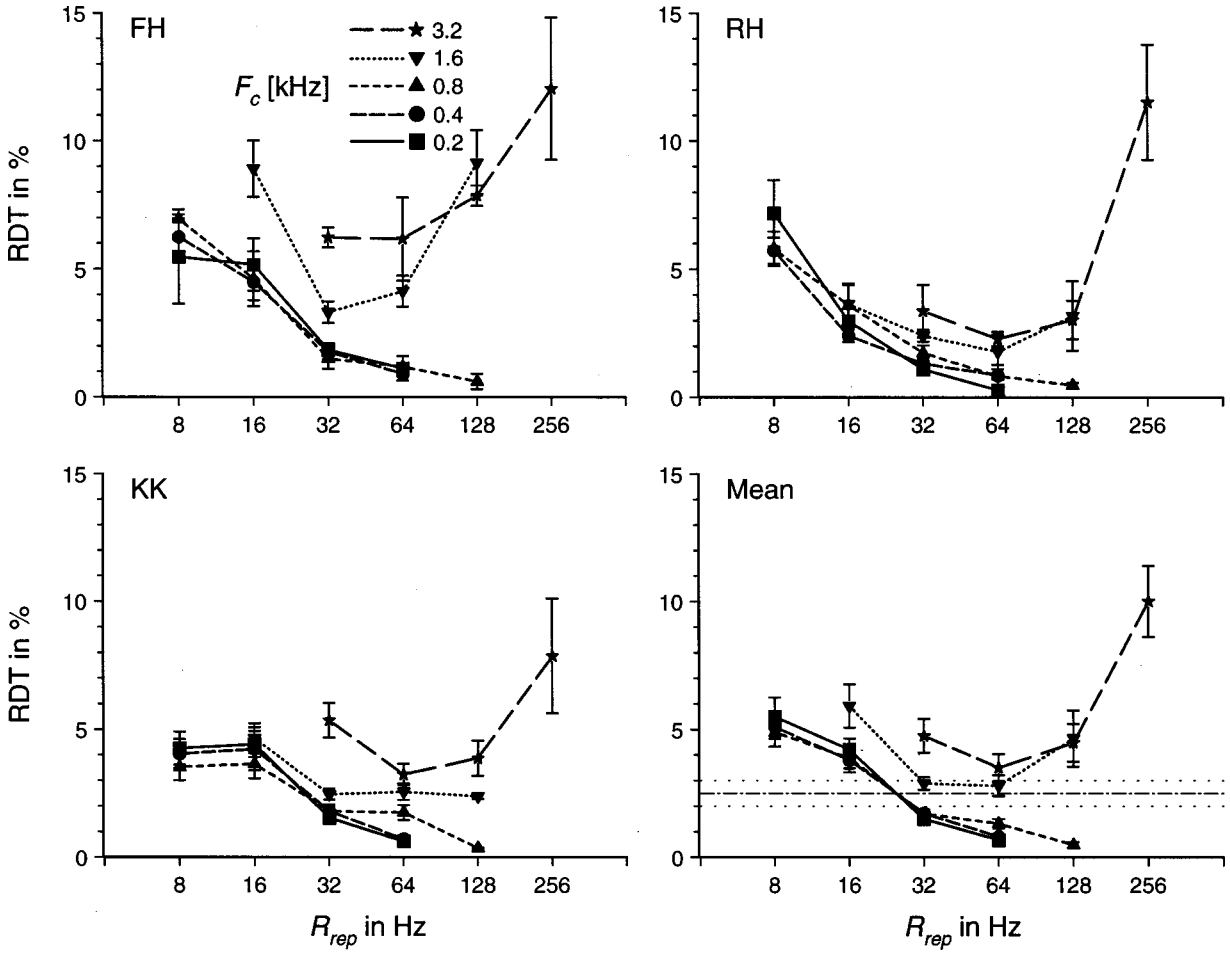


FIG. 8. Individual RDT functions for three listeners, and the average RDT function for APH tones, plotted in the same format as the CPH data in Fig. 4.

exhibit an extra peak of activity midway through each period. They analyzed the time intervals between the neural activity peaks by computing the autocorrelation function of each frequency channel and summing across channels. The summary autocorrelation function (SACF) for spectrally unresolved APH tones has its largest peak at the lag corresponding to the period of the wave, but it also has a secondary peak at a lag corresponding to the envelope rate of the stimulus.

If the lower limit of pitch is determined by a limitation in a temporal pitch mechanism, for example, a limit on the maximum time interval between neural activity peaks that is accurately represented, then the pitch limit for APH tones will be one octave lower than for CPH tones. This is because the peaks in the neural response to spectrally unresolved APH tones occur at twice the rate of those in the CPH response. If the same process determines rate discrimination threshold, then we would expect the transition from large to small discrimination thresholds to be shifted downwards by one octave in the APH RDT functions. Alternatively, if rate discrimination threshold is determined by the spectral resolution of the harmonic components, then we would expect the transitions for APH tones to occur at the same repetition rates as for CPH tones.

A. Methods

The stimuli were filtered in the same way as in experiment 1. Only the first five filter conditions were used (F_c $=0.2$, 0.4, 0.8, 1.6, and 3.2 kHz). Rate discrimination thresholds for APH and CPH tones were measured for subsets of a total of six repetition rates ($R_{\mathrm{rep}}=8$, 16, 32, 64, 128, and 256 Hz). Listeners FH, RH, and KK participated in this experiment. RH's and KK's CPH thresholds had already been measured in experiment 1. Both APH and CPH thresholds were measured for listener FH. The stimulus generation and procedure were the same as in experiment 1 (see Sec. I A). In the APH tones, the even and the odd harmonics were in sine- and cosine phase, respectively.

B. Results and discussion

The four panels of Fig. 8 show the RDT functions of three listeners and their average for APH tones. The discrimination functions in filter conditions 1, 2, and 3 (F_c $=0.2$, 0.4, and 0.8 kHz) exhibit a transition (from more than 5% at low repetition rates to about 1% at higher rates) similar to that in the respective CPH discrimination functions (Figs. 4 and 5). However, the transitions are shifted towards lower repetition rates by almost one octave. The average APH transition rates are shown by the filled squares in Fig. 6. The APH transition rates were calculated for a criterion of 2.5% and a range of 1%, as for the CPH data (see the dash-dotted line and the two dotted lines in the lower-right panel of Fig. 8). In the three lowest filter conditions ($F_c=0.2$, 0.4, and 0.8 kHz), the transition rates for APH tones are nearly one octave below the corresponding transition rates for CPH

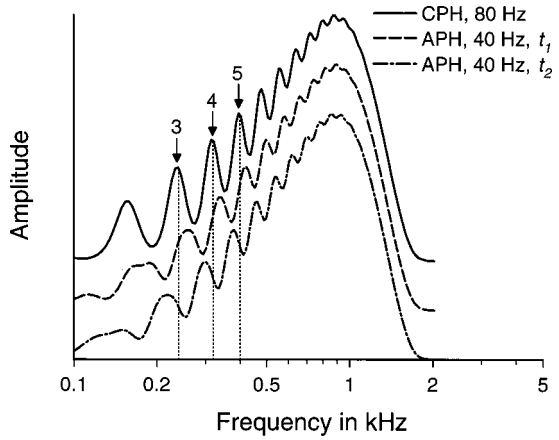


FIG. 9. Short-term spectral representations from the SPINET model of Cohen *et al.* (1995). The solid line shows the short-term spectrum for a CPH tone with a repetition rate of 80 Hz. The dashed and dash-dotted lines show short-term spectra for an APH tone with a repetition rate of 40 Hz at two different times, t_1 and t_2, separated by 12.5 ms. For clarity, the spectra are offset vertically by a constant amount. The vertical, dotted lines show the third, fourth, and fifth harmonics of 80 Hz. Note that the positions of the peaks in the spectrum for the APH tone change over the period of the stimulus.

tones. In the fourth and fifth filter conditions ($F_c = 1.6$ and 3.2 kHz), the average, and most of the individual, APH discrimination functions are nonmonotonic and do not reach the 2.5% criterion over the range of repetition rates tested. Nevertheless, the RDT functions in the fourth filter condition still exhibit a marked decrease between the 16- and 32-Hz repetition rate. The rapid increase of the APH thresholds at 128 and 256 Hz in filter conditions 4 and 5 is not due to pitch ambiguity arising from the limited bandwidth of the stimuli (as discussed in Sec. I B). Shackleton and Carlyon (1994) observed similarly elevated rate discrimination thresholds for APH tones with comparable spectral composition and repetition rate. They attribute their results to an ambiguity of the pitch of APH tones at the boundary between spectrally resolved and unresolved harmonics which mediate conflicting pitch cues.

The results for APH tones reinforce the hypothesis that the transitions in the RDT functions for the lower filter conditions reflect a limitation in a temporal pitch mechanism rather than the spectral resolution of harmonics. The effect of phase on the discrimination functions in Fig. 8 would be very difficult to explain with most spectral models, where the excitation pattern is derived from the magnitude of the long-term Fourier spectrum of the stimulus (Goldstein, 1973; Wightman, 1973; Terhardt, 1979). In the remainder of this section, we consider whether short-term spectral differences between CPH and APH tones might explain the experimental findings. Whereas the long-term Fourier spectra of CPH and APH tones with the same repetition rate are identical, their short-term spectra can differ, if the duration of the integration window is not too long with respect to the repetition rate. We used the **S**patial **PI**tch **NET** (SPINET) model of Cohen *et al.* (1995) to calculate short-term spectral representations for the stimuli used in the current experiments. The SPINET model is the most recent and, perhaps, the most comprehensive computational version of Goldstein's (1973) optimum processor model. The first stage of the SPINET model consists of a gammatone filterbank[5] (Patterson *et al.*, 1995) which simulates the spectral analysis performed in the cochlea. The second stage calculates the root-mean-square (rms) integral of the temporal output of each gammatone filter over a short, exponential window with a 7.24-ms time constant. The short-term spectral representation is produced by averaging the instantaneous rms values over 5 ms and multiplying the average with frequency-dependent weights which represent the combined effects of the outer- and middle-ear transfer function and the loss of phase locking at high frequencies. Cohen *et al.* present the short-term spectral representation as the input to their pitch mechanism which is a harmonic sieve preceded by spectral sharpening.

The dashed line in Fig. 9 shows the short-term spectral representation from SPINET for an APH tone in the first filter condition ($F_c = 0.2$ kHz) with a repetition rate of 40 Hz. For comparison, the solid line above it shows the spectral representation for a CPH tone with a repetition rate of 80 Hz. Both of these spectra were calculated at the end of a period of the stimulus, and both exhibit a comb-filter profile with peaks separated by 80 Hz, even though the frequency separation of the Fourier components in the APH tone is 40 Hz. Thus, the short-term spectral representation of the APH stimulus reflects the envelope rate rather than the repetition rate of the waveform. Note, however, that the alternating phase shift of the peaks in the APH wave (see Fig. 3) means that the peaks in the comb-filter profile for the APH tone are frequency shifted upwards from the harmonic relation illustrated by the CPH profile and the dotted vertical lines in Fig. 9 which show the positions of the third, fourth, and fifth harmonics of 80 Hz. Moreover, the positions of these peaks oscillate over time. This is shown by the dash-dotted line which shows the short-term representation for the APH tone calculated 12.5 ms, or half a repetition period, later in time. Thus, the peaks in the comb-filter profile for the APH tone shift back and forth over the course of the repetition period, and this oscillating inharmonic shift means that the short-term spectral information is unlikely to provide a simple explanation for the difference between the APH and CPH discrimination functions.

III. COMPARISON BETWEEN TRANSITION RATE AND LOWER LIMIT OF MELODIC PITCH

Pressnitzer *et al.* (1999) have measured the lower limit of melodic pitch (LLMP) with a melody task and the same stimuli as used in the present study. They defined the LLMP as the lowest repetition rate for which listeners could identify a semitone change in a novel four-note melody. The notes were randomly chosen from a chromatic scale and had a restricted range of a major third; the base note was adaptively varied to determine threshold. Pressnitzer *et al.* showed that the LLMP for CPH stimuli was about 35 Hz for the first and second filter conditions ($F_c = 0.2$ and 0.4 kHz); it increased with increasing filter cutoff for conditions 3 to 5 ($F_c = 0.8$, 1.6, and 3.2 kHz); and it was unmeasurable in the sixth filter condition ($F_c = 6.4$ kHz). For APH stimuli, the LLMP was almost an octave lower than that for CPH stimuli in the two lowest filter conditions; it was similar to the CPH values in the third and fourth filter conditions, and it was

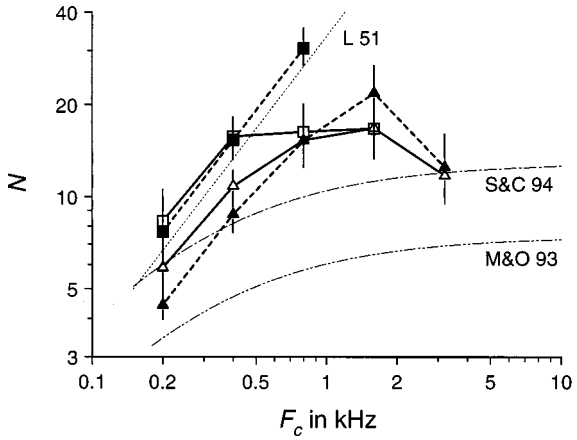


FIG. 10. Data on the lower limit of melodic pitch from Pressnitzer *et al.* (1999) for CPH tones (open triangles) and APH tones (open squares), respectively. The limit is expressed in terms of the harmonic number, *N*, associated with the filter cutoff frequency, F_c. The filled triangles and squares show the transition rates for CPH and APH tones. They are reproduced from Fig. 6 and expressed in terms of *N*. The dotted line shows the harmonic number corresponding to a constant repetition rate of 30 Hz. The dash-dotted and the dash-dot-dotted lines show the two definitions of the limit of spectral resolution proposed by Shackleton and Carlyon (1994) and by Moore and Ohgushi (1993), respectively.

unmeasurable in the fifth and sixth filter conditions. The open triangles and squares in Fig. 10 show the LLMP for CPH and APH tones as a function of the cutoff frequency, F_c. The LLMP is expressed in terms of the harmonic number, *N*, associated with F_c. As in Fig. 1(b), the dotted line shows the harmonic number corresponding to a constant repetition rate of 30 Hz, and the dash-dotted line shows the limit of spectral resolution proposed by Shackleton and Carlyon (1994). The dash-dot-dotted line shows the more traditional definition of the limit of spectral resolution based on the audibility of individual components in complex tones (Plomp, 1964; Plomp and Mimpen, 1968; Moore and Ohgushi, 1993). The region above the lines representing the two definitions of the limit of spectral resolution would be unresolved according to the respective criterion. The filled triangles and squares show the transition rates for the CPH and APH rate discrimination functions, reproduced from Fig. 6 and expressed in terms of *N*. The open and filled triangles in Fig. 10 show that, for CPH tones, there is a close correspondence between the LLMP and the transition in the RDT function for the entire frequency region from 0.2 to 3.2 kHz. Thus, in this frequency region, the LLMP for CPH tones corresponds to an approximately constant level of performance in rate discrimination of about 2.5% (the criterion level used to calculate the transition rates). For APH tones (open and filled squares), the correspondence between the LLMP and transition rate exists for the two lowest filter conditions ($F_c = 0.2$ and 0.4 kHz), but in the third filter condition ($F_c = 0.8$ kHz), the transition in the APH rate discrimination function occurs at about 26 Hz, whereas the LLMP is almost equal to the value for CPH stimuli (about 50 Hz). In the fourth filter condition ($F_c = 1.6$ kHz), only one of the three subjects managed to achieve the 2.5% discrimination level with APH tones, whereas the LLMP for APH tones is still measurable and approximately equal to the value for CPH tones. The comparison between the LLMP and transition rates for APH discrimination is probably complicated by the fact that the RDT functions become nonmonotonic in the higher filter conditions (Fig. 8).

IV. COMPARISON OF SPECTRAL AND TEMPORAL EXPLANATIONS OF THE LOWER LIMIT OF PITCH

The data from the current study indicate that the transitions in the RDT functions for CPH tones reflect the boundary of the pitch region. At low frequencies, this boundary seems to be determined by temporal factors rather than the spectral resolution produced by cochlear filtering. The first experiment showed that the transitions in the RDT functions for the lower filter conditions occur at a constant repetition rate rather than a constant fraction of filter cutoff frequency, F_c. The second experiment showed that the transitions for filter conditions 1 to 4 are determined by the temporal structure of the stimuli rather than the spectral separation of harmonics. Together, these results suggest that, in the frequency region below 1 kHz, the transition in rate discrimination is associated with a limitation in a temporal pitch mechanism which calculates time intervals between peaks in the phase-locked activity of the auditory nerve. It is as if the mechanism is unable to process time intervals longer than about 33 ms (corresponding to a pitch limit of about 30 Hz). Licklider (1951) hypothesized that pitch is processed by an array of neural delay lines and coincidence detectors. In his model, the lowest perceptible pitch would correspond to the longest delay that can be realized with sufficient precision in a neural delay line.

In the frequency region above 1 kHz, the lower limit of pitch increases sharply with increasing filter cutoff frequency, F_c. In this frequency region, the data from the present study are compatible with the data from previous studies (Ritsma and Hoekstra, 1974; Cullen and Long, 1986; Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994), which indicate that the transition in the RDT functions occurs at a roughly constant fraction of filter cutoff frequency, F_c. It should be noted, however, that this does not necessarily mean that above 1 kHz, the lower limit of pitch is determined by a limit in spectral resolution. As was pointed out in Sec. I C, many of the transition rates measured in the present and previous studies fall in a region where the harmonics are assumedly unresolved even by the more generous definition of resolution proposed by Shackleton and Carlyon (1994). This casts serious doubt on the presumed relationship between transition rate and spectral resolution.

If the lower limit of pitch is determined by a limitation in a temporal pitch mechanism in both the low- and high-frequency regions, then the increase of the pitch limit with increasing frequency above 1 kHz might mean that the range of time intervals which can be accurately represented varies with frequency. Moore's (1997) time-interval model of pitch includes such an assumption; in this model, the longest time interval which can be analyzed in a neural frequency channel with frequency, *F*, is given by $15/F$. Pressnitzer *et al.* (1999), however, have shown that the assumption of a frequency-dependent limit in time-interval processing is not necessary for predicting the lower limit of pitch in the high-

frequency region. They were able to model the pitch limit quantitatively with a modified version of Meddis and Hewitt's (1991) autocorrelation model of auditory temporal processing. In order to simulate the pitch limit in the low-frequency region, Pressnitzer *et al.* restricted the range of time intervals in the autocorrelation analysis to 33 ms by applying a linear sloping window to the SACFs. With this modification, the model correctly predicts not only a lowest perceptible pitch of just over 30 Hz for CPH tones and an octave shift for APH tones, but also an increase of the lower limit of pitch with increasing filter cutoff frequency, F_c. The success of the model in the high-frequency region is due to the fact that the pitch-related peaks in the SACF become ever smaller and broader with increasing frequency region (Meddis and O'Mard, 1997; see also Carlyon, 1998). Ultimately, the SACF will probably prove inadequate as a simulation of the auditory system's neural representation of periodic sounds. Nevertheless, the analysis shows that a temporal pitch mechanism can provide a unitary explanation of the lower limit of pitch in both low- and high-frequency regions.

ACKNOWLEDGMENTS

We would like to thank Lucy Jessop and Alexandra Smith for showing the way with their pilot study performed as an undergraduate research project in the Department of Physiology at the University of Cambridge. This research was supported by the UK Medical Research Council (No. G9703469).

[1]ERB refers to the equivalent rectangular bandwidth of the cochlear filter at a given frequency, F, which is $24.7+0.108*F$ according to the definition of Glasberg and Moore (1990).

[2]The discrimination function for the broadband condition is presented in Fig. 2(c); it was omitted in Fig. 2(d) because the effective lower cutoff frequency, F_c, of the broadband stimuli is not well defined.

[3]The envelope was calculated by filtering the squared waveform with a second-order Butterworth low-pass filter with a cutoff frequency of 0.6 kHz.

[4]The frequency modulation might have prohibited Shackleton and Carlyon's (1994) listeners from integrating pitch-related information over time. White and Plack (1998) have shown that the integration time for pitch is longer for harmonic tones consisting of unresolved, high-frequency harmonics than for harmonic tones with resolved, low-frequency harmonics. Thus, due to the frequency modulation, the listeners might have required a greater repetition rate to achieve criterion discrimination performance.

[5]In our implementation of the SPINET model, the gammatone filterbank had 256 filters whose center frequencies ranged from 0.1 to 8 kHz, evenly spaced on an ERB scale.

Carlyon, R. P. (**1998**). ''Comments on A unitary model of pitch perception,'' J. Acoust. Soc. Am. **104**, 1118–1121.

Cohen, M. A., Grossberg, S., and Wyse, L. L. (**1995**). ''A spectral network model of pitch perception,'' J. Acoust. Soc. Am. **98**, 862–879.

Cullen, Jr., J. K., and Long, G. (**1986**). ''Rate discrimination of high-pass filtered pulse trains,'' J. Acoust. Soc. Am. **79**, 114–119.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched noise data,'' Hear. Res. **47**, 103–138.

Goldstein, J. L. (**1973**). ''An optimum processor theory for the central formation of the pitch of complex tones,'' J. Acoust. Soc. Am. **54**, 1496–1516.

Hoekstra, A. (**1979**). ''Frequency discrimination and frequency analysis in hearing,'' Doctoral dissertation (Institute of Audiology, University Hospital, Groningen, The Netherlands).

Houtsma, A. J. M., and Smurzynski, J. (**1990**). ''Pitch identification and discrimination for complex tones with many harmonics,'' J. Acoust. Soc. Am. **87**, 304–310.

Kaernbach, C., and Demany, L. (**1998**). ''Psychophysical evidence against the autocorrelation theory of temporal processing,'' J. Acoust. Soc. Am. **104**, 2298–2306.

Kohlrausch, A., and Houtsma, A. J. M. (**1992**). ''Pitch related to spectral edges of broadband signals,'' Philos. Trans. R. Soc. London, Ser. B **336**, 375–382.

Licklider, J. C. R. (**1951**). ''A duplex theory of pitch perception,'' Experientia **7**, 128–133.

Levitt, H. (**1971**). ''Transformed up–down methods in psychophysics,'' J. Acoust. Soc. Am. **49**, 467–477.

Lyon, R. F. (**1984**). ''Computational models of neural auditory processing,'' IEEE ICASSP (March).

Meddis, R., and Hewitt, M. J. (**1991**). ''Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification,'' J. Acoust. Soc. Am. **89**, 2866–2882.

Meddis, R., and O'Mard, L. (**1997**). ''A unitary model of pitch perception,'' J. Acoust. Soc. Am. **102**, 1811–1820.

Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing*, 4th ed. (Academic, New York).

Moore, B. C. J., and Ohgushi, K. (**1993**). ''Audibility of partials in inharmonic complex tones,'' J. Acoust. Soc. Am. **93**, 452–461.

Patterson, R. D. (**1973**). ''The effects of the relative phase and number of components on residue pitch,'' J. Acoust. Soc. Am. **53**, 1565–1572.

Patterson, R. D., and Wightman, F. L. (**1976**). ''Residue pitch as a function of component spacing,'' J. Acoust. Soc. Am. **59**, 1450–1459.

Patterson, R. D. (**1987**). ''A pulse ribbon model of monaural phase perception,'' J. Acoust. Soc. Am. **82**, 1560–1586.

Patterson, R. D., Allerhand, M. H., and Giguère, C. (**1995**). ''Time-domain modeling of peripheral auditory processing: A molecular architecture and a software platform,'' J. Acoust. Soc. Am. **98**, 1890–1894.

Plomp, R. (**1964**). ''The ear as a frequency analyser,'' J. Acoust. Soc. Am. **36**, 1628–1636.

Plomp, R. (**1967**). ''Pitch of complex tones,'' J. Acoust. Soc. Am. **41**, 1526–1533.

Plomp, R. (**1976**). *Aspects of Tone Sensation* (Academic, London).

Plomp, R., and Mimpen, A. M. (**1968**). ''The ear as a frequency analyzer. II,'' J. Acoust. Soc. Am. **43**, 764–767.

Preisler, A. (**1993**). ''The influence of spectral composition of complex tones and of musical experience on the perceptibility of virtual pitch,'' Percept. Psychophys. **54**, 589–603.

Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (**1999**). ''Lower limit of melodic pitch for filtered harmonic complexes,'' J. Acoust. Soc. Am. **105**, 1152.

Ritsma, R. J. (**1962**). ''Existence region of the tonal residue. I,'' J. Acoust. Soc. Am. **34**, 1224–1229.

Ritsma, R. J. (**1967**). ''Frequencies dominant in the perception of the pitch of complex sounds,'' J. Acoust. Soc. Am. **42**, 191–198.

Ritsma, R. J., and Hoekstra, A. (**1974**). ''Frequency selectivity and the tonal residue,'' in *Facts and Models in Hearing*, edited by E. Zwicker and E. Terhardt (Springer, Berlin), pp. 156–163.

Schouten, J. F. (**1938**). ''The perception of subjective tones,'' Proc. K. Ned. Akad. Wet. **41**, 1086–1093.

Schouten, J. F., Ritsma, R. J., and Cardozo, B. L. (**1962**). ''Pitch of the residue,'' J. Acoust. Soc. Am. **34**, 1418–1424.

Shackleton, T. M., and Carlyon, R. P. (**1994**). ''The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination,'' J. Acoust. Soc. Am. **95**, 3529–3540.

Terhardt, E. (**1974**). ''Pitch, consonance, and harmony,'' J. Acoust. Soc. Am. **55**, 1061–1069.

Terhardt, E. (**1979**). ''Calculating virtual pitch,'' Hear. Res. **1**, 155–182.

Thurlow, W. R., and Small, Jr., A. M. (**1955**). ''Pitch perception for certain periodic auditory stimuli,'' J. Acoust. Soc. Am. **27**, 132–137.

Warren, R. M., and Bashford, Jr., J. A. (**1981**). ''Perception of acoustic iterance: Pitch and infrapitch,'' Percept. Psychophys. **29**, 395–402.

White, L. J., and Plack, C. J. (**1998**). ''Temporal processing of the pitch of complex tones,'' J. Acoust. Soc. Am. **103**, 2051–2063.

Wightman, F. L. (**1973**). ''The pattern transformation model of pitch,'' J. Acoust. Soc. Am. **54**, 407–416.

Characterizing frequency selectivity for envelope fluctuations

Stephan D. Ewert
Carl von Ossietzky Universität Oldenburg, AG Medizinische Physik, D-26111 Oldenburg, Germany

Torsten Dau
Carl von Ossietzky Universität Oldenburg, AG Medizinische Physik, D-26111 Oldenburg, Germany and Boston University, Hearing Research Center, Department of Biomedical Engineering, 44 Cummington Street, Boston, Massachusetts 02215

(Received 22 October 1999; revised 21 March 2000; accepted 8 June 2000)

Three experimental paradigms were used to specify the auditory system's frequency selectivity for amplitude modulation (AM). In the first experiment, masked-threshold patterns were obtained for signal-modulation frequencies of 4, 16, 64, and 256 Hz in the presence of a half-octave-wide modulation masker, both applied to the same noise carrier with a bandwidth ranging from 1 to 4 kHz. In the second experiment, psychophysical tuning curves (PTCs) were obtained for signal-modulation frequencies of 16 and 64 Hz imposed on a noise carrier as in the first experiment. In the third experiment, masked thresholds for signal-modulation frequencies of 8, 16, 32, and 64 Hz were obtained according to the ''classical'' band-widening paradigm, where the bandwidth of the modulation masker ranged from $\frac{1}{8}$ to 4 octaves, geometrically centered on the signal frequency. The first two experiments allowed a direct derivation of the shape of the modulation filters while the latter paradigm only provided an indirect estimate of the filter bandwidth. Thresholds from the experiments were predicted on the basis of an envelope power-spectrum model (EPSM) which integrates the envelope power of the modulation masker in the passband of a modulation filter tuned to the signal-modulation frequency. The Q-value of second-order bandpass modulation filters was fitted to the masking patterns from the first experiment using a least-squares algorithm. Q-values of about 1 for frequencies up to 64 Hz suggest an even weaker selectivity for modulation than assumed in earlier studies. The same model also accounted reasonably well for the shape of the temporal modulation transfer function (TMTF) obtained for carrier bandwidths in the range from 1 to 6000 Hz. Peripheral filtering and effects of peripheral compression were also investigated using a multi-channel version of the model. Waveform compression did not influence the simulated results. Peripheral bandpass filtering only influenced thresholds for high modulation frequencies when signal information was strongly attenuated by the transfer function of the peripheral filters. © *2000 Acoustical Society of America.* [S0001-4966(00)03509-8]

PACS numbers: 43.66.Ba, 43.66.Dc [SPB]

I. INTRODUCTION

Effects of frequency selectivity in the audio-frequency domain are well established and associated with the existence of independent frequency channels, or critical bands (Fletcher, 1940; Zwicker *et al.*, 1957). Several researchers have suggested a similar spectral decomposition in the envelope-frequency domain (Kay, 1972; Bacon and Grantham, 1989; Houtgast, 1989; Dau *et al.*, 1997a, b). Houtgast (1989) performed modulation-masking experiments where the subject's task was to detect a sinusoidal signal modulation in the presence of a fixed narrow-band masker modulation, both applied to a broadband noise carrier. He found that for each half-octave-wide masker, centered at 4, 8, and 16 Hz, respectively, the masking pattern showed a peak at the masker frequency. In addition, Houtgast (1989) performed an experiment adopted from the classical band-widening paradigm, as described by Fletcher (1940). The modulation masker was a noise of variable bandwidth centered at a signal frequency of 8 Hz. The spectral density of the masker was held constant in all conditions. Signal modulation thresholds increased up to a masker bandwidth of about $\frac{1}{2}$ octave and remained constant for larger bandwidths. Bacon and Grantham (1989) measured detection thresholds for sinusoidal signal modulation in the presence of a sinusoidal masker modulation, both imposed on the same broadband noise carrier. Most modulation masking generally occurred when the signal frequency was at the masker frequency, resulting in bandpass-shaped masking patterns. The results were interpreted as indicating frequency selectivity in the envelope-frequency domain, analogous to the frequency selectivity in the audio-frequency domain. No quantitative derivations of the bandwidth and shape of the hypothetical filter mechanism were undertaken in these studies. Furthermore, the above experiments were restricted to relatively low signal modulation frequencies, up to 64 Hz.

In a more recent study, Dau *et al.* (1997a, b) examined AM detection thresholds using narrow-band noise carriers (see also Fleischer, 1982). The noise carrier bandwidths were 3, 31, and 314 Hz, respectively. The rate of the random envelope fluctuations inherent in the noise carrier increased with increasing bandwidth, resulting in a very different shape of the temporal modulation transfer function (TMTF) for these conditions. These findings could be described qualitatively by assuming that detection of the signal modulation

was hardest when its frequency fell within the spectral range of the inherent envelope fluctuations of the carrier. Dau *et al.* (1997a, b) explained the results quantitatively in terms of a model based on a modulation filterbank. Further evidence for such a mechanism was demonstrated in a more recent study by Dau *et al.* (1999). TMTFs were obtained with different types of narrow-band noise as the carrier, having the same overall level and a bandwidth of 50 Hz. The three carriers, Gaussian noise, multiplied noise, and low-noise noise, produced very different TMTFs for modulation rates from 3 to 100 Hz. The results could be explained reasonably well by taking into account the different shapes of the carrier envelope spectra and by assuming that the amount of inherent envelope power of the carrier that passes a specific modulation filter determines detection threshold.

Thus, the modulation filterbank concept appears to be a powerful tool for describing TMTFs in various stimulus configurations. However, the filterbank parameters have been determined in experiments which only represent an indirect measure of the frequency selectivity. Furthermore, the parameters were originally derived in the framework of a relatively complex signal processing model (Dau *et al.*, 1996a, b, 1997a, b). Thus, the shape of the resulting modulation filters was probably influenced by the transfer characteristic of the adaptation stage prior to the modulation filterbank. Although the adaptation stage represents a major processing stage within the framework of the whole processing model, the role of its time-dependent compressive properties for the processing of modulations is still unclear.

The first part of the current article presents new experimental data which reflect more direct measures of the modulation filter shape. All experimental paradigms were adopted from classical studies performed in the audio-frequency domain: masked-threshold patterns, psychophysical tuning curves, and masking patterns according to the band-widening paradigm. Data were obtained for a large range of signal frequencies (up to 256 Hz) to examine the ''existence region'' of modulation filters.

In the second part of the study, simulations are presented which are based on a model referred to here as the ''envelope power spectrum model'' (EPSM) of modulation masking. The Q-value of the modulation filters assumed in this model are fitted to the experimental masking data. The same model parameters are used for the simulation of experimental TMTF data from the literature. The question is whether the shape and amount of modulation masking as well as the shape of the TMTF and the overall sensitivity to AM can be explained by the same underlying mechanisms. Furthermore, the influence of peripheral filtering and peripheral (waveform) compression are investigated in the framework of a multi-channel version of the EPSM. Capabilities and limitations of the model are discussed.

II. METHOD

A. Procedure and subjects

Modulation detection thresholds were obtained using an adaptive three-interval forced-choice (3IFC) procedure. The stimuli were presented in three consecutive intervals separated by 500-ms silent intervals. Two of the three intervals contained the reference stimulus. Depending on the specific experiment, this reference stimulus was either an unmodulated noise (TMTF condition) or a noise on which a stochastic masker modulation was imposed (masking condition). In one randomly chosen interval, the sinusoidal signal modulation was imposed on the same carrier. The subject's task was to identify the interval containing the signal modulation. During a threshold run, the modulation index m, in dB ($20 \log m$), was adjusted according to a two-down one-up rule (Levitt, 1971) which provides an estimate of the modulation index necessary for 70.7% correct responses. The modulation depth was varied by using a 4-dB step size at the beginning of the adaptive track. After each second reversal the step size was divided by 2 until it reached 1 dB. At this step size, eight reversals were obtained and threshold was calculated as the median value of $20 \log m$ at these reversals. The subjects received visual feedback after each response. The measurement procedure was repeated at least three times for each configuration and subject. For each subject, the final modulation detection threshold was taken as the average across the last three threshold estimates. Unless otherwise stated, all figures show mean data and standard deviations calculated by averaging across three subjects. All subjects had experience in psychoacoustic experiments and had clinically normal hearing. They were between 24 and 33 years old and were paid for participation in the study.

B. Apparatus

All acoustic stimuli were digitally generated at a sampling rate of 32 kHz. The stimuli were converted to analog signals by a two-channel 16-bit D/A converter including reconstruction filtering. The stimuli were presented diotically via headphones (Sennheiser HD 25) in a sound-attenuating booth. Signal generation and presentation during the experiments were computer controlled using a signal-processing software package developed at the Drittes Physikalisches Institut at the university of Göttingen. Analytical calculations and computer simulations were performed with Matlab.

C. Stimuli

In all experiments, the carrier duration was 600 ms including 50-ms $\cos^2$ onset and offset ramps. Both the sinusoidal signal modulation and the masker modulation had a duration of 500 ms including 50-ms $\cos^2$ ramps and were temporally centered in the carrier. In all conditions the signal modulation started at a positive going zero-crossing. Gaussian noise (GN) and low-noise noise (LNN) were used as modulation maskers. The GN was generated as a sequence of random samples from a Gaussian distribution and was restricted to the required bandwidth by setting the Fourier coefficients outside the desired passband to zero. The LNN was originally described by Pumplin (1985) and Hartmann and Pumplin (1988, 1991). They generated LNN by defining the amplitude of its spectral components, while optimizing the phase of these components using a gradient search procedure to achieve a temporal waveform with minimized envelope

(power) fluctuations. This procedure results in a temporal waveform with a smooth envelope. It is, however, quite time consuming and not easily implemented. In the present study, a similar waveform was generated in a way described by Kohlrausch *et al.* (1997): the temporal waveform of a band-limited GN was divided by its envelope on a sample-by-sample basis. Spectral splatter introduced by this process was then removed by restricting the stimulus to its original bandwidth. These two steps were iterated ten times. The resulting waveform provides a smooth envelope and small power fluctuations. Although the generation algorithm differs considerably from the algorithm proposed by Hartmann and Pumplin, the noise waveform will be referred to as LNN in the following.

The rationale for using LNN as masker in the present study was to introduce a stimulus which can also be applied at large energies without producing overmodulation when applied to the signal. This should, for instance, allow the measurement of masking patterns with a maximum dynamic range (in the modulation domain). This is interesting since, for broadband noise carriers, the reference detection thresholds (as reflected in the TMTF) increase with increasing signal frequency, causing a reduced available dynamic range for observing masking at these higher frequencies.

In the case of the LNN as masker, a long-duration realization of approximately 8 s (2^{18} samples) was generated only once in advance and was stored to disk. In the case of the GN as masker, a realization of approximately 2 s (2^{16} samples) was generated before each threshold run. In each presentation interval during the experimental run, the masker was cut out randomly from the long realization.

Signal and masker were applied consecutively to a wide-band noise carrier, resulting in the envelope $(1+M_m(t))(1+M_s(t))$, where $M_m(t)$ reflects the masker modulation and $M_s(t)$ reflects the signal modulation. This multiplicative approach was also used by Houtgast (1989). Unless otherwise stated, the carrier was restricted to the 1- to 4-kHz band after applying the modulation. The reduction in effective modulation depth was less than 1 dB for the largest signal modulation frequency applied. An unmodulated and uncorrelated inverse bandstop-filtered noise of the same type was added at the same spectrum level. Thus, the resulting stimuli in the signal and the nonsignal interval(s) always had the same long-term power spectrum. The spectral manipulations were performed in the audio-frequency domain by adjustment of the Fourier coefficients of the 2-s signal realizations.

The average power of a sinusoidally amplitude-modulated stimulus is increased by the factor $1+m^2/2$ compared to the unmodulated stimulus. For large modulation depths, detection might therefore be based on changes in the overall intensity rather than on the presence or absence of modulation. To eliminate level cues, the waveforms were adjusted to have equal energy in each interval of the forced-choice trial. The overall level of the stimuli was 65 dB SPL, equivalent to a spectrum level of 23 dB.

III. EXPERIMENTS

A. Masked-threshold patterns

Masked-threshold patterns (MTPs) are commonly considered as a very useful paradigm to investigate the auditory system's frequency selectivity in the audio-frequency domain. Generally, the threshold for a sinusoidal signal is obtained in the presence of a narrow-band masker of fixed frequency as a function of the signal frequency. Houtgast (1989) and Bacon and Grantham (1989) adopted this experimental paradigm to the modulation domain. They measured detection thresholds for a sinusoidal amplitude modulation in the presence of an additional masker modulation, applied to a broadband noise carrier. The masker modulation was a narrow-band noise in the Houtgast study and a tone in the study by Bacon and Grantham. The spectral position of the masker modulation was fixed in the two studies. Both studies showed most masking when the signal frequency was at the masker frequency. The peaked patterns obtained for a certain masker frequency may be considered as reflecting the ''modulation excitation pattern'' evoked by the modulation masker, analogous to the excitation patterns commonly described in the audio-frequency domain (e.g., Zwicker and Feldtkeller, 1967; Moore and Glasberg, 1986). Thus, the shape of the masked-threshold pattern for a particular masker frequency does not directly reflect the shape of a specific modulation filter. A more direct approach to characterize the modulation-filter shape at a specific ''position'' would be to measure the threshold for the (fixed) signal frequency in the presence of the masker frequency whose spectral position is varied relative to the signal frequency. This fixed signal modulation frequency approach was used by Takahashi and Bacon (1992) and Lorenzi *et al.* (1997), and was also used in the present study for deriving MTPs. Such a paradigm has sometimes also been used in the audio-frequency domain (Chistovich, 1957; Rodenburg *et al.*, 1974).

1. Conditions

The GN was used as the carrier. The signal frequency was 4, 16, 64, or 256 Hz. For each signal frequency, the spectral position of the masker band was varied in the range from -2 to $+2$ octaves relative to the signal frequency, using a step size of $\frac{2}{3}$ octave. The bandwidth of the masker was 1.4, 5.6, 22.3, and 89.2 Hz for the signal frequencies of 4, 16, 64, and 256 Hz, respectively. In each of the on-frequency conditions (where the masker was centered geometrically at the signal frequency) this corresponds to a $\frac{1}{2}$-octave bandwidth. We have chosen such a configuration because previous studies suggested a logarithmic scaling of the modulation filter bandwidth. In the case of GN as the modulation masker, the rms modulation depth of the masker was -10 dB in all conditions. In the case of LNN as the masker, the rms modulation depth of the masker was -10 dB for the signal frequencies 4 and 16 Hz while it was -7 dB for the signal frequencies of 64 and 256 Hz. Since LNN exhibits less envelope fluctuations than GN the masker level could be increased by 3 dB without producing overmodulation when applied to the signal.[1]

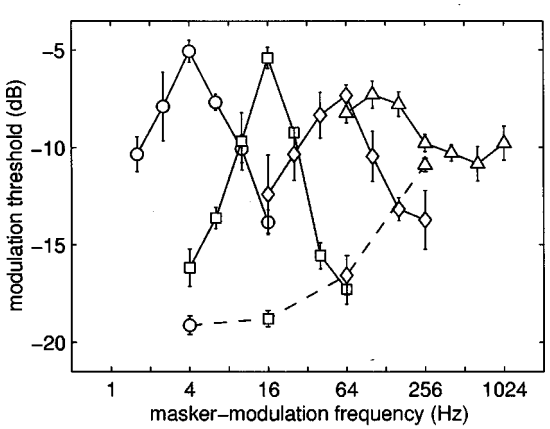


FIG. 1. Average masked threshold patterns (MTPs) for signal frequencies of 4 Hz (circles), 16 Hz (squares), 64 Hz (diamonds), and 256 Hz (triangles). The signal modulation depth at threshold is plotted as a function of the masker frequency. The masker level was always −10 dB. In each on-frequency condition, the GN masker was $\frac{1}{2}$ octave wide. The absolute bandwidth was held constant when the masker was shifted in the range from −2 to +2 octaves relative to the signal frequency (see text). The dashed curve represents the ''reference'' thresholds without masker modulation. The carrier was a bandpass GN ranging from 1 to 4 kHz.

In addition, the TMTF was measured for the same carrier in order to obtain reference thresholds for the signal frequencies without any external masker modulation.

2. Results

Figure 1 shows average masked-threshold patterns, obtained with a carrier noise band in the range from 1 to 4 kHz. Signal modulation depth at threshold is plotted as a function of the masker frequency. In this condition, the masker was a GN. The four peaked patterns represent the data for the signal frequencies of 4 Hz (circles), 16 Hz (squares), 64 Hz (diamonds), and 256 Hz (triangles).

The patterns for 4, 16, and 64 Hz show a maximum masking effect when the masker is centered at the signal frequency. The shape of the pattern is approximately symmetric on a logarithmic frequency scale. The pattern for 256 Hz looks quite different from the patterns at the lower signal frequencies. Masking can only be observed for masker frequencies below the signal frequency. For masker frequencies at and above the signal frequency, hardly any masking effect can be observed; masked thresholds are about the same as the reference threshold for 256 Hz in the TMTF condition. The maximum amount of masking, given by the difference between the masked threshold and the corresponding reference threshold, is about 14 dB for the signal frequencies 4 and 16 Hz, and 10 and 4 dB for the signal frequencies 64 and 256 Hz.

Figure 2 shows corresponding data obtained with LNN instead of GN as the masker. The masking patterns for 4, 16, and 256 Hz are very similar to those obtained with the GN masker. For 64 Hz, the pattern is somewhat broadened compared to the GN condition and has a slightly asymmetric shape. At this frequency the peak of the pattern is shifted by $-\frac{2}{3}$ octave towards lower envelope frequencies. The slope of the low-frequency tail of the pattern is somewhat reduced compared to that obtained with GN. For all signal frequencies the peak threshold is at about −5 dB. Note that the patterns for 64 and 265 Hz are shifted by about 3 dB towards higher values compared to the patterns obtained with GN, as a consequence of the higher level for the LNN masker. Again, since the modulation thresholds of the TMTF increase with increasing frequency, the resulting masking effect at the peak frequencies decreases with increasing frequency.

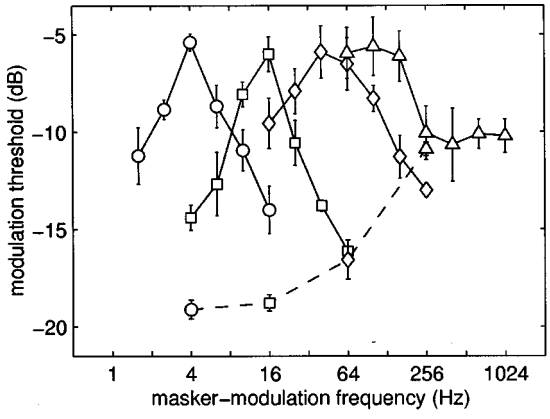


FIG. 2. Same as in Fig. 1 but with LNN as the masker modulation instead of GN. The carrier was the same as in the previous experiment.

Figure 3 shows masked-threshold patterns for 16 and 64 Hz obtained with a carrier noise band in the range from 4 to 7 kHz. Thresholds connected by the dashed line represent the TMTF condition (as reference condition without masker modulation). Both masked thresholds as well as reference thresholds are very similar to those obtained in the low-frequency-carrier condition from the previous figure.

B. Psychophysical tuning curves

This experiment measures psychophysical tuning curve (PTCs) in the evelope-frequency domain, which are also adopted from a popular measure of the frequency selectivity in the audio-frequency domain. The PTC is meant to represent a perceptual correlate of a standard physiological measure, the frequency threshold curve (FTC), commonly referred to as the tuning curve. The FTC is obtained by measuring the level of a sinusoidal stimulus at a certain cri-

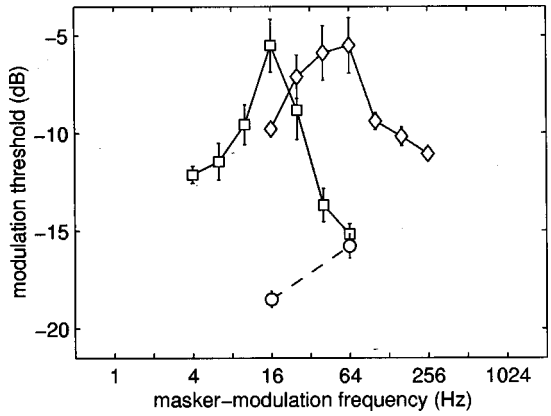


FIG. 3. Same as in Fig. 2 but with a GN carrier ranging from 4 to 7 kHz instead of 1 to 4 kHz.

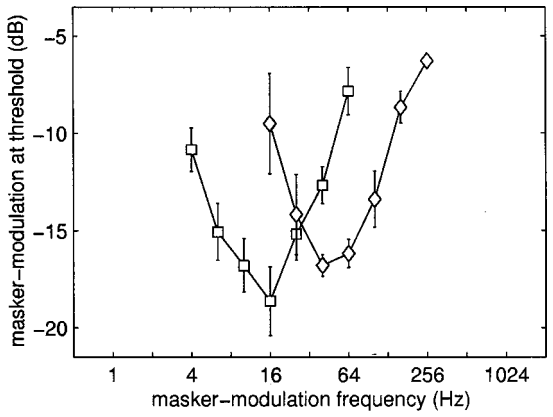


FIG. 4. Average psychophysical tuning curves (PTCs) for the signal frequencies 16 Hz (squares) and 64 Hz (diamonds). The level of the LNN masker at signal threshold is plotted as a function of the masker frequency. The signal was presented at a fixed level of 4 dB above the average detection threshold. Masker bandwidth and center frequency were the same as in the corresponding MTP experiment from Fig. 2.

terion response in a primary auditory neuron, as a function of the frequency of the stimulus (Kiang *et al.*, 1965). The frequency requiring the lowest level is referred to as the characteristic frequency (CF) of the auditory neuron. As the stimulus frequency diverges from the CF, the level needed to achieve a constant activity of the neuron usually rises.

To obtain a PTC, the signal is presented to the subject at a fixed level and frequency in the presence of a narrow-band masker whose spectral position is varied. The PTC thus reflects the level of the masker at detection threshold for the signal as a function of the masker frequency. It is generally assumed that when the signal level is close to its threshold, it will stimulate only a small group of neurons with similar CF and the PTC paradigm might become analogous to a corresponding physiological tuning curve.

1. Conditions

The PTCs in the modulation domain were obtained for 16 and 64 Hz. The level of the signal modulation was kept constant at 4 dB above the average detection threshold (-15 dB for 16 Hz and -13 dB for 64 Hz). At this modulation depth the signal modulation was clearly detectable for all subjects. As for the MTPs from the previous experiment, the masker was a $\frac{1}{2}$-octave-wide LNN (relative to the signal frequency) which was presented in the range from 2 octaves below to 2 octaves above the signal frequency. The GN was not tested. The subject's task was to adjust the masker level so that it just masked the signal. The starting level of the masker was 5 dB below the signal level (near threshold).[2]

2. Results

Figure 4 shows average PTCs. Masker modulation depth at signal threshold is plotted as a function of the masker frequency. The shape of the PTCs largely corresponds to the shape of the corresponding MTPs from Fig. 2. For 64 Hz, the peak of the PTC is shifted by one data point towards lower frequencies as it was also observed for the MTP. The dynamic range of the PTC is 3 dB larger (10 dB) than that of the MTP (7 dB). For 16 Hz, the PTC is tuned slightly broader than the corresponding MTP. For both signal frequencies (16 and 64 Hz), the two paradigms do not show differences in the symmetry of the patterns.

C. Band-widening experiment

The band-widening experiment represents a classical experimental approach to characterizing the bandwidth of a peripheral filter and was first described by Fletcher (1940). The basic idea behind this paradigm is to determine the filter bandwidth from the shape of the threshold curve obtained with a pure tone presented in noise with variable bandwidth. The noise is generally centered at the signal frequency. Since the spectral power density is held constant, the total power of the masker increases as the bandwidth increases. Assuming auditory filters to be rectangular in shape as a first approximation, and assuming thresholds to be proportional to the noise power at the output of the filter, the following shape of the threshold curve should be expected: As long as the noise bandwidth is smaller than the filter bandwidth threshold will increase by 3 dB per doubling of the noise bandwidth, while threshold should remain constant for larger bandwidths. The threshold curve could then be fitted by two lines and the bandwidth corresponding to the point of intersection provides an estimate of the ''critical bandwidth'' (Fletcher, 1940; Bos and deBoer, 1966).

In the present study, as in the study by Houtgast (1989), the band-widening technique was transposed to the modulation domain to estimate the width of the modulation channels. In fact, modulation filters (as well as peripheral filters) are far from rectangular in shape. Nevertheless, as long as the filter function exhibits a relatively sharp roll-off on both sides of its passband, the threshold function obtained in the band-widening experiment should show a narrow transition between the increasing and constant part of the function.

1. Conditions

The band-widening experiment was performed for signal frequencies of 8, 16, 32, and 64 Hz. A LNN masker was used which was centered geometrically at the signal frequency. The bandwidth of the masker ranged from $\frac{1}{2}$ to 4 octaves. The signal and masker were imposed on a broadband GN carrier. For a signal frequency of 8 Hz, the experiment was repeated with a GN masker and a pink-noise carrier to most closely reproduce the experimental conditions described by Houtgast (1989). As in the Houtgast study, the stimuli were restricted to 1–4 kHz after modulation. In addition, the inverse bandstop-filtered carrier noise was added at the same spectrum level.

2. Results

Figure 5 shows the data obtained with LNN as the masker. All data points represent the difference in thresholds between the masked condition and the reference condition (TMTF), expressed as masking (in dB). Masking is plotted for the signal frequencies 8 Hz (circles), 16 Hz (squares), 32 Hz (diamonds), and 64 Hz (triangles), respectively.

For each signal frequency, threshold increases with increasing masker bandwidth. This increase is observed for the

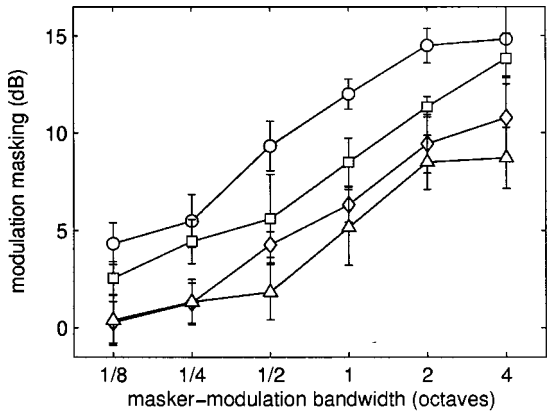


FIG. 5. Average data obtained in the band-widening experiment. Masking is plotted as a function of the masker bandwidth for the signal frequencies 8 Hz (circles), 16 Hz (squares), 32 Hz (diamonds), and 64 Hz (triangles). Masking was derived by subtracting the reference threshold from the masked threshold. The masker was a LNN which was geometrically centered at the signal frequency. As in most of the previous conditions, the carrier was a GN ranging from 1 to 4 kHz.

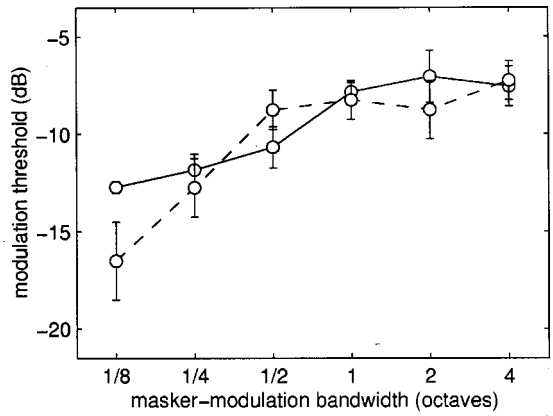


FIG. 6. Comparison of the data from the present study and the data of Houtgast (1989). The masked threshold for 8 Hz is plotted as a function of the masker bandwidth. The masker was GN instead of LNN as in the previous experiment. The solid curve represents our own data while the dashed curve is redrawn from Houtgast (1989). The experimental conditions were the same in both studies.

whole range of masker bandwidths. For 8 and 64 Hz, thresholds appear to asymptote at the largest masker bandwidths (2 and 4 octaves). For each bandwidth masking decreases with increasing signal frequency. For the largest bandwidth condition, masking amounts to 15 dB for 8 Hz and only 8 dB for 64 Hz.

Figure 6 shows the data for 8 Hz obtained with a GN masker as in the Houtgast study. For direct comparison, the data from Houtgast are indicated by the dashed curve. In the data of Houtgast, thresholds first increase at a rate of 3–4 dB per octave increase of the noise bandwidth, and then remain constant for bandwidths larger than $\frac{1}{2}$ octave. In contrast, our data (indicated by the solid curve) do not show such a behavior. Thresholds show an almost continuous increase with increasing bandwidth, as was also observed for the LNN masker.

D. Discussion

The data obtained in the present study provide further psychoacoustical evidence for frequency-selective mechanisms involved in the processing of amplitude modulation. The MTPs showed a basically symmetric shape on a logarithmic frequency scale for signal frequencies up to 64 Hz. The shape and bandwidth of the patterns correspond very well with those of the PTCs, suggesting that the principal underlying mechanism is the same. No tuning was observed for the signal frequency of 256 Hz. Probably, two different processes are involved, one responsible for modulation tuning and a second one responsible for the general decrease of sensitivity, or sluggishness, at frequencies above about 100–150 Hz. This will be further investigated in the modeling section of the present article.

Differences were observed between the MTPs obtained with LNN and GN as the modulation masker. The patterns appeared to be slightly more sharply tuned for GN than for LNN. Although the use of LNN has some appeal because of its low-peak factor, it also may complicate the interpretation of the data, since additional detection cues may become available. As we realized during the experiments, modulation beatings between the signal frequency and the instantaneous frequency of the masker can occur and influence the shape of the masking pattern. Thresholds appeared to be most influenced in the on-frequency condition where the perceptual salience of the beats was strongest. This is also in line with recent data by Moore *et al.* (1999), who showed that the perceptual salience of modulation beating at low rates is strong enough to influence the detection threshold for signal modulation at a similar rate. Thus, the resulting pattern obtained with LNN as a modulation masker may underestimate the amount of frequency selectivity for modulation. Using a GN as the masker, modulation beatings probably play a minor role since they are at least partly masked by the intrinsic fluctuations of the masker.

The experimental data obtained in the band-widening experiment differ from those published by Houtgast (1989). While the shape of the threshold curve in the Houtgast study was quite similar to that generally obtained in corresponding experiments in the spectral domain, the current study only showed a general increase of masking with increasing masker bandwidth. In fact, one might argue that only the data point at a masker bandwidth of $\frac{1}{8}$ octave clearly differs in both studies while all other data points lie within 2 dB (Fig. 6). However, this exactly reflects the problem behind the band-widening experiment: differences in single data points can lead to very different estimates of the underlying selectivity which makes this paradigm unreliable. Our own data do not allow any direct estimation of the underlying modulation filter characteristics. Qualitatively, the filters appear to broaden with increasing center frequency since the amount of masking decreases with increasing signal frequency. The fact that the slope of the curves is similar for all signal frequencies as a function of the masker bandwidth indicates that the filter bandwidth does not change much on a logarithmic frequency scale. These observations are consistent with the observations from the previous experiments. Thus, the data obtained in the band-widening experiment

may still be interpreted in favor of modulation channels. However, these data alone, of course, would hardly allow a reliable estimation of absolute bandwidth and shape of the filters, a problem which also has been described in a similar way for the audio-frequency domain (Patterson and Moore, 1986).

In order to derive quantitative values for the modulation filter parameters and to test whether the experimental data of the study can be described consistently, model predictions are needed. These are presented in the following section.

IV. MODEL PREDICTIONS

A. Envelope power spectrum model (EPSM)

In this section, the data obtained in the present study are compared with model predictions. The model contains three basic stages: envelope extraction of the stimuli, bandpass filtering in the envelope-frequency domain, and calculation of the ac-coupled power at the output of a specific (modulation) bandpass filter. Since the principal structure of the model is very similar to the structure of the well-known power-spectrum model of masking (Fletcher, 1940; Patterson and Moore, 1986) but transposed to the envelope-frequency domain, it will be referred to as the ''envelope power spectrum model'' (EPSM) in the following. Within the EPSM it is assumed that AM-detection thresholds are related to the total envelope noise power that passes through the modulation filter centered at the signal frequency.

This modeling approach is equivalent to the calculations presented in a recent study by Dau *et al.* (1999), where TMTFs were predicted for a large set of carrier bandwidths and for different types of noise as the carrier. In contrast to the calculations in Dau *et al.* (1999), the present study generally deals with modulation *masking* conditions where the nonsignal interval contains a carrier which is already modulated. Thus, in these conditions the envelope power spectrum of the nonsignal stimulus is given by the convolution of the envelope power spectrum of the carrier alone with the envelope power spectrum of the modulator (see the Appendix).

To calculate the envelope power of modulated noise falling in the transfer range of a specific modulation filter (centered at a frequency cf_{env}), the envelope power spectrum of the modulated noise, $N_{(m)}$, must be multiplied by the squared transfer function of the modulation filter, $W_{cf}(f_{\mathrm{env}})$, and then integrated across envelope frequency f_{env}. The ac-coupled envelope power, $P_{\mathrm{env}}(cf_{\mathrm{env}})$, can then be written as

$$P_{\mathrm{env}}(cf_{\mathrm{env}}) = \frac{1}{N_{(m)}(0)} \int_{f_{\mathrm{env}}>0}^{\infty} N_{(m)}(f_{\mathrm{env}}) W_{cf}(f_{\mathrm{env}}) df_{\mathrm{env}}. \quad (1)$$

Here $N_{(m)}(0)$ denotes the dc value of the envelope power. If the carrier is an unmodulated bandlimited GN, its envelope power spectrum, N, can be calculated according to the formula by Lawson and Uhlenbeck (1950). They showed that, assuming a rectangular shape of the power spectrum of a Gaussian noise, its envelope power spectrum approximately has a triangular shape besides the dc peak and stretches up to the frequency that equals the noise bandwidth (for details, see also Dau *et al.*, 1999).

One way to relate the calculated envelope power to an experimentally obtained AM-detection threshold is to assume a certain signal-plus-noise-to-noise $(S+N)/N$ ratio at threshold. In the calculations presented here, a 1-dB $(S+N)/N$ criterion was used, assuming that a 1-dB increase in the envelope power at the output of the modulation filter tuned to the signal modulation frequency is necessary to detect the signal modulation. For computational reasons, the $(S+N)/N$ criterion was converted into a S/N ratio at threshold. Under the assumption that signal and noise are statistically independent (which is the case in the present study), the corresponding signal-to-noise ratio is $S/N \approx -6$ dB. In order to introduce a lower limit for the integrated envelope power, a value corresponding to an AM-detection threshold of -30 dB was finally added after integration (see Dau *et al.*, 1999). This value is intended to reflect the empirically found ''absolute'' detection threshold for a signal modulation imposed on a sinusoidal carrier at a medium level (e.g., Viemeister, 1979; Kohlrausch *et al.*, 2000).[3]

The present model does not incorporate the influences of spectral filtering after modulation. Filtering after modulation reduces the effective modulation depth of the modulated stimulus—something that is not included in the calculations. Furthermore, the inverse bandstop-filtered masker noise, which was added in the experiments, was not considered in the calculations: the auditory system is assumed to be ''focused'' at the region between lower and upper spectral edge of the stimuli. More importantly, the current version of the EPSM does not include peripheral filtering, adaptation, or any influence of the statistics of the integrated envelope power which is a random variable for short-term intervals. In this respect, the EPSM model differs substantially from the more complex processing model described in Dau *et al.* (1997a, b) despite the similar modulation filterbank approach. In particular, the complex model contains an adaptation stage which allows the description of both simultaneous and nonsimultaneous masking (Dau *et al.*, 1996a, b; Derleth and Dau, 2000) as well as modulation detection data (Dau *et al.*, 1997a, b). Such a stage is not contained within the EPSM. In addition, the detection stage within the complex model performs a kind of pattern recognition of the whole temporal course of the internal representation of the stimuli. Decisions are based on the cross correlation between the internal representation of the actual stimulus and a normalized suprathreshold template, and an adaptive tracking procedure is used as in the real experiment. The EPSM is a model which is restricted to amplitude modulation processing and makes only a few simple assumptions about the processing of modulated stimuli. Nevertheless, even if this model probably oversimplifies the ''real'' processing in the auditory system (more than does the complex model), this analytical approach might be helpful for the understanding of what processing stages are essential for successfully describing modulation detection and masking phenomena.

1. Fitting modulation filters to the masking data

The shape of the experimental MTP, obtained in Sec. III, was basically symmetrical on a logarithmic frequency

TABLE I. Best fitting Q-values for the second-order bandpass filters assumed within the framework of the EPSM. The Q-value was fitted to the masking patterns derived from Figs. 1–3. A least-square fitting procedure was used. Corresponding −3-dB bandwidths of the filters are given in parentheses. The asterisks indicate that a Q-value could not be derived from the data. The bars denote conditions where no experimental data were obtained.

	Type of masker noise, carrier range		
cf (Hz)	GN, 1–4 kHz	LNN, 1–4 kHz	LNN, 4–7 kHz
4	0.714 (5.6)	0.711 (5.6)	—
16	1.713 (9.3)	1.255 (12.7)	0.981 (16.3)
64	0.728 (87.9)	0.411 (155.7)	0.315 (203.2)
256	*	*	—

scale, for signal frequencies up to 64 Hz. In a previous study it was observed that the general sensitivity for modulation decreases above about 150 Hz (Kohlrausch *et al.*, 2000). Such a behavior was found for sinusoidal carriers at high frequencies where neither intrinsic carrier fluctuations nor spectral cues can influence detection (at least up to several hundred Hertz). Based on these observations, the modulation filterbank proposed here consists of an array of second-order bandpass filters combined with an additional first-order lowpass filter with a cutoff frequency of 150 Hz. With the Q-value as the only free parameter within the model, the best-fitting filter function for each empirical MTP was obtained using a least-squares fitting procedure. The squared deviation between predicted and measured masking was minimized using a Nelder-Mead-type simplex search method.

Table I shows the best fitting Q-values for the different signal frequencies (cf) 4, 16, 64, and 256 Hz for all conditions tested in the experiments, i.e., with GN and LNN as the modulation masker and with the low- and high-frequency noise band as the carrier, respectively. The values in parentheses represent the corresponding −3-dB bandwidths of the filter.

The asterisk for 256 Hz indicates that it was not possible to derive a Q-value from these data. The derived Q-value for the 4-Hz filter is about the same for GN and LNN while for the signal frequencies 16 and 64 Hz it differs considerably for the two noises, indicating sharper tuning for GN than for LNN. However, as already addressed in Sec. III D, the LNN pattern might have been influenced by cues resulting from modulation beating, particularly in the on-frequency condition, which would result in an underestimation of the system's ''real'' amount of envelope frequency selectivity.

The upper panel (a) of Fig. 7 shows model predictions (closed symbols) of the masking patterns, obtained with the best-fitting values for the modulation filters, together with the corresponding experimental masking data (open symbols, from Fig. 1). The predicted patterns agree very well with the experimental data. The middle panel (b) of Fig. 7 shows corresponding predictions where, for reasons of simplicity, a constant Q-value of 1 was assumed for all modulation bandpass filters. The correlation between model and data still is quite reasonable, except perhaps for 16 Hz, where the predicted masking pattern is too broad.

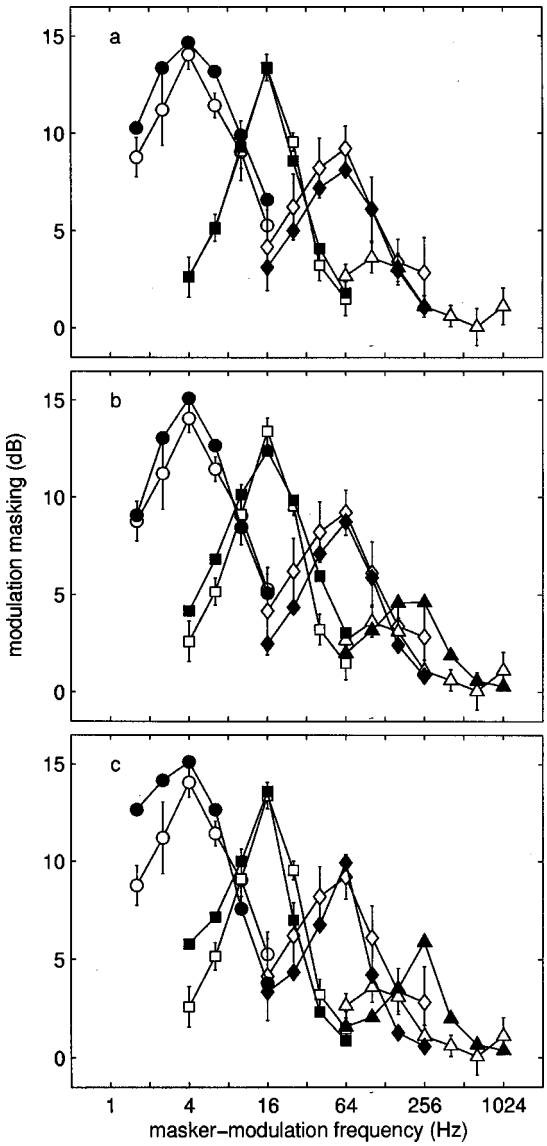


FIG. 7. Predicted masking patterns (filled symbols) in comparison with the experimental data (open symbols) from Fig. 1. Masking is plotted as the difference between the thresholds obtained in the masked and the reference conditions. The experimental data are the same in each panel. Different realizations of the modulation filterbank were used for the simulations. The uppel panel (a) represents simulations obtained with the best-fitting Q-values for the modulation filters. The middle panel (b) shows simulations obtained with constant-Q-bandpass filters (Q=1) as an approximation of the best-fitting filters from the upper panel. In the lower-panel (c), the resonance filters proposed in earlier studies (Dau *et al.*, 1997a, b, 1999) were used for the simulaitons.

In earlier studies (Dau *et al.*, 1997a, b, 1999) where the shape of modulation filters was estimated only indirectly, first-order resonance filters were assumed with a constant absolute bandwidth of 5 Hz for modulation frequencies up to 10 Hz and a Q-value of 2 for filters centered at frequencies larger than 10 Hz. The lower panel (c) of Fig. 7 shows model predictions where these filters were incorporated in the EPSM. For the frequencies 64 and 256 Hz, the patterns show too sharp tuning while for 4 and 16 Hz the low-frequency skirt is too shallow. For direct comparison, Fig. 8 shows the transfer functions of the filters (dashed lines) suggested in the earlier studies in the framework of the more complex signal processing model together with those for the band-

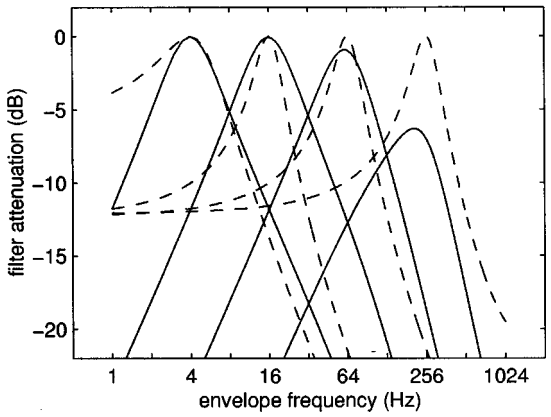


FIG. 8. Transfer functions of the second-order bandpass filters with constant Q=1 (solid functions) and of the first-order resonance filters suggested in earlier studies (dashed functions).

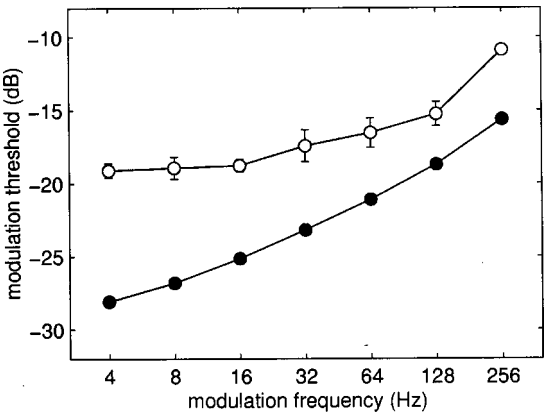


FIG. 10. TMTFs for broadband (1 to 4 kHz) noise as the carrier. The open symbols represent average experimental data, the filled symbols represent model predictions.

pass filters with a Q-value of 1 (solid lines), as suggested in the present study.

Finally, Fig. 9 shows model predictions from the band-widening experiment described in Sec. II. In this and in the following simulations, a constant Q of 1 was assumed for all signal frequencies. First, as observed in the experimental data from Fig. 5, thresholds increase continuously with increasing masker bandwidth over the entire range of bandwidths tested. Second, masking decreases with increasing signal frequency, which also agrees well with the data. The model slightly overestimates the amount of masking for the signal frequencies 16, 32, and 64 Hz for the narrow-bandwidth maskers ($\leqslant$1 octave) by about 1–3 dB.

Overall, the predictions demonstrate that second-order modulation bandpass filters account very well for the shape of the MTPs as well as for the masking data obtained in the band-widening experiment. The filter shape appears to play a larger role than the exact choice for the Q-value. A constant Q-value of 1 as an approximation to the best-fitting values still describes the data reasonably well.

2. Modeling modulation transfer functions

While the model parameters have been optimized to describe the amount of modulation masking correctly, it is not clear in advance whether the same model can also account for the shape of the TMTF, which serves as the reference condition for the masking data.

Figure 10 shows the TMTF (open symbols) for the 3-kHz-wide noise carrier referred to as the broadband carrier in the following. In addition to the data shown in Fig. 1, thresholds were also obtained at the intermediate signal frequencies 8, 32, and 128 Hz. The filled symbols indicate predictions obtained with the EPSM. Each simulated threshold represents the envelope power of the carrier in the passband of the filter tuned to the signal frequency. As a consequence of the constant Q-value of the filters, the TMTF increases continuously at a rate of about 3 dB per octave with increases in signal frequency. This is in contrast to the empirical data which show a plateau up to about 16–32 Hz before threshold increases. The data are in line with those from many other studies (e.g., Rodenburg, 1972, 1977; Viemeister, 1977, 1979; Eddins, 1993; Strickland and Viemeister, 1997; Eddins, 1999).[4]

To get a broader picture of the model's capabilities and limitations in predicting TMTFs, Fig. 11 shows data (open symbols) and model predictions (filled symbols) for a large

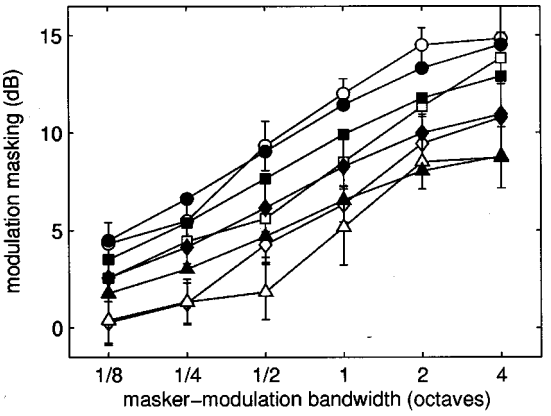


FIG. 9. Predicted masking (filled symbols) in the band-widening experiment. For direct comparison, the experimental data from Fig. 5 are replotted and indicated as open symbols.

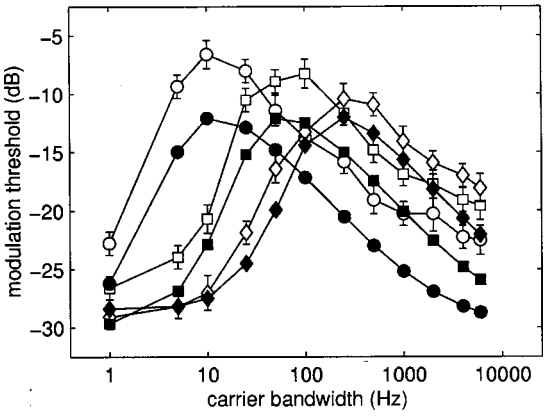


FIG. 11. AM thresholds for 5 Hz (circles), 25 Hz (squares), and 100 Hz (diamonds) as a function of the carrier bandwidth. The upper cutoff frequency of the carrier was 6 kHz. The open symbols indicate experimental data and are replotted from Dau *et al.* (1999). The filled symbols represent model predictions.

set of carrier bandwidths ranging from 1 Hz to 6 kHz, with 6 kHz as the upper cutoff frequency of the carrier. The data are replotted from Dau *et al.* (1999). The AM thresholds for 5 Hz (circles), 25 Hz (squares), and 100 Hz (diamonds) are plotted as a function of the carrier bandwidth. For a fixed carrier bandwidth, the figure indirectly reflects the principal shape of the TMTF (for details, see Dau *et al.*, 1999). That is, for large carrier bandwidths ($\geqslant$250 Hz) thresholds increase with increasing signal frequency; for small carrier bandwidths ($\leqslant$25 Hz) thresholds decrease with increasing signal frequency; and for bandwidths in the region between 25 and 250 Hz threshold is largest for the intermediate signal frequency (25 Hz).

The model accounts for the general shape of the threshold patterns. In particular, the shape of the predicted patterns is much closer to the data than that of the calculations presented in Dau *et al.* (1999), especially for small carrier bandwidths. This is a further indication of the appropriateness of the filter parameters chosen here. However, there remain some discrepancies between data and predictions. The predicted thresholds are generally lower than the corresponding experimental thresholds. The effect is small (1–3 dB) for a signal frequency of 100 Hz, but increases with decreasing signal frequency. For 5 Hz, the deviations between model and data amount to 3–8 dB. At large carrier bandwidths ($\geqslant$250 Hz) the model predicts a decrease of about 3 dB per octave increase of the carrier bandwidth, independent of signal frequency. In contrast, the experimental data show a smaller change in threshold (1–2 dB per octave) with increasing bandwidth beyond 250 Hz. Thus, the predicted TMTF for a wideband carrier (e.g., 3 kHz) is much too sensitive at 5 Hz with sensitivity decreasing too rapidly with increasing signal frequency. This was also observed in Fig. 10. The discrepancy between predicted and obtained TMTF will be discussed in detail in Sec. III C 1.

B. Effects of peripheral filtering

1. Multi-channel model

The EPSM in its present state does not include the influence of peripheral filtering on modulation detection. Several studies suggested that AM thresholds do not depend much on the frequency region of a bandpass noise carrier. In other words, thresholds obtained with a fixed-bandwidth carrier centered at low frequencies are about the same as those for a carrier shifted towards higher frequencies (e.g., Viemeister, 1979; Eddins, 1993; Dau *et al.*, 1997b). These data would suggest that information about the signal modulation in the low-frequency region will be combined optimally across channels without loss of information. In fact, this would imply that the data could be modeled very efficiently by simply assuming one broad ''predetection'' filter as was already done by Viemeister (1979) as well as in the present model. However, even if such a simple description has some appeal, a more realistic model would be of more general interest. The ''predetection'' filter approach would not account for data in more complex conditions such as, e.g., modulation detection interference (MDI) (e.g., Yost and Sheft, 1988, 1989; Yost *et al.*, 1989) or comodulation masking release (CMR) (e.g., Hall *et al.*, 1984; McFadden, 1986; Hall *et al.*, 1989) where across-peripheral-channel interactions seem to play a role. In the following this ''predetection'' filter EPSM is referred to as the single-channel EPSM.

In order to investigate the role of peripheral filtering on AM detection within the framework of the EPSM, a multi-channel version of the model was implemented in which the envelope fluctuations are processed in all excited peripheral channels. Since the bandwidth of the envelope spectrum of a noise stimulus is directly related to the bandwidth of the stimulus' audio-frequency spectrum, peripheral filtering produces narrowband envelope spectra. Assuming rectangular filter shapes for the peripheral filters, the corresponding envelope spectra can be derived analytically, as in the single-channel EPSM presented earlier. However, if one assumes more realistic filter shapes such as Gammatone filters (Patterson, 1987), the corresponding envelope spectra can no longer be easily derived analytically. In such a case, numerical simulations are needed where the envelope power spectrum is estimated for the output of each peripheral channel by averaging over a large number of stimulus realizations.

In the following, sample TMTF simulations are presented, which were obtained with a multi-channel EPSM. The average envelope power of the carrier was calculated for each peripheral filter at the output of the modulation filter tuned to the signal frequency. Comparable to the single-channel model, AM threshold was estimated by adjusting the signal-modulation depth to match a 1-dB$(S+N)/N$ decision criterion, corresponding to a signal-to-noise ratio of $S/N = d'^2/2 \approx -6$ dB for the combined observation. To satisfy this ''overall'' criterion within the multi-channel model, observations were combined across audio-frequency channels by assuming that

$$d' = \left(\sum_{i=1}^{n} (d_i')^2 \right)^{1/2} = \left(\sum_{i=1}^{n} \frac{2S_i}{N_i} \right)^{1/2}, \tag{2}$$

where d_i' denotes the sensitivity index for peripheral channel i, which is proportional to the square root of the signal-to-noise ratio S_i/N_i in this channel (Tanner and Sorkin, 1972), and d' denotes the overall sensitivity index. This is equivalent to a linear combination of n independent observations according to the ''integration model'' by Green and Swets (1966). Thus, it is assumed within the multi-channel EPSM that information is combined without any loss, and that decisions are based on this combined information. Analogous to the analytical approach, resolution was finally limited by adding a -30-dB threshold.

As long as external noise is the limiting factor for detection, the independence of observations will depend on the correlation between the carrier waveform in the different peripheral channels. The correlation depends in turn on the shape and spacing of the filters. This was discussed in the framework of the study by Dau *et al.* (1997b). For instance, by allowing some overlap between adjacent filters, the observations in these filters are not statistically independent. By using a wider spacing, the information gain from combining a certain number of filters is larger. However, with a wider spacing, the number of filters that could be placed within a given spectral range would decrease. This decrease in the

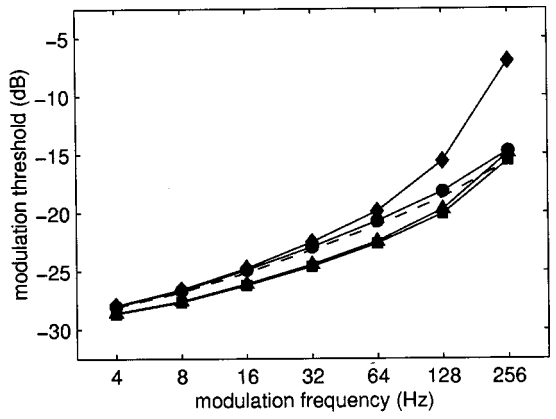


FIG. 12. Predicted TMTFs obtained with different versions of the EPSM. Dashed curve: Analytically derived thresholds obtained with the ''single-channel'' model as in the predictions from the previous figures. Filled circles: Numerically derived single-channel predictions based on averaged stimulus realizations (for details, see text). Filled diamonds: Numerically derived predictions obtained with the multi-channel EPSM assuming rectangular-shaped peripheral filters with bandwidths corresponding to the ERB values suggested by Moore and Glasberg (1983). Filled triangles: Corresponding multi-channel predictions assuming overlapping Gammatone filters. Filled squares: As for the triangles but, in addition, an instantaneous waveform compression was assumed in the preprocessing (for details, see text).

number of observations would affect thresholds more strongly than the gain from the statistical independence. Overall, it turned out that the typically chosen overlap at the −3-dB points of the transfer functions, which was also chosen here, allow a close to optimal detection of modulation imposed on a broadband noise carrier with a minimum computational load, i.e., for such a filter configuration, performance was comparable to that for nonoverlapping (rectangular) filters.

2. Predictions

Figure 12 shows predictions of the TMTF for broadband carriers. Four different model realizations were used. The dashed threshold curve represents results obtained with the original single-channel EPSM where thresholds were derived analytically. The filled circles represent single-channel predictions where thresholds were derived numerically from the averaged ''real'' envelope spectra instead of the approximated ones. The two curves differ by less than 1 dB. This small difference reflects the influence of bandwidth restriction of the carrier after modulation. While the numerical approach considers exactly the same stimulus conditions as in the real experiments, the analytical approach represents a (very good) approximation.

Results from the multi-channel predictions are indicated by the remaining symbols. The filled diamonds represent thresholds obtained with a model which assumes rectangular peripheral filters (within 1–4 kHz) with bandwidths according to the equivalent rectangular bandwidth (ERB) suggested by Moore and Glasberg (1983). For modulation frequencies below 32 Hz, peripheral filtering has no influence on thresholds. For frequencies larger than 32 Hz peripheral filtering starts having an influence. Thresholds are increased by 1, 3, and 8 dB for the frequencies 64, 128, and 256 Hz, respectively, relative to the single-channel prediction (filled circles). The reason for the large effect at high frequencies is that only a few peripheral channels (the upper 5 out of 11 filters in case of a modulation frequency of 256 Hz) pass the signal modulation while the remaining channels in the lower carrier-frequency region infinitely attenuate the signal.

Results change considerably when Gammatone filters are used instead of the unrealistic rectangular filter shapes. Corresponding modeling results are indicated by the filled squares. Thresholds are decreased by about 1.5 dB relative to the single-channel simulation. This shift is caused by the partly correlated observations at the output of the overlapping Gammatone filters which are combined under the assumption of statistical independence within the model. Apart from this effect, peripheral filtering only has an effect at the highest signal frequency (256 Hz), where threshold is increased by about 2 dB.

Finally, the filled triangles in Fig. 12 represent predictions where, in addition, an instantaneous compressive nonlinearity was assumed for each Gammatone channel. A power-law function with an exponent of 0.4 was assumed. This is reasonably consistent with physiological estimates of the amount of compression on the basilar membrane for mid-range levels (Ruggero *et al.*, 1997). It has been shown that such a compressive nonlinearity can have major influence on masked thresholds as, e.g., in conditions of forward-masking (Oxenham and Moore, 1995; Moore and Oxenham, 1998; Oxenham and Plack, 1997; Plack and Oxenham, 1998). It also has been suggested that a peripheral nonlinearity affects the threshold for detecting a signal modulation in the presence of a pair of masker modulators beating at the rate of the signal frequency: the nonlinearity introduces a distortion component at the beat rate in the internal representation of the envelope which might interact with the signal component in a phase-sensitive way (Moore *et al.*, 1999). In each peripheral channel, the compressed stimulus $c(t)$ was calculated from the input stimulus $s(t)$ according to

$$c(t) = \text{sign}\,(s(t)) \cdot |s(t)|^{0.4}, \tag{3}$$

where sign( ) denotes the signum function.

The negligible differences between thresholds obtained with (filled triangles) and without compression (filled squares) indicates that peripheral compression has no effect on the results in the framework of the current model. Since both signal modulation and inherent carrier modulations are compressed in the same way, the signal-to-noise ratio at the output of the model does not change. These predictions are consistent with the observation that sensorineural hearing-impaired listeners often show about the same sensitivity for modulation independent of the amount of hearing loss (e.g., Bacon and Viemeister, 1985; Formby, 1987; Bacon and Gleitman, 1992; Derleth *et al.*, 2000), at least for narrowband noise carriers, and also for broadband noise carriers as long as the hearing loss is relatively flat. Sensorineural hearing impairment is commonly associated with a loss or reduction of basilar-membrane compression (e.g., Yates, 1990; Glasberg and Moore, 1992; Ruggero and Rich, 1991).

C. Discussion

The envelope power spectrum model (EPSM) accounts well for both modulation masking and modulation detection data. While the model parameters have been fitted to the masked-threshold patterns (MTPs) the model also accounts for the data from the band-widening experiment as well as for the shape of the threshold patterns (from Fig. 11) for most of the carrier-bandwidth conditions. The use of second-order bandpass filters led to better results than the filters proposed in the earlier studies. The modulation filters appear to have a constant Q-value of about 1 for frequencies up to about at least 64 Hz. While frequency selectivity can clearly be observed up to this frequency, the effects are much weaker at higher frequencies. Within the model, an additional first-order low-pass filter with a cutoff frequency of 150 Hz introduces a certain sluggishness into the system which reduces the available dynamic range of detectable modulation depths at high frequencies. This corresponds, in principle, to the modeling strategy suggested by Viemeister (1979). However, Viemeister (1979) and Strickland and Viemeister (1997) suggested a much lower cutoff frequency (of 60–70 Hz) to fit their model to the TMTF data obtained with broadband carriers when a change in the envelope power was used as the decision statistic. A cutoff frequency of 150 Hz has also been suggested in a recent study on TMTFs and beat-detection thresholds for sinusoidal carriers (Kohlrausch *et al.*, 2000).

1. The role of the decision variable

As described in Sec. III A 2, some discrepancies between predictions and data were observed for the shape of the TMTF obtained with a broadband noise carrier (Fig. 10). Due to the logarithmic scaling of the modulation filter bandwidth the predicted TMTF increases at a rate of about 3 dB per doubling of the signal frequency. Obviously, this does not agree with the experimental data at low signal frequencies where threshold is constant up to about 30–60 Hz.

The question is what mechanisms may be involved which are not reflected in the EPSM. In the present state of the EPSM, modulation detection thresholds are directly related to the average integrated envelope power. It was assumed in the model that a 1-dB increase in power is required for detection. In terms of signal detection theory, this is equivalent to assuming that the envelope power (at the output of the modulation filter) forms a random variable, whose standard deviation and mean are proportional to one another. Such a decision statistic was also suggested by Viemeister (1979) who used the ac-coupled rms power of the output of a 64-Hz low-pass filter (instead of a bandpass modulation filter tuned to the signal frequency), calculated over the duration of the observation interval. Instead of the rms power, some studies have used the ratio of the largest to the smallest instantaneous value, max/min (Forrest and Green, 1987), the fourth moment (Hartmann and Pumplin, 1988), the crest factor (Hartmann and Pumplin, 1988), or the average magnitude of the slope of the envelope (Richards, 1992) as the decision statistic (for details, see Strickland and Viemeister, 1996). Some models of modulation detection for noise carriers (e.g., Viemeister, 1979; Wakefield and Viemeister, 1990; Dau *et al.*, 1997a, b) incorporate the ''true'' variation in the stimulus as the limiting quantity for detection by using the same adaptive tracking procedure as in the experiments.

The use of a constant S/N or $(S+N)/N$ ratio as the decision criterion, as assumed within the EPSM, appears to be in contrast to experimental data on modulation depth discrimination (Fleischer, 1990; Wakefield and Viemeister, 1990). In Fig. 3 of Wakefield and Viemeister (1990), discrimination thresholds are expressed as a Weber fraction which directly corresponds to the S/N ratio defined in our study. They found small Weber fractions in the range from -4 to -7 dB for standards with modulation depths larger than -10 dB. A value of -6 dB equals a $(S+N)/N$ ratio of 1 dB, as assumed in the present study. However, for modulation depths near detection thresholds, Weber fractions increased strongly with decreasing standard modulation depth corresponding to S/N ratios up to as much as 6 dB. Thus, Weber's law does not hold over a large range of modulation depth. Wakefield and Viemeister (1990) were able to predict part of their discrimination data with Viemeister's (1979) leaky integrator model (using a cascade of a bandpass filter with a 2-kHz bandwidth, a half-wave rectifier, and a time constant of 2.5 ms corresponding to a cutoff frequency of 64 Hz). Thresholds were estimated using a 2IFC tracking procedure, which means that the statistical properties of the stimuli were taken into account. For low standard modulation depths, this model could predict the data reasonably well while it was much too sensitive at higher standard depths. In contrast, our S/N criterion would account for the data for high standard modulation depths, where Weber's law holds, but would be much too sensitive at low standard depths. These findings suggest that a combination of both criteria might be appropriate for describing modulation detection and discrimination data.

In the following, model predictions are presented where decisions were based on the *statistics* of the integrated envelope power. Each single component in the envelope spectrum of a bandpass Gaussian noise is known to closely follow an exponential distribution (Green *et al.*, 1992). This distribution has the specific property that the mean and standard deviation are equal. The mean of the envelope power passing through a specific modulation filter equals the sum over the means of a number of spectral components which are weighted by the (power) transfer function of the modulation filter. For a sufficiently large number of components, this integrated envelope power becomes a Gaussian random variable. The variance of this random variable can be computed by adding up the variances of the individual spectral components which were assumed to equal their squared means to a good first approximation.[5] It is assumed in the following that detection requires an increase in power which is proportional to three times its standard deviation. This value was chosen to closely reproduce the empirically found sensitivity to modulation. In addition, as an internal limit of resolution, it is assumed that the required increase in power cannot become smaller than 1 dB, which corresponds to our original decision criterion. Again, an absolute threshold corresponding to a modulation depth of -30 dB was assumed.

Figure 13 shows model predictions for three of the ex-

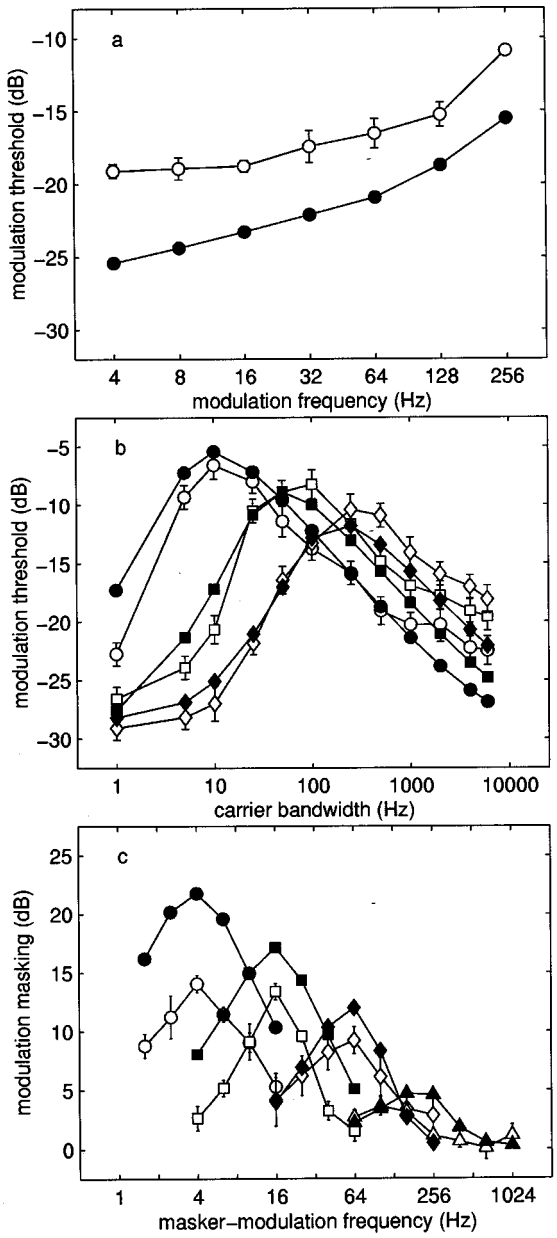


FIG. 13. Predictions (filled symbols) obtained with a ''statistical'' decision criterion. Decisions are based on the change of the standard deviation of the envelope power rather than on the change of the envelope power itself (for details, see text). Upper panel: Prediction of the TMTF for a broadband carrier (compare to Fig. 10). Middle panel: Prediction of the AM-detection data from Fig. 11. Lower panel: Prediction of the masking patterns derived from Fig. 1 (compare to Fig. 7). The open symbols indicate experimental data and are replotted from previous figures.

perimental conditions from the present study. The upper panel (a) of the figure shows results for the TMTF obtained with a broadband (1 to 4 kHz) carrier (compare to Fig. 10). The middle panel (b) shows results for the detection data from Fig. 11. Finally, the lower panel (c) represents predictions of the masking patterns derived from Fig. 1 [compare to panel (c) in Fig. 7]. The model accounts for several aspects in the data which were not accounted for by the original model. First, the slope of the broadband TMTF (upper panel) shows a 1.5-dB increase per octave increase of signal frequency for frequencies up to about 64 Hz, and an increase of 3–4 dB for larger frequencies. Thus the shape of the curve is more similar to that observed in the data. Predicted thresholds are still systematically lower than in the data, but the maximum deviation amounts to about 6 dB instead of 9 dB in case of the original model. Second, the model accurately accounts for the threshold patterns (middle panel) for 5, 25, and 100 Hz (except for very large bandwidths; see below). The statistical criterion leads to a decreasing peak modulation threshold with increasing modulation frequency, in accordance with the experimental data. Finally, the statistical approach also accounts for the shape of the MTP (lower panel) even though the modulation filters used in the simulations have been determined in combination with the original decision criterion. However, the amount of masking in the patterns cannot be described correctly anymore. Masking is overestimated by as much as 8, 4, 3, and 1 dB for the signal frequencies 4, 16, 64, and 256 Hz, respectively.

Thus, it appears that the use of the statistical criterion allows an accurate description of the TMTF in various bandwidth conditions. In particular, the model predicts more realistic thresholds for low signal frequencies than the original approach does. At the same time, however, the statistical approach predicts too much masking for these low frequencies when an additional modulation masker is presented. Based on these results it cannot be finally decided which criterion might be more appropriate. It should be noted, however, that the *shape* of the masking patterns can be described well—independent of the decision criterion.

2. Single-channel versus multi-channel model

Predictions obtained with the multi-channel model of the present study hardly differed from those obtained with the single-channel model. Within the multi-channel model peripheral filtering only has an effect on AM threshold when the signal is strongly attenuated by the transfer characteristic of one or more channel(s) assuming Gammatone filters as peripheral filtering stage. The model predicts an increase of threshold for 256 Hz of about 2 dB relative to the single-channel prediction while thresholds remain unchanged for lower signal frequencies. This was obtained for a noise carrier in the range from 1 to 4 kHz; however, the results may change somewhat when the frequency region of the carrier is shifted, since the transfer characteristic of the peripheral filters changes. Instantaneous waveform compression was shown to have no influence on the results since both signal and inherent modulation of the carrier are transformed in the same way, resulting in the same signal-to-noise ratio at the output of the model. However, peripheral compression may have an effect in conditions of modulation masking produced by beating modulators. Moore *et al.* (1999) showed that the threshold for detecting 5-Hz signal modulation was affected by the presence of a pair of masker modulators beating at a 5-Hz rate. Their results suggested that such a nonlinearity introduces a distortion component in the internal representation of the envelope of the stimuli. The results could be fitted well using a model based on the concept of a modulation filterbank, following peripheral filtering, compression, and envelope extraction (Moore *et al.*, 1999). Compression might also influence results if one would assume a dynamic compression stage that acts differently for different modulation

rates, as realized in the adaptation stage within the auditory processing model by Dau *et al.* (1996a, b, 1997a, b). Such a stage compresses low-frequency envelope components more than high-frequency components. Its transfer characteristic, however, is more complex than that of a linear high-pass filter (for details, see Dau *et al.*, 1999).

3. Combination across peripheral filters

The present model assumes optimal combination of signal information across frequency. Independent of whether decisions were based on the envelope power or on its standard deviation, the model predicts a 3-dB decrease of threshold per doubling of the carrier bandwidth, as observed for large bandwidths ($\geqslant$250 Hz) in Figs. 11 and 13(b), respectively. This differs from the data which only show a smaller decrease of 1–2 dB. The conditions at the higher carrier bandwidths reflect situations in which several peripheral channels are excited by the stimulus while for smaller bandwidths activity is mainly restricted to the peripheral channel at or near 6 kHz. It appears that the real observer can combine signal information from different auditory filters but only less effectively than an ideal observer. Except for the study by Strickland and Viemeister (1997) who found a 3–4-dB effect, our data are consistent with several studies where smaller effects were observed (Maiwald, 1967; Viemeister, 1979; Eddins, 1993, 1999; Strickland, 1999). For example, Eddins (1999) and Strickland (1999) showed that the TMTF for bandpass noise centered at low frequencies is shifted towards higher thresholds relative to carriers (of the same bandwidth) centered at high frequencies. Furthermore, Eddins (1999) observed an increase in sensitivity for 8-Hz SAM of about 6 dB as the carrier bandwidth increased from 400 to 3200 Hz, corresponding to a 2-dB change per doubling the carrier bandwidth. Optimal detection would lead to a 9-dB difference. Thus, the results obtained in the present study as well as those from the other studies might be accounted for by assuming a suboptimal integration process across different peripheral channels. This observation is consistent with results from van den Brink and Houtgast (1990a, b) who showed that across-frequency integration in signal detection is *less* efficient for long-duration ($>$100 ms) broadband stimuli than for short-duration ($<$100 ms) broadband signals presented in a broadband noise masker. It appears that similar mechanisms play a role in modulation detection conditions. In the present study, the signal modulation (which is coherent across frequency) had a duration of 500 ms and deviations from optimal performance were observed for broadband carriers.

V. SUMMARY AND CONCLUSIONS

(i) Peaked masked-threshold patterns (MTPs) with about the same shape on a logarithmic frequency scale were obtained for signal modulation frequencies from 4 to 64 Hz using half-octave-wide noise as the modulation masker imposed on the same noise carrier as the signal. The shape of the MTPs did not change markedly when the carrier band was shifted towards higher frequencies. The shape of the MTPs corresponded well with the shape of psychophysical tuning curves (PTCs) obtained with the same masker modulation and carrier. The data obtained in the band-widening experiment also suggested a constant relative scaling of the modulation filter bandwidth but did not allow any assumptions about the underlying filter shape.

(ii) Within the framework of the envelope power spectrum model (EPSM) a modulation filterbank consisting of second-order bandpass filters and an additional first-order lowpass filter at 150 Hz was fitted to the data. The AM threshold was estimated by calculating the envelope power of the carrier and masker modulation passing the modulation filter tuned to the signal frequency. A constant Q-value of 1 for the bandpass filters accounted very well for the masking data. The overall lowpass filter accounted for the general decrease in sensitivity for frequencies larger than about 150 Hz.

(iii) The same model also accounted reasonably well for the shape of the TMTF obtained for carrier bandwidths in the range from 1 to 6000 Hz. However, for low signal frequencies the model generally underestimated thresholds in all bandwidth conditions by 3 to 5 dB. In addition, for large carrier bandwidths the model generally underestimated thresholds for all signal frequencies tested. Two possible explanations for these deviations were discussed. First, a decision criterion based on the standard deviation of the envelope power (instead of the average envelope power itself) was shown to account very well for the TMTF for low signal frequencies. Second, a ''sub-optimal'' signal-integration process across carrier frequency will probably account for the data for large carrier bandwidths.

(iv) Peripheral filtering influenced modeling results only in conditions when the signal modulation was strongly attenuated by the transfer function of the filters. Using Gammatone filters within the multi-channel EPSM, for a carrier ranging from 1 to 4 kHz, threshold was unaffected for signal frequencies up to 128 Hz and was increased by 2 dB for 256 Hz, relative to the predictions without peripheral filtering. Instantaneous waveform compression at a peripheral level did not have any effect on the simulations in the conditions tested in the present study.

ACKNOWLEDGMENTS

We thank Sid Bacon and two anonymous reviewers for very helpful comments on an earlier version of the manuscript. We also want to thank Birger Kollmeier, Jesko Verhey, Brian Moore, and Andrew Oxenham for critical remarks and suggestions. This work was supported by the Deutsche Forschungsgemeinschaft and the Max Kade Foundation.

APPENDIX: Envelope spectra of modulated stimuli

We consider a modulated waveform $s(t)$:

$$s(t)=c(t)(1+ma(t)), \tag{A1}$$

with $c(t)$ and $a(t)$ denoting the carrier and the modulator, respectively, and m representing the modulation depth. As long as the Fourier transform of the modulating waveform $A(f)=\mathcal{F}\{a(t)\}$ is restricted to frequencies below the lowest frequency component of the carrier spectrum $C(f)=\mathcal{F}\{c(t)\}$, the Hilbert envelope $e_s(t)$ of $s(t)$ can be expressed as

$$e_s(t)=|\hat{s}(t)|=|\hat{c}(t)||(1+ma(t))|, \tag{A2}$$

where $\hat{s}(t)$ represents the analytic signal of $s(t)$. The envelope spectrum $E_s(f_{\text{env}})$ is the Fourier transform of the envelope, $\mathcal{F}(e_s(t))$, and can be written as

$$E_s(f_{\text{env}})=E_c(f_{\text{env}})*\mathcal{F}\{|(1+ma(t))|\}, \tag{A3}$$

where $E_c(f_{\text{env}})$ represents the envelope spectrum of the carrier. Thus, the envelope spectrum of a modulated stimulus can be calculated via convolution of the carrier envelope spectrum and the spectrum of the absolute value of the modulator.

[1]This was done in an attempt to obtain about the same level for the masked thresholds at the peak frequencies, for all signal frequencies tested. Preliminary experiments showed that, in case of the LNN masker, this could be obtained with about a 3-dB higher masker level for the signal frequencies 64 and 256 Hz than for the signal frequencies 4 and 16 Hz, without producing overmodulation.

[2]If GN were used as the masker instead of LNN, in this experiment, masker levels larger than −10 dB would have led to overmodulation when combined with the signal. A slightly larger dynamic range in case of the GN masker may have been obtained with a decreased signal level. However, the signal level was already only 4 dB above threshold.

[3]Sinusoidal carriers do not exhibit inherent fluctuations like noise carriers. Thus modulation detection must be limited by *internal* noise.

[4]It should be noted that the current stimulus configuration represents a gated-carrier conditions. As was shown earlier (e.g., Viemeister, 1979; Sheft and Yost, 1990; Yost and Sheft, 1997), detection thresholds for very low modulation rates can be decreased when a continuous carrier is used instead of a gated one. The subjects of the present study did not show the typical band-pass shape of the TMTF, with increased thresholds at very low modulation rates. Possibly, these subjects might show such an effect for rates lower than 4 Hz, the smallest rate tested in the present study. In any case, the basis for these differences is still controversial. According to Sheft and Yost (1990), threshold differences between the two presentation modes still occur for ''fringe'' durations of the carrier up to as much as 500 ms, which cannot be explained in terms of neural adaptation due to the carrier onset. Whatever the reason for these effects might be, we assume that the *amount of masking*, reflecting the difference between masked and reference threshold, roughly remains the same for both presentation modes.

[5]We computed the envelope spectrum of a Gaussian noise carrier which was modulated by a bandlimited Gaussian noise (masker). The mean-to-sigma ratio of the envelope spectrum components, based on 10 000 independent signal realizations, was close to unity.

Bacon, S. P., and Viemeister, N. F. (**1985**). ''Temporal modulation transfer functions in normal-hearing and hearing-impaired listeners,'' Audiology **24**, 117–134.

Bacon, S. P., and Grantham, D. W. (**1989**). ''Modulation masking: Effects of modulation frequency, depth, and phase,'' J. Acoust. Soc. Am. **85**, 2575–2580.

Bacon, S. P., and Gleitman, R. M. (**1992**). ''Modulation detection in subjects with relatively flat hearing losses,'' J. Speech Hear. Res. **35**, 642–653.

Bos, C. E., and deBoer, E. (**1966**). ''Masking and discrimination,'' J. Acoust. Soc. Am. **82**, 708–715.

Chistovich, L. A. (**1957**). ''Frequency characteristics of masking effect,'' Biophysics (Eng. Transl.) **2**, 743–755.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997a**). ''Modeling auditory processing of amplitude modulation. I. Modulation detection and masking with narrowband carriers,'' J. Acoust. Soc. Am. **102**, 2892–2905.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997b**). ''Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration in modulation detection,'' J. Acoust. Soc. Am. **102**, 2906–2919.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996a**). ''A quantitative model of the ''effective'' signal processing in the auditory system: I. Model structure,'' J. Acoust. Soc. Am. **99**, 3615–3622.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996b**). ''A quantitative model of the ''effective'' signal processing in the auditory system: II. Simulations and measurements,'' J. Acoust. Soc. Am. **99**, 3623–3631.

Dau, T., Verhey, J. L., and Kohlrausch, A. (**1999**). ''Intrinsic envelope fluctuations and modulation-detection thresholds for narrowband noise carriers,'' J. Acoust. Soc. Am. **106**, 2752–2760.

Derleth, R. P., and Dau, T. (**2000**). ''On the role of envelope fluctuation processing in spectral masking,'' J. Acoust. Soc. Am., in press.

Derleth, R. P., Dau, T., and Kollmeier, B. (**2000**). ''Amplitude modulation detection in normal and hearing-impaired listeners,'' submitted to J. Acoust. Soc. Am.

Eddins, D. A. (**1993**). ''Amplitude modulation detection of narrow-band noise: effects of absolute bandwidth and frequency region,'' J. Acoust. Soc. Am. **93**, 470–479.

Eddins, D. A. (**1999**). ''Amplitude-modulation detection at low- and high-audio frequencies,'' J. Acoust. Soc. Am. **105**, 829–837.

Fleischer, H. (**1980**). ''Subjective Größe von Unterschieden im Amplitudenmodulationsgrad von Sinustönen,'' Acustica **46**, 31–37.

Fleischer, H. (**1982**). ''Modulationsschwellen von Schmalbandrauschen,'' Acustica **51**, 154–161.

Fletcher, H. (**1940**). ''Auditory patterns,'' Rev. Mod. Phys. **12**, 47–65.

Formby, C. C. (**1987**). ''Modulation threshold functions for chronically impaired maniere patients,'' Audiology **26**, 89–102.

Forrest, T. G., and Green, D. M. (**1987**). ''Detection of partially filled gaps in noise and the temporal modulation transfer function,'' J. Acoust. Soc. Am. **82**, 1933–1943.

Glasberg, B. R., and Moore, B. C. J. (**1992**). ''Effects of envelope fluctuations on gap detection,'' Hear. Res. **64**, 81–92.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Green, D. M., Berg, B. G., Dai, H., Eddins, D. A., Onsan, Z., and Nguyen, Q. (**1992**). ''Spectral discrimination of narrow-band sounds,'' J. Acoust. Soc. Am. **92**, 2586–2597.

Hall, J. W., Grose, J. H., and Haggard, M. P. (**1989**). ''Effects of flanking band proximity, number, and modulation pattern on comodulation masking release,'' J. Acoust. Soc. Am. **87**, 269–283.

Hall, J. W., 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 fluctuation and the masking of sine signals,'' 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**, 1310–1317.

Houtgast, T. (**1989**). ''Frequency selectivity in amplitude-modulation detection,'' J. Acoust. Soc. Am. **85**, 1676–1680.

Kay, R. H., and Matthews, D. R. (**1972**). ''On the existance in human auditory pathways of channel selectivity to the modulation present in frequency modulated tones,'' J. Physiol. (London) **225**, 657–667.

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,'' Research Monographs No. 35 (MIT, Cambridge, MA).

Kohlrausch, A., Fassel, R., and Dau, T. (**2000**). ''The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers,'' J. Acoust. Soc. Am., in press.

Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A., and Püschel, D. (**1997**). ''Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations,'' Acust. Acta Acust. **83**, 659–669.

Lawson, J. L., and Uhlenbeck, G. E. (**1950**). *Threshold Signals*, Vol. 24 of Radiation Laboratory Series (McGraw-Hill, New York).

Levitt, H. (**1971**). ''Transformed up–down procedures in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Lorenzi, C., Micheyl, C., Berthommier, F., and Portalier, S. (**1997**). ''Modulation masking with sensorineural hearing loss,'' J. Speech Lang. Hear. Res. **40**(1), 200–207.

Maiwald, D. (**1967**). ''The calculation of modulation thresholds with a model,'' Acustica **18**, 193–207.

McFadden, D. M. (**1986**). ''Comodulation masking release: Effects of vary-

ing level, duration, and time delay of the cue band,'' J. Acoust. Soc. Am. **80**, 1658–1667.
Moore, B. C. J., and Glasberg, B. R. (**1983**). ''Suggested formulae for calculating auditory filter bandwidths and excitation patterns,'' J. Acoust. Soc. Am. **74**, 750–753.
Moore, B. C. J., and Glasberg, B. R. (**1986**). ''The role of frequency selectivity in the perception of loudness, pitch and time,'' in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London), pp. 251–308.
Moore, B. C. J., and Oxenham, A. J. (**1998**). ''Psychoacoustic consequences of compression in the peripheral auditory system,'' Psychol. Rev. **105**, 108–124.
Moore, B. C. J., Sek, A., and Glasberg, B. R. (**1999**). ''Modulation masking produced by beating modulators,'' J. Acoust. Soc. Am. **106**, 938–945.
Oxenham, A. J., and Moore, B. C. J. (**1995**). ''Additivity of masking in normally hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **98**, 1921–1934.
Oxenham, A. J., and Plack, C. J. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listers with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.
Patterson, R. D., and Moore, B. C. J. (**1986**). ''Auditory filters and excitation patterns as representations of frequency resolution,'' in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London).
Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1987**). ''An efficient auditory filterbank based on the gammatone function,'' in paper presented at a meeting of the IOC Speech Group on Auditory Modeling at RSRE, 14–15 December.
Plack, C. J., and Oxenham, A. J. (**1998**). ''Basilar-membrane nonlinearity and the growth of forward masking,'' J. Acoust. Soc. Am. **103**, 1598–1608.
Pumplin, J. (**1985**). ''Low-noise noise,'' J. Acoust. Soc. Am. **78**, 100–104.
Richards, V. M. (**1992**). ''The detectability of a tone added to narrow bands of equal-energy noise,'' J. Acoust. Soc. Am. **91**, 3424–3435.
Rodenburg, M. (**1972**). ''Sensitivity of the auditory system to differences in intensity,'' doctoral thesis, Medical Faculty of Rotterdam.
Rodenburg, M. (**1977**). ''Investigation of temporal effects with amplitude modulated signals'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 429–437.
Rodenburg, M., Verschuure, J., and Brocaar, M. P. (**1974**). ''Comparison of two masking methods,'' Acustica **31**, 99–106.
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., 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.
Sheft, S., and Yost, W. A. (**1990**). ''Temporal integration in amplitude modulation detection,'' J. Acoust. Soc. Am. **88**, 796–805.
Strickland, E. A., and Viemeister, N. F. (**1996**). ''Cues for discrimination of envelopes,'' J. Acoust. Soc. Am. **99**, 3638–3646.
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.
Strickland, E. A. (**2000**). ''The effects of frequency region and level on the temporal modulation transfer function,'' J. Acoust. Soc. Am. **107**, 942–952.
Takahashi, G. A., and Bacon, S. P. (**1992**). ''Modulation detection, modulation masking, and speech understanding in noise in the elderly,'' J. Speech Hear. Res. **35**(6), 1410–1421.
Tanner, W. P., and Sorkin, R. D. (**1972**). ''The theory of signal detectability,'' in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York).
van den Brink, W. A. C., and Houtgast, T. (**1990a**). ''Efficient across-frequency integration in short-signal detection,'' J. Acoust. Soc. Am. **87**, 284–291.
van den Brink, W. A. C., and Houtgast, T. (**1990b**). ''Spectro-temporal integration in signal detection,'' J. Acoust. Soc. Am. **88**, 1703–1711.
Viemeister, N. F. (**1977**). ''Temporal factors in audition: A systems analysis approach,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 419–427.
Viemeister, N. F. (**1979**). ''Temporal modulation transfer functions based upon modulation thresholds,'' J. Acoust. Soc. Am. **66**, 1364–1380.
Wakefield, G. H., and Viemeister, N. F. (**1990**). ''Discrimination of modulation depth of sinusoidal amplitude modulation (SAM) noise,'' J. Acoust. Soc. Am. **88**, 1367–1373.
Yates, G. K. (**1990**). ''Basilar-membrane nonlinearity and its influence on auditory nerve rate-intensity functions,'' Hear. Res. **45**, 203–220.
Yost, W. A., and Sheft, S. (**1988**). ''Detecting amplitude modulation of sinusoidal carriers,'' J. Acoust. Soc. Am. Suppl. 1 **83**, S35.
Yost, W. A., and Sheft, S. (**1989**). ''Across-critical band processing of amplitude-modulated tones,'' J. Acoust. Soc. Am. **85**, 848–857.
Yost, W. A., and Sheft, S. (**1997**). ''Temporal modulation transfer functions for tonal stimuli: Gated versus continuous conditions,'' Aud. Neurosci. **3**, 401–414.
Yost, W. A., Sheft, S., and Opie, J. (**1989**). ''Modulation interference in detection and discrimination of amplitude modulation,'' J. Acoust. Soc. Am. **86**, 2138–2147.
Zwicker, E., and Feldtkeller, R. (**1967**). *Das Ohr als Nachrichtenempfänger* (Hirzel Verlag, Stuttgart).
Zwicker, E., Flottorp, G., and Stevens, S. S. (**1957**). ''Critical bandwidth in loudness summation,'' J. Acoust. Soc. Am. **29**, 548–557.

The use of visible speech cues for improving auditory detection of spoken sentences

Ken W. Grant[a] and Philip-Franz Seitz[b]
Walter Reed Army Medical Center, Army Audiology and Speech Center, Washington, D.C. 20307-5001

(Received 29 June 1999; revised 17 April 2000; accepted 12 June 2000)

Classic accounts of the benefits of speechreading to speech recognition treat auditory and visual channels as independent sources of information that are integrated fairly early in the speech perception process. The primary question addressed in this study was whether visible movements of the speech articulators could be used to improve the detection of speech in noise, thus demonstrating an influence of speechreading on the ability to detect, rather than recognize, speech. In the first experiment, ten normal-hearing subjects detected the presence of three known spoken sentences in noise under three conditions: auditory-only (A), auditory plus speechreading with a visually matched sentence (AV_M) and auditory plus speechreading with a visually unmatched sentence (AV_{UM}). When the speechread sentence matched the target sentence, average detection thresholds improved by about 1.6 dB relative to the auditory condition. However, the amount of threshold reduction varied significantly for the three target sentences (from 0.8 to 2.2 dB). There was no difference in detection thresholds between the AV_{UM} condition and the A condition. In a second experiment, the effects of visually matched orthographic stimuli on detection thresholds was examined for the same three target sentences in six subjects who participated in the earlier experiment. When the orthographic stimuli were presented just prior to each trial, average detection thresholds improved by about 0.5 dB relative to the A condition. However, unlike the AV_M condition, the detection improvement due to orthography was not dependent on the target sentence. Analyses of correlations between area of mouth opening and acoustic envelopes derived from selected spectral regions of each sentence (corresponding to the wide-band speech, and first, second, and third formant regions) suggested that AV_M threshold reduction may be determined by the degree of auditory-visual temporal coherence, especially between the area of lip opening and the envelope derived from mid- to high-frequency acoustic energy. Taken together, the data (for these sentences at least) suggest that visual cues derived from the dynamic movements of the fact during speech production interact with time-aligned auditory cues to enhance sensitivity in auditory detection. The amount of visual influence depends in part on the degree of correlation between acoustic envelopes and visible movement of the articulators.
[S0001-4966(00)03709-7]

PACS numbers: 43.66.Dc, 43.66.Mk [SPB]

I. INTRODUCTION

Numerous studies have demonstrated the benefits of auditory-visual speech perception over either listening alone or speechreading alone. The addition of visual cues to audition can be effectively equivalent to an improvement in the speech-to-noise (S/N) ratio of as much as 15 dB (Sumby and Pollack, 1954). Depending on the speech materials, each 1-dB improvement in S/N can correspond to a 5–10 percentage point increase in intelligibility (e.g., Miller *et al.*, 1951; Grant and Braida, 1991). Needless to say, the addition of speechreading can mean the difference between failure to understand and near-perfect comprehension, especially in noisy environments.

Summerfield (1987) hypothesized three possible roles for visual cues in improving speech understanding in noise. First, speechreading provides segmental (e.g., consonant and vowel) and suprasegmental (e.g., intonation, stress, rhythmic pattering, etc.) information which is *redundant* to cues provided acoustically (i.e., cues equally available to the auditory modality). Second, speechreading provides segmental and suprasegmental information *complementary* to cues provided acoustically (i.e., cues not available to the auditory modality, usually because of noise or other form of interference in signal clarity). Some speech elements, especially those of relatively long duration, can often be identified by speechreading alone (e.g., certain vowels, certain stress patterns, etc.). This type of visual speech information is usually redundant with auditorily derived speech information (provided that the acoustic signal is presented at a fairly favorable signal-to-noise ratio and is not overly distorted). Such information can be useful in reducing the attentional demands placed on the auditory channel and can increase the confidence of a particular recognition response. However, if visual and auditory modalities provide essentially the same information, that is, highly redundant, then only a small improvement in speech recognition occurs when the two modalities are combined (Grant and Walden, 1995; Grant *et al.*, 1998).

[a]Author to whom correspondence should be addressed; electronic mail: grant@tidalwave.net
[b]Present address: Foreign Service, U.S. Department of State.

Fortunately, complete redundancy between auditory and visual sources of information is rare. Instead, visual cues tend to complement auditory cues. In particular, acoustic cues for place-of-articulation that help distinguish most consonants are relatively easy to resolve through speechreading, but are the first to be disrupted by noise (Miller and Nicely, 1955) and hearing impairment (Walden *et al.*, 1975). In contrast, voicing cues are relatively robust acoustically, but nearly impossible to resolve through speechreading. Thus when audition and speechreading are combined, the integrated *AV* percept is typically far more complete and less ambiguous than that obtained from either of the unimodal conditions.

The third role of speechreading hypothesized by Summerfield (1987) is one that has received relatively little attention in the past and is the focus of this study. When a listener watches a talker speak, the acoustic speech signal and the visible movements of the talker's lips share common spatial and temporal properties which may help direct auditory analyses to the speech signal of interest rather than the surrounding background noise. With regard to spatial commonalities, seeing the location of the talker may help binaural processes localize and separate the target signal from the noise. For example, Spence *et al.* (2000) recently demonstrated that attending to visual lipreading from a spatial location proximal to a source of speech distractors makes it harder to hear speech targets coming from a distal location, even though the lipread motion was consistent with the target speech and not the distractor speech. Thus spatial proximity of acoustic and optical sources causes subjects to selectively attend to the distractor sound source even when instructed to ignore that source and attend to a second sound source coming from different direction.

In this study, however, we are interested mainly in the potential benefits of *temporal* rather than spatial coincidence between visible lip movements and speech sound modulations. Summerfield (1987) has shown that the displacements of the upper and lower inner margins of the lips at midline are ''generally loosely, and occasionally tightly, related to the overall amplitude contour [of the speech signal].'' If the visually perceived movements of the lips are related to the overall amplitude contour of speech as Summerfield suggests, then watching the changes in area of lip opening should serve as a cue for the acoustic envelope of speech. This might lead to a reduction in the temporal uncertainty associated with the onset of syllables and words similar to that observed in signal detection experiments when a light or other means is used to mark the onset of a signal masked by noise (Egan *et al.*, 1961a, 1961b; Watson and Nichols, 1976; Marshall and Jesteadt, 1986). Specifically, when a brief tone is presented in a background of noise, it is easier to detect the tone when the temporal interval in which the tone occurs is clearly marked (i.e., the tone is presented in a *defined observation interval*). When subjects have to attend to relatively long intervals without knowing exactly when the tone signal will occur (i.e., the tone is presented in an *undefined observation interval*), detection is significantly worse. The difference between the noise level that just masks a tone in these two types of presentations determines the amount of masking release. The amount of masking release due to the reduction in temporal uncertainty for tone detection has been shown to be as much as 2–3 dB (e.g., Watson and Nichols, 1976). A similar release from masking for the detection of speech when the peaks in the waveform are cued by speechreading might help explain some of the very large benefits to speech recognition under auditory-visual conditions. Considering the typical steepness of speech intelligibility functions relating S/N and intelligibility, a 2–3-dB improvement in masked thresholds for speech could lead to an intelligibility improvement of roughly 20–30 percentage points, or almost half of the improvement that is typically observed when comparing auditory and auditory-visual speech recognition.

Earlier studies by Frost *et al.* (1988), Frost (1991), and Repp *et al.* (1992) examined the influence of visual speech cues on the detection of acoustic speech signals. In the first two studies, printed words and nonwords were presented simultaneously with speech masked by amplitude-modulated noise. In the study by Repp *et al.* (1992), video recordings of words and nonwords showing the talker's face were presented concurrently with masked speech. In all three studies, speech detection thresholds were examined with respect to the influence of visual cues. When visual cues (orthographic or speechread) matched the acoustic signals, and the speech signals were real words, speech detection performance did not improve. Instead, there was a strong tendency for subjects to hear speech whenever a matching visual word was presented regardless of whether the word was presented acoustically or not, reflecting a strong bias toward perceiving speech. In terms of the Theory of Signal Detection (Luce, 1963), the bias index β (beta), but not the sensitivity index d' (*d*-prime), was affected by the visual manipulation.

The failure to demonstrate a change in sensitivity for the detection of speech when accompanied by visual cues in these earlier studies may have been related to the particular methodologies used and may not be a true indication of the potential benefits of speechreading to reduce temporal uncertainty and improve speech detection. Orthographically presented words are assumed to make contact with the subject's mental lexicon and evoke a memory representation of the acoustically presented word (Frost *et al.*, 1988). The correspondence between the temporal properties of a particular acoustic utterance and the memory representation of that same word, however, need not be highly correlated. Particular instantiations of a spoken word have natural variations in amplitude, duration, intonation, etc., and may have substantially different properties from the memory representations of words evoked from text. In order for visual speech cues to affect the auditory detection of speech signals, it may be that the two sources of information have to be highly correlated. Furthermore, because detection of speech is most likely precategorical (that is before any phonetic label is applied), the two sources of information (visual and auditory) need to be physically correlated at a level of analysis prior to lexical access if the visual cues are to have an influence on masked thresholds.

This criticism was addressed by Repp *et al.* (1992) by using videotaped speech productions of talker's faces rather than orthographic representations. Under these conditions,

perception of temporal comodulation between the visual and auditory percepts does not depend on lexical access (as it must with orthographic cues) and may be obtained directly at the level of the signal. However, certain properties of the masking noise used in the Repp *et al.* study, and technical limitations in aligning audio and video productions, may have favored the finding of a shift in bias at the expense of showing a change in sensitivity. In the Repp *et al.* study, a single-interval, yes/no paradigm was used in which subjects were asked to indicate whether a speech signal was present (or not) in a background of noise that was modulated by the envelope of the target speech signal. This type of noise has certain advantages in auditory masking studies. The speech-to-noise (S/N) ratio is held constant over time because the speech and noise have highly similar amplitude envelopes. However, in auditory-visual masking studies, this type of masking noise has enough speechlike properties to encourage response bias, thus making it more difficult to detect changes in sensitivity. In fact, numerous studies have shown that speech-modulated noise by itself is a substantial aid to speechreading (e.g., Grant *et al.*, 1991, 1994). Therefore, by using a masking noise that was modulated by the speech envelope, the changes of showing a *release from masking* or significant *masking protection* may have been greatly reduced since the speech target signal, the matched visual signal, and the noise masker all had the same temporal structure (Gordon, 1997a, 1997b). In the present study, we used a masking noise that did not share the temporal properties of the target speech signal. In this way, visual cues from speechreading represented the only source of cues that were temporally comodulated with the speech signal.

The test materials used by Repp *et al.* (1992) included three visual conditions: (1) the visual and acoustic speech signals were the same disyllable word, (2) the visual and acoustic speech signals were different disyllabic words, and (3) the visual signal was a still image o the talker's face. In all three cases, auditory recordings of the speech materials were digitized and realigned with the video productions. Because the equipment necessary to ensure exact auditory-visual synchrony was not available to these researchers, they used a manual switch to start and stop the audio output sequence until it appeared in synchrony with the visual stimulus. Subsequent measurements on the accuracy of these alignments revealed that the auditory and visual components were within ± 100 ms. Whereas this degree of precision may be suitable for speech recognition experiments (e.g., McGrath and Summerfield, 1985), it may not be suitable for auditory detection experiments. The failure to synchronize auditory and visual stimulus components accurately weakened the correlation between the acoustic and optical signals, which is likely to have weakened the amount of masking release (Hall *et al.*, 1984). In the present study stimulus alignment problems were eliminated by aligning precisely the digitized waveforms to the original sound tracks recorded on the optical disk.

II. EXPERIMENT 1: AUDITORY DETECTION WITH MATCHED AND UNMATCHED MOVING FACES

In experiment 1, the level of masking required to just detect the presence of three different sentences was determined independently for auditory-alone and auditory-visual conditions with matched and unmatched motion video. The use of sentence materials rather than isolated words offers many opportunities for the temporal correspondence between auditory and visual signals to emerge. The masker was a continuous white noise. Precise auditory-visual alignments within ± 2 ms were accomplished by specialized hardware and software which have been used in previous studies (Grant and Walden, 1996; Grant *et al.*, 1998). The primary question addressed was whether the detectability of a masked speech signal is improved by the addition of temporally comodulated visual speech information, or possibly reduced by the presence of incongruent visual stimuli.

A. Methods

1. Subjects

The subjects were ten normally hearing adults (four men, six women) with a mean age of 49 years (range = 27 to 75). Subjects were screened bilaterally to assure pure-tone air-conduction thresholds of $\leqslant$20 dB HL at audiometric test frequencies 0.25–4.0 kHz and $\leqslant$30 dB HL at 6.0 and 8.0 kHz (ANSI, 1989). All subjects were native English speakers with normal or corrected-to-normal vision (visual acuity equal to or better than 20/30 as measured with a Snellen chart). All subjects provided informed written consent before participating in the study, and were paid when eligible according to Walter Reed Army Medical Center guidelines.

2. Stimuli

Speech materials consisted of six sentences spoken by a female talker. Three of these were selected as target sentences with appropriately synchronized video. The video recordings of the remaining three sentences were used as incongruent (i.e., mismatched) visual speech stimuli as described below. The audio portions of the three nontarget sentences were not used in the experiment. The six sentences were drawn from the IEEE/Harvard (1969) sentence lists comprising 720 phonetically balanced low-context sentences. IEEE sentences were chosen because they are constructed to have roughly equal intelligibility and have approximately the same length, number of syllables, grammatical structure, and intonation. The sentences were videotaped and the auditory and visual images were transferred to an optical disk recorder (Panasonic TQ-3031F). The audio portion of each target sentence was digitized (16 bit A/D, 20 kHz sampling rate), normalized in level, and stored on a personal computer.

The three target sentences were ''To make pure ice, you freeze water'' (sentence 1), ''Both brothers wear the same size'' (sentence 2), and ''Watch the log float in the wide river'' (sentence 3). The unmatched video sentences were ''The barrel of beer was a brew of malt and hops,'' ''Nine men were hired to dig the ruins,'' and ''The node on the stalk of wheat grew daily.''

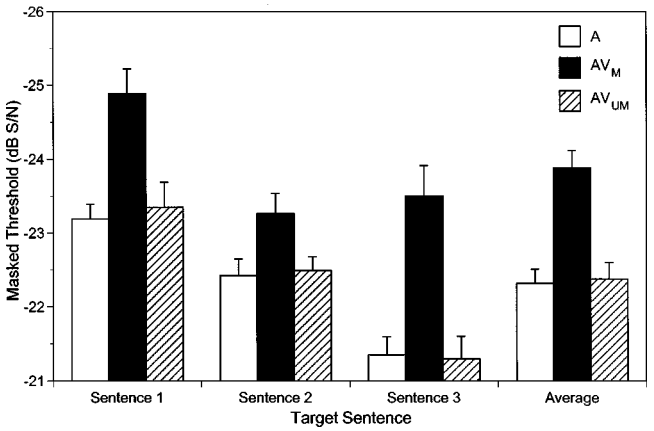


FIG. 1. Signal-to-noise (S/N) ratio for speech detection thresholds as a function of listening condition and target sentence. A=audio only; AV_M=matching video; AV_{UM}=mismatched video. Error bars show +1 standard error.

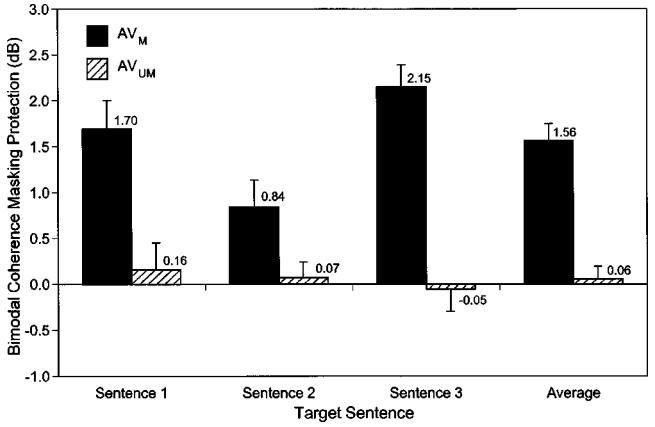


FIG. 2. Bimodal coherence masking protection in dB relative to the A only condition. AV_M=black bars; AV_{UM}=striped bars. Error bars show +1 standard error.

Masking noise consisted of a 10-sec sample of digitized (20-kHz, 16-bit) white noise that was stored on computer. During each trial of a two-interval forced-choice (2 IFC) procedure, a different random sample of this noise was presented. The duration of the noise masker on each trial was equal to the target sentence duration (between 2.5 and 3.0 sec long) plus a random amount ranging from 100–300 ms added equally to the beginning and end of the noise. The same duration of noise was presented on both target and nontarget intervals of a trial. When speech was presented, the target sentence was temporally centered in the noise. The purpose of the random leading and lagging edge of the noise was to introduce temporal uncertainty as to the start and end of the speech target signals, especially in the auditory-only conditions (see below).

3. Procedure

Subjects were tested binaurally with headphones (Beyer Dynamic DT770) in a sound-treated booth using an adaptive 2IFC procedure. Masked thresholds for detecting speech were obtained under three conditions: auditory alone (A), auditory-visual with matching visual stimulus (AV_M), and auditory-visual with unmatched visual stimulus (AV_{UM}). In each test block, there were nine interleaved tracks (3 conditions $\times$ 3 target sentences). In the AV_M condition, target audio sentences were presented along with simultaneous congruent visual speech information. In the AV_{UM} condition, target audio sentences were presented along with simultaneous incongruent visual speech information. The video from one of three nontarget sentences was selected randomly on each trial in the AV_{UM} conditions. For both AV conditions, the video speech information was identical in both observation intervals. Thus in the noise-alone interval, there was video plus white noise, whereas in the speech plus noise interval, there was the same video, plus white noise, plus target sentence. Video signals were displayed on a 19-in color monitor (SONY PVM 2030) positioned approximately 1.5 m from the subject.

The subject's task was to identify the interval containing the target sentence. The speech signal level was held constant at approximately 50 dB SPL when measured in a 2-cc coupler (Larsen-David Model 800B, linear weighting). The intensity of the noise masker was varied independently for each track according to a three-up, one-down adaptive tracking procedure targeting the 79% point on the psychometric function (Levitt, 1971). Each track was controlled independently and selected randomly. The initial step size of the digital noise attenuator (TDT PA4) was 3 dB and the final step size was 1 dB. Threshold estimates for each track were computed as the mean of the noise levels between reversal points for each of the last six reversals in the direction of the track obtained with the smaller step size. Final threshold values for each of the target sentences in each of the three conditions were the average of three separate threshold estimates. If the standard error of the three estimates exceeded 1 dB a fourth estimate was obtained and the final threshold value was the average of all four estimates.

III. RESULTS AND DISCUSSION

Figure 1 shows the average masked thresholds for each condition and each sentence in terms of the S/N ratio averaged across subjects. Unlike the earlier study by Repp *et al.* (1992), for each of the three target sentences there was a reduction in masked threshold for the AV_M condition relative to the other two conditions. This is indicated in the figure by noting that the AV_M threshold values are associated with the poorest S/N (i.e., the greatest amount of noise). No difference in S/N was observed between the A and AV_{UM} conditions. That is, the presentation of incongruent visual information appears to have had little effect on auditory detection thresholds, suggesting that subjects may be able to weight the auditory information more heavily under conditions when the visual information is uninformative.

Using the detection thresholds obtained from the A condition as a reference, we computed the average masking difference for the AV_M and AV_{UM} conditions, as well as the masking difference for each of the three target sentences (Fig. 2). A masking level difference or *bimodal coherence masking protection* (BCMP) was observed for all three sentences in the AV_M condition (black bars). In contrast, there was no appreciable BCMP for the AV_{UM} condition (striped

bars). The average improvement in masked thresholds for the three target sentences observed in the AV_M condition was 1.56 dB. The range of improvements across sentences was 0.84 (sentence 2) to 2.15 (sentence 3) dB.

To assess the statistical significance of these effects, a repeated-measures analysis of variance (ANOVA) was run with condition and sentence as within-subject factors. The main effects of condition [$F(2,18)=51.20$, $p<0.0001$] and sentence [$F(2,18)=37.64$, $p<0.0001$] were significant. The interaction of condition and sentence was also significant [$F(4,36)=5.19$, $p=0.002$]. *Post hoc* comparisons confirmed that the A and AV_{UM} conditions required a significantly higher speech-to-noise ratio than did the AV_M condition for all three sentences. The amount of BCMP observed in the AV_M condition was significantly smaller for sentence 2 than for either sentence 1 or sentence 3. Finally, the difference in BCMP observed between sentences 1 and 3 was not significant.

These data show that cross-modality comodulation can offer protection from noise masking in much the same manner as shown by Gordon (1997a, 1997b) for acoustic only stimuli. The magnitude of this protection (between 1 and 2 dB, depending on the particular target sentence) is consistent with the studies by Gordon, as well as early studies that show threshold improvements when a light or other visual signal is used to mark the onset of an acoustic signal masked by noise (e.g., Egan *et al.*, 1961a, 1961b; Watson and Nichols, 1976; Marshall and Jesteadt, 1986). These studies showed improved detection thresholds on the order of 2–3 dB either by marking observation intervals in time, or by presenting additional signals in the noise alone and noise plus signal intervals that were comodulated with the target. The slightly smaller effect of speechreading facilitation on auditory detection thresholds obtained for these three known sentences might be expected because the degree of coherence between the movement of the lips and the amplitude envelope of the speech signals is determined by a loosely coupled system (Summerfield, 1987). This last point will be revisited in greater detail in the general discussion.

IV. EXPERIMENT 2: AUDITORY DETECTION WITH AND WITHOUT MATCHING ORTHOGRAPHY

The masking protection of auditory speech signals offered by matched speechread information may be thought of in similar terms as the phenomenon of coherence masking protection (CMP) described by Gordon (1997a, 1997b). Similar in many respects to comodulation masking release (CMR), CMP occurs when coherence between two signals (one clearly perceptible and the other barely audible) acts to make the near-threshold signal stand out against a background masker. In the bimodal version of CMP, the movements of the lips and other surface features of the face and tongue provide information pertaining to the temporal and spectral properties of the target speech signals. The temporal information probably relates to the timing location of amplitude peaks and valleys in the time-amplitude envelope, whereas the spectral information derives from the shape of the mouth and front cavity of the vocal tract. Exactly how much information about the temporal and spectral envelope can be gleaned via speechreading is not clear, although a recent study by Yehia *et al.* (1998) suggests that 70%–80% of the variance in the rms amplitude can be recovered by nonlinear transformations of facial motion.

Before pursuing this explanation in greater detail, however, it is important to rule out a much simpler interpretation of the first experiment, namely that the BCMP observed under matched *AV* conditions was due simply to knowing what the target sentence was on each AV_M trial. With only three target sentences, it is likely that, at the beginning of each AV_M trial, the subjects received enough visual cues to know unambiguously which target sentence was being presented. This was especially likely when the target was presented in the second interval, giving the subject a full look at the moving face during the noise-alone interval. In contrast, in the A condition, there was much greater uncertainty as to which target was being presented. If the subject knows the identity of the target sentence beforehand, it is possible that he or she can anticipate the approximate temporal location of amplitude peaks, thus reducing the stimulus uncertainty. This possibility was explicitly addressed in a second experiment where the identity of the target sentence was given via a written display of the text prior to each trial.

A. Methods

1. Subjects

Six of the ten adult subjects who had participated in the first experiment participated in the second experiment. The subject's age ranged from 27–53 years. All subjects had normal hearing (ANSI, 1989) and normal or corrected-normal vision.

2. Stimuli

The same three target sentences used in the first experiment were used again. There were six interleaved tracks: three auditory (A) tracks corresponding to each of the three target sentences, and three audio plus orthography (AV_O) tracks. The noise masker and positioning of the target with respect to the noise were identical to the first experiment. The only difference between A and AV_O trials was that the text of the target sentence was displayed for 0.5 sec prior to the start of every AV_O trial. Thus the subjects were told explicitly which target sentence would be presented. In this sense, the subject's knowledge of the identity of the target sentence was assured in this experiment.

3. Procedure

The procedures for this experiment were identical to those used in the first experiment. The only differences were that there were no mismatched *AV* trials and that each test block consisted of six (rather than nine) interleaved tracks (three auditory target sentences and the same three target sentences preceded by an orthographic string). When orthography was present, it was displayed simultaneously on the subject's touch screen monitor as well as the video monitor prior to the onset of each AV_O trial.

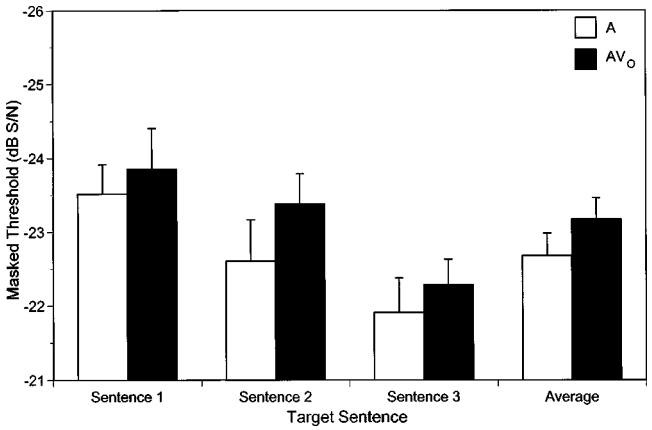


FIG. 3. Signal-to-noise (S/N) ratio for speech detection thresholds as a function of listening condition and target sentence. A=audio only; AV_O=matching orthography. Error bars show +1 standard error.

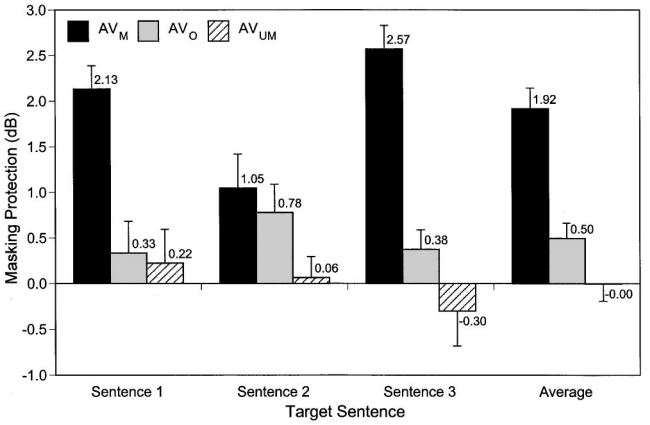


FIG. 5. Masking protection in dB ($n=6$) comparing matched speechreading (AV_M), matched orthography (AV_O), and unmatched speechreading (AV_{UM}). Error bars show +1 standard error.

V. RESULTS AND DISCUSSION

The average masked thresholds (expressed in dB S/N) for the A and AV_O conditions, and the average magnitude of the masking protection (in dB) for the AV_O condition, are displayed in Figs. 3 and 4, respectively. There was a small but significant reduction in masked thresholds when the subjects knew the identity of the target sentence. The magnitude of the masking protection for each sentence varied between 0.33 and 0.78 dB, with a mean of 0.5 dB.

To assess the significance of these results, and to compare the reduction in masked thresholds for matched speechreading and matched orthography conditions, we reexamined the data from the first experiment for the six subjects who participated in experiment 2. Before comparing the data across the two experiments, it was important to establish that the A thresholds for the six subjects (essentially a replication across the two experiments) did not show any effects of learning or context (nine interleaved tracks versus six). A repeated-measures analysis of variance (ANOVA) was run with test and sentence as within-subject trial factors. The main effect of test (experiment 1 versus experiment 2) was not significant [$F(1,5)=0.15$, $p=0.71$], nor was the interaction between test and sentence [$F(2,10)=0.26$, $p=0.77$]. The main effect of sentence was significant [$F(2,10)=24.72$, $p<0.001$], showing for both experiments, that sentence 1 was the easiest to hear and sentence 3 was the hardest to hear. Thus for these three target sentences and six subjects, there was no evidence of a change in A masked thresholds across the two experiments.

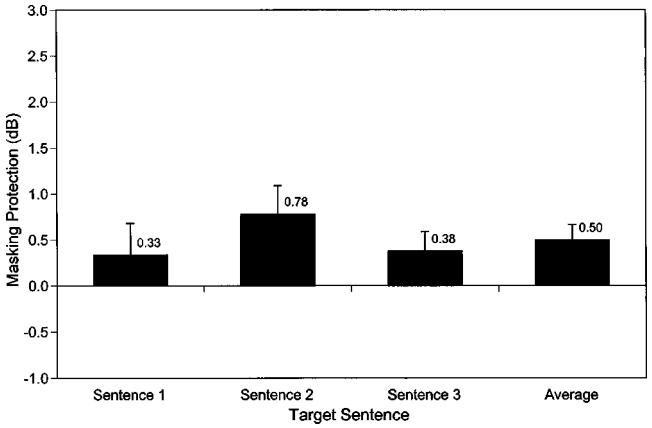


FIG. 4. Masking protection in dB for the orthography (AV_O) condition ($n=6$). Error bars show +1 standard error.

Analysis of the results of experiment 2 showed that there was a significant effect of condition (A versus AV_O), with the AV_O condition requiring greater masker levels at threshold [$F(1,5)=35.79$, $p=0.002$]. Similar to experiment 1, there was also a significant effect of sentence [$F(2,10)=32.92$, $p<0.001$]. However, unlike the results from the first experiment, there was no significant interaction between condition and sentence [$F(2,10)=0.50$, $p=0.62$], suggesting that the amount of threshold reduction was approximately equivalent across sentences.

Given the absence of significant differences between the auditory thresholds across the two experiments, a direct comparison of masking protection for the speechreading conditions of experiment 1 (match and unmatched) and orthographic condition from experiment 2 was carried out on the six subjects who participated in both experiments. Figure 5 shows the results. The average BCMP provided by matched speechread information was significantly greater than the masking protection provided by orthography, even though the orthographic display was less ambiguous with regard to the identity of the target sentence. Furthermore, the fact that the amount of masking protection due to orthography was not dependent on the target sentence (unlike the BCMP for speechreading) suggests that the processes that lead to masked threshold reductions for speechreading and for orthography may be different (at least under the condition employed here). In the orthography condition, the subject is provided with explicitly knowledge of the sentence identity prior to the presentation of each trial. Because the sentences are mostly audible at the beginning of the track (when the S/N is the most favorable), the subject may also learn the cadence and intonation of each specific sentence production. Thus it may be possible for the subject to say the sentence internally while the noise masker is presented. However, the exact starting time of the target sentence is not known due to the random duration of leading noise before the sentence

begins. Apparently, this knowledge can be applied equally to any target sentence (hence the lack of interaction between condition and sentence),[1] and provides a small reduction in stimulus uncertainty resulting in a significant threshold improvement. On the other hand, whereas matched speechreading may be more ambiguous with regard to the sentence identity, it is more explicit with regard to the temporal (and possibly spectral) structure of the sentence.[2] The correspondence between facial movements and the location of acoustic amplitude peaks in the sentence may provide the listener with timing cues as to when to listen. The fact that the BCMP in the matched speechreading condition depended on the particular target sentence suggests that the degree of bimodal coherence may vary across sentences. Assuming that the peaks in the amplitude envelope of the target sentences are the primary cues for auditory detection, we hypothesized that when the bimodal correspondence between facial movements and amplitude envelope is high, a large BCMP will result. In contrast, when the correspondence is low, a small BCMP will result. Stated differently, the degree of correlation between visible movement of the articulators and variations in acoustic energy over time should differ for different sentences, being greater for those sentences that give rise to more BCMP. This hypothesis was evaluated in a third experiment.

VI. EXPERIMENT 3: CORRELATION BETWEEN ACOUSTIC AND VIDEO MEASURES

A correlation analysis of facial movement and amplitude envelope was performed for the sentence which produced the greatest BCMP (sentence 3) and the sentence which produced the smallest BCMP (sentence 2) in experiment 1. Individual video frames (frame length equal to 33.33 ms) from sentence 3 (''Watch the log float in the wide river'') and sentence 2 (''Both brothers wear the same size'') were digitized and analyzed for area of mouth opening, height of mouth opening, and width of mouth opening. All measures were made using SigmaScan Software. Although there are many potentially important parameters of the face that could have been extracted and used, the area of lip opening was chosen for the subsequent correlational analyses because of the work by Fairbanks (1950) and Stevens and House (1955, 1961) showing that cross-sectional area of the front portion of the vocal tract (controlled primarily by the mandible and lips) is proportional to intensity.

Prior to acoustic envelope extraction, the sentences were filtered into four distinct spectral regions corresponding roughly to $F1$ (100–800 Hz), $F2$ (800–2200 Hz), and $F3$ (2200–6500 Hz) formant regions, as well as a wide-band condition (100–6500 Hz). The rms output from the resulting filtered waveforms was computed for successive 33.33-ms time intervals, corresponding to the timing of each video frame. The reason for examining the amplitude envelopes from different bands of speech was twofold. First, the energy fluctuations in each sentence are different for different bands of speech. And second, based on past results from speechreading experiments showing that subjects extract primarily place-of-articulation cues (known to be associated primarily with $F2$ and $F3$ transitions), it is quite possible that speechreading provides selective information about the amplitude envelope of speech, primarily that part conditioned by the higher formant frequencies associated with the cross-sectional area of the front cavity. Thus in addition to our main prediction that sentences with greater BCMP will have higher correlations between area of lip opening and acoustic envelope than sentences with smaller BCMP, we also predict that visible lip area functions will be correlated best with the $F2$, $F3$ envelope bands and correlated least with the $F1$ envelope band.

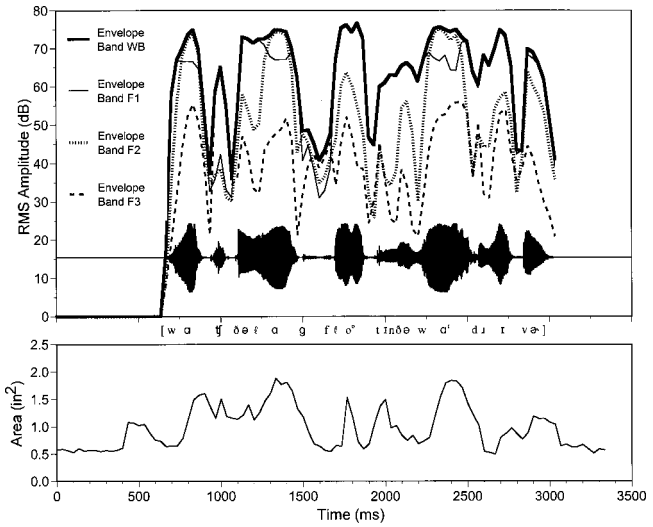


FIG. 6. Amplitude-time waveform, rms envelope functions, and area of mouth opening for sentence 3, ''Watch the log float in the wide river.'' *WB* envelope band (100–6500 Hz); $F1$ envelope band (100–800 Hz); $F2$ envelope band (800–2200 Hz); $F3$ envelope band (2200–6500 Hz).

The time-intensity waveform, amplitude envelopes, and area of mouth opening for sentence 3 and sentence 2 are displayed in Figs. 6, and 7, respectively. For sentence 3, the peak in the wide-band envelope function occurs on the word ''float'' with several other peaks of slightly less amplitude on the words ''wide,'' ''log,'' and ''watch.'' For sentence 2, the peak in the wide-band envelope occurs on the first syllable of the word ''brothers,'' with secondary peaks on the

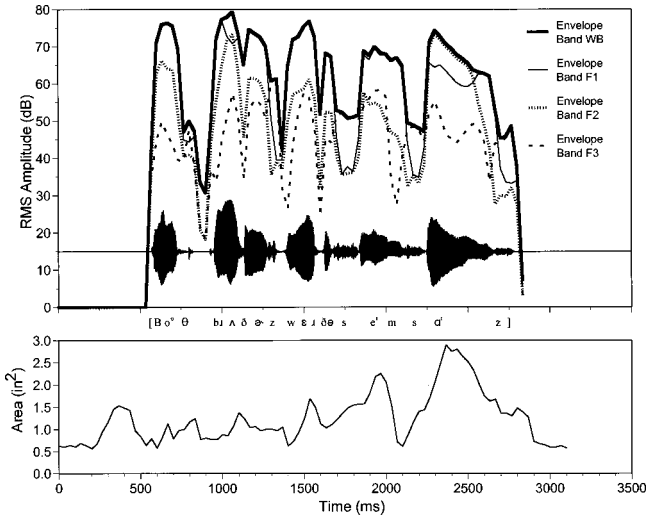


FIG. 7. Same as Fig. 3 except for sentence 2, ''Both brothers wear the same size.''

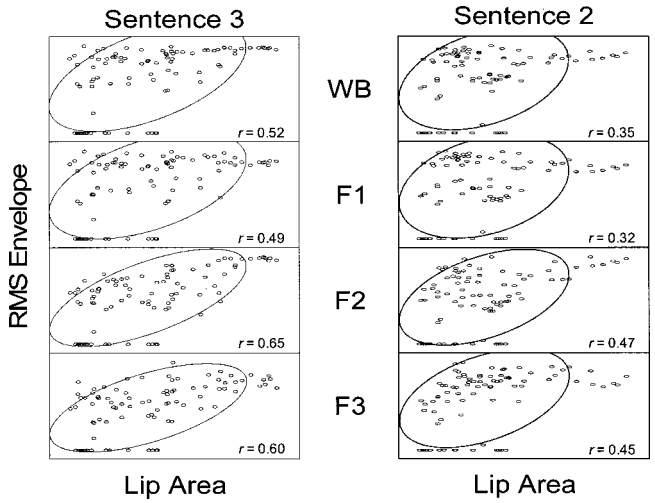


FIG. 8. Scatter plot showing the degree of correspondence between area of mouth opening and rms envelope functions for sentence 3 and sentence 2. A 75% ellipse assuming a Gaussian bivariate distribution is also shown along with the Pearson product moment correlation (after appropriate time shifts, see text). All correlations are significant ($p<0.01$). Within each sentence, correlation with $F2$ envelope was best, whereas correlation with $F1$ envelope was worst. Differences in correlations across sentences for each envelope function are significant ($p<0.05$).

words ''both'' and ''wear.'' The location of these peaks probably marks the time when the sentence first emerges from the noise masker and is detected. According to our working hypothesis, the correlation between area of mouth opening and envelope should be higher for sentence 3 than for sentence 2, especially in the region of the envelope peak, and higher in the $F2$ and $F3$ region that in the $F1$ region.

Given that visual movement of the speech articulators generally precedes acoustic output, we used cross-correlation plots between the are function and envelope functions for each sentence to help identify the lag (in frames) that resulted in the maximum correlation. Lags of up to ± 5 frames ($\approx$165 ms) were deemed appropriate based on previous data on the effect of *AV* asynchrony on speech intelligibility (Grant and Seitz, 1998a) which show that, on average, subjects are able to tolerate about 150–200 ms of audio delay relative to the video signal without a detrimental effect on intelligibility. Inspection of the cross-correlation plots for each of the two sentences revealed that the maximum correlation was obtained with a lag of -1 frame, or about 33 ms.

Figure 8 shows scatter plots and Pearson product-moment correlations between area of lip opening and rms envelope (with a -1 frame lag between *V* and *A* signals) for sentence 3 and sentence 2. As predicted, sentence 3 had a stronger correlation than sentence 2 for each of the four frequency bands tested. The frequency band showing the highest correlation for both sentences was the $F2$ envelope band, whereas the frequency band showing the lowest correlation was the $F1$ envelope band. Differences between correlations for each envelope function across sentences were significant ($p<0.05$), as were differences between correlations for $F2$ and $F1$ envelope functions within sentences (Bruning and Kintz, 1968).

The correlational analyses between lip-area function and rms amplitude envelopes were conducted on only two sentences, making it difficult to generalize the claims to other materials. To address this concern, we analyzed the relationship between lip movement and acoustic amplitude envelope using additional speech materials from the University of Wisconsin x-ray Microbeam database (Westbury, 1994). In particular, we chose a single 30-sec passage produced by a male speaker in which movement data were obtained from the upper and lower lip along with the tongue and jaw. For the purposes of this analysis, the passage was partitioned into six segments ranging between 2 and 9 sec in length and the inter-lip distance in the midsagittal plane was extracted every 34 ms. Inter-lip distances were chosen because our video measures of lip area for the two sentences studied indicated that the lip area and vertical lip distance were highly correlated. The acoustic utterances were filtered as described earlier into the four bandpass regions (*WB*, $F1$, $F2$, and $F3$). To adjust for temporal asynchronies that were introduced by the different sampling rates of acoustic and kinematic records and by the processing used to extract the acoustic speech envelopes, a cross-correlational analysis using the *WB* envelope and inter-lip distance was performed for each speech segment separately to determine the time lag (± 5 frames) yielding the strongest correlation. This lag (usually 2 or 3 frames) was then used in computing the Pearson correlation between the rms acoustic envelopes (averaged over a 34-ms window) and inter-lip distance over the entire length of each of the six utterances. As observed earlier for correlations between lip-area functions and acoustic envelope (see Fig. 8), the correlations between inter-lip distance and acoustic amplitude envelope showed, in every case, that the strongest correlations were found for envelopes extracted from the $F2$ and $F3$ regions ($0.39<r<0.54$), whereas the weakest correlations were found for envelopes extracted from the $F1$ region ($0.05<r<0.42$). Thus using either lip-area functions or inter-lip distance functions as a measure of facial kinematics, it appears that there is a consistent relationship between acoustic energy and lip movement, especially in the higher frequency regions above 2200 Hz.

As noted earlier, there was no effect on masked thresholds, either positive or negative, in the AV_{UM} conditions. According to the reasoning above, this could have been due to low correlations between acoustic amplitude envelopes and lip-area functions. To estimate the correlation for the AV_{UM} condition, we computed the correlation between the lip-area function for each nontarget sentence and the envelope functions derived from sentences S2 and S3 separately and averaged the results. The average AV_{UM} correlations ranged from 0.32 to 0.41, which were significantly lower ($p<0.05$) than the correlations for the AV_M condition with sentence 3 for all four envelope functions, and for sentence 2 for the $F2$ envelope function. However, because there was considerable overlap between these unmatched correlations and the matched correlations for sentence 2, it is reasonable to ask why the AV_{UM} conditions were unsuccessful in generating any BCMP when significant BCMP was observed for sentence 2. To answer this question, the degree of coherence between acoustic and optical signals over much shorter time frames than the 2–3 sec comprising each of the target sentences was computed. The rationale and results of these

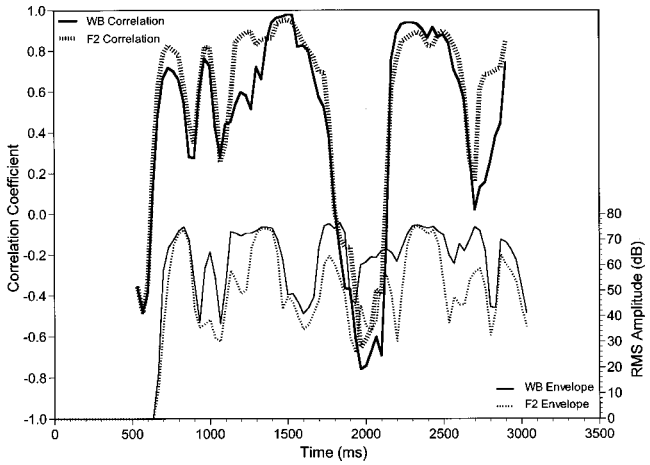


FIG. 9. Ten-point moving correlation between area of mouth opening and *WB* and $F2$ envelope functions for sentence 3. The left axis shows the Pearson product-moment correlation. The right axis shows the rms amplitude in dB. Heavy black line and heavy dashed line=correlation coefficients for *WB* and $F2$ envelopes, respectively; thin black line=*WB* envelope; thin dashed line=$F2$ envelope. Note the high correlation in the vicinity of the envelope peaks.

analyses are discussed in greater detail below.

In the detection experiment (experiment 1), the masked thresholds were presumably based on relatively brief segment(s) of each sentence having the greatest S/N. Thus high correlations across the entire sentence may not be necessary for obtaining BCMP. Perhaps what is more important is the degree of local correlation at or near the envelope peaks for each sentence. In other words, if the correlation between area of mouth opening and envelope is high in the region of envelope maxima, we might expect a relatively large amount of BCMP. On the other hand, if the correlation between area of mouth opening is low, we might expect a relatively small amount of BCMP.

Local correlations between area of mouth opening and envelope were computed by selecting the first 10 frames from each function (333 ms), calculating the Pearson product moment correlation, and then sliding the 10-point window by 1 frame and repeating the calculation. Plots of this moving 10-point correlation are shown in Figs. 9 and 10 for sentence 3 and sentence 2, respectively. Also shown in the figures are the wide-band and $F2$ envelope function so that the points of correlation maxima can be compared to points of envelope maxima. For convenience, only the correlation with the *WB* and $F2$ envelope functions are displayed.

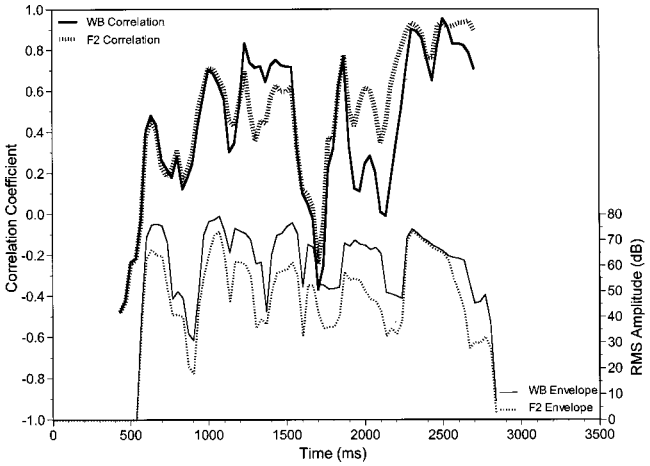


FIG. 10. Same as in Fig. 9 except for sentence 2. Note the relatively low correlation in the vicinity of the envelope peaks, especially at the beginning of the sentence.

In Fig. 9 (sentence 3), the local correlations are quite high, especially for the $F2$ envelope ($r>0.8$) in the region of the maximum envelope peak (''log float''), as well as at several other envelope peaks. In comparison, the local correlations shown in Fig. 10 (sentence 2) were lower ($r<0.75$) near the maximum envelope peaks (''both brothers''). Note also, that even though the local correlations for both the *WB* and $F2$ envelope bands reached a high value near the end of the sentence on the word ''size,'' this high degree of comodulation between the area of mouth opening and envelope may be of little use because the overall energy at the end of the sentence is approximately 5 dB less than that at the beginning of the sentence. Subjects may not have been able to take advantage of this cross-modality correspondence because the S/N at the end of the sentence was too low compared to more intense parts of the sentence.

The hypothesis that the local correlation between lip-area function and acoustic envelope has to be high (e.g., >0.8) in the vicinity of an energy peak in order to achieve a significant BCMP may provide a clue as to why the AV_{UM} conditions failed to produce any masking protection. As noted earlier, correlations computed over the entire length of the sentence for the AV_{UM} conditions were not that different from correlations computed for the AV_M conditions with sentence 2 (the sentence producing the least amount of, but nevertheless significant, BCMP). However, when local correlations were computed between sentence 2 (or sentence 3) as the acoustic target and lip-area functions derived from the mismatched sentences, none of the correlations occurring in the vicinity of the highest envelope peaks exceeded 0.6. Another factor that may be important in explaining the lack of BCMP for the AV_{UM} conditions is that the coherence between lip-area function and acoustic amplitude envelope were somewhat unreliable due to the methodology used in setting up these conditions. Because each acoustic target was paired randomly on every trial with one of three mismatched video productions, the coherence between video movement and acoustic energy fluctuations of any particular acoustic target sentence varied from trial to trial. Thus this on-again, off-again coherence between optical and acoustic components in the AV_{UM} conditions was probably insufficient to produce a reliable effect on speech detection thresholds.

VII. GENERAL DISCUSSION AND CONCLUSION

In speech, the visible modulation of the area of mouth opening (or inter-lip distances) is correlated, to varying degrees depending on the particular sentence, with the modulation of overall amplitude. Our measures extend the earlier work by Summerfield (1987) by highlighting the relation between the area of mouth opening and the acoustic energy modulations in the $F2$ and $F3$ regions. This correspondence is exactly what one might predict given speechreaders' abilities to extract primarily place-of-articulation information

from visible oral and facial movements, known to be carried by information in the $F2$ and $F3$ spectral regions. Thus it appears that the information contained in the $F2$, and to a lesser extent $F3$ region of speech, is partially visible to the speechreader through variations in the lip kinematics over time.

One account of the possible benefit of this cross-modality comodulation may be to inform the observer that the visual and acoustic information are bound to the same articulatory event and should be processed together. By observing the visual stimulus in the AV_M condition, the subject is alerted to temporal, and possibly spectral, locations of the acoustic noise-plus-speech stimulus where the S/N is most favorable for detecting the speech. By this account, visually congruent speech information may serve to direct auditory attention, thereby reducing temporal and spectral uncertainty. In recent publications, the reduction in detection threshold as a result of adding a second signal to both noise alone and noise-plus-signal intervals that is comodulated with the target signal has been termed coherence masking protection (Gordon, 1997a, 1997b).

Gordon (1997b) offered two accounts to explain the phenomenon of coherence masking protection (CMP) for auditory alone stimulation. The first was an adaptation of earlier CMR explanations whereby the detection of peaks in the comodulated signal (in the present case, the visual speechread signal) facilitates the detection of peaks in the target signal. The second account of CMP is based on the auditory system's ability to detect correlations in envelope fluctuations across different frequency bands (Richards, 1987). In the case of BCMP, this would imply that listeners can also detect correlations in envelope fluctuations across different modalities. Thus the detection of significant cross-modal comodulation could be used as a cue for selecting the interval containing the target sentence as opposed to the noise alone, and the amount of BCMP would be conditioned by the degree of cross-modal correlation.

Recent neurophysiological findings have demonstrated that auditory cortex is greatly influenced by speechreading (Sams *et al.*, 1991; Calvert *et al.*, 1997). Further, there is a wide variety of multisensory cells in the superior colliculus whose firing rates depend on the degree of convergence of stimuli arriving from the different senses (Meredith *et al.*, 1987; Stein and Meredith, 1993). Based on the evidence that the nervous system is sensitive to converging input from multiple senses, the second of the two explanations offered by Gordon (1997b) seems preferable. Synchrony of auditory and visual speech cues appears to be an important characteristic in determining the magnitude of BCMP Multisensory neurons described by Meredith *et al.* (1987) also have an optimum stimulus onset asynchrony, with some cells preferring simultaneous A and V input, and others preferring A input to be desynchronized by as much as 50–100 ms. The specifics of the temporal preference varies with neurons and may depend on the populations in different areas of the brain. It is quite possible that populations involved in speech perception prefer synchrony. The close proximity between talker and listener under most communication settings, and especially during development in the first years of language learning, means that auditory and visual speech cues will be highly synchronized. It is therefore not unreasonable to assume that multisensory cells subserving speech perception might be tuned preferentially to synchronous inputs. In addition, multisensory cells in the superior colliculus tend to be maximally excited under conditions of weak unimodal input, a principle referred to as inverse effectiveness. Thus under synchronized AV conditions with a relatively weak or degraded auditory target signal, as in the present experiments, we can assume that the activity from these multisensory cells is enhanced for the signal-plus-noise interval relative to the noise-alone interval, leading to a different pattern of sensory activation. Furthermore, the greatest activation of these cells would presumably occur in the AV_M condition where the correlation between the A and V modality is highest. The result of this increase in AV activity would be an enhanced ability to extract an acoustic signal from a noise background, as evidenced by the significant BCMP.

The results from the present investigation must be viewed with some caution, however, because of the limited number of familiar target sentences studied. This is especially true for the results presented in experiment 3 regarding the correlation of lip-area functions with acoustic envelope functions derived from selected spectral regions of speech. As a way to generalize these particular findings, one could measure the correlation between area of lip opening and acoustic envelope for a wider range of materials and then, on the basis of those measures, categorize speech tokens into those that would be the most likely to produce either high or low BCMP. Predictions could then be verified by obtaining masked thresholds under A and AV_M conditions. In spite of these limitations, the main result of the present investigation, that speechreading can improve auditory detection, suggests a role for speechreading that is not usually considered in models of auditory-visual integration. Most models of auditory-visual integration assume independence between the auditory and visual sources of information (Massaro, 1987, 1998; Braida, 1991). A more recent model presented by Massaro (1998) includes a third independent source of information which is itself cross modal. Information from this domain includes temporal asynchrony between auditory and visual sources, and intermodal timing cues associated with voicing (Massaro, 1998). Massaro is quick to point out, however, that the presence of cross-modal information does not violate the assumption of independence at the evaluation stage. It is simply treated as another source of independent information and integrated in the same multiplicative manner as assumed by the fuzzy logical model of integration (FLMP). Although the FLMP has been mostly used to model suprathreshold speech identification, the basic framework appears capable of modeling the present results. Masked detection thresholds under synchronous, congruent AV conditions would be predicted to be lower than thresholds measured under A alone conditions because there would be increased support for selecting the correct interval (the one containing the target sentence plus noise) due to increased coherence between visible lip movements and acoustic amplitude envelope in the speech plus noise interval.

The observations of BCMP for speech detection suggest

that visual information may be able to alter auditory perception at a very basic level. Exactly where this interaction occurs in the processing of speech is not clear. A model that combines the integration of unimodal and bimodal information (e.g., Massaro's FLMP, 1987, 1998) with bimodal sensory activation (e.g., Stein and Meredith, 1993) may help better explain the great advantages readily observed in *AV* speech recognition with regard to intelligibility, speed of processing (see, for example, Seitz and Grant, 1999), and enhanced speech detection in noise.

ACKNOWLEDGMENTS

This work was supported by Grant Nos. DC00792 and DC01643 from the NIDCD and the Department of Clinical Investigation, Walter Reed Army Medical Center under Work Unit #2556. A preliminary report of this work was presented at the 16th International Congress on Acoustics and the 135th Meeting of the Acoustical Society of America, Seattle, WA, June 25, 1998: (Grant and Seitz, 1998b). All subjects participating in this research provided written informed consent prior to beginning the study. We would like to thank Steve Tasko for his assistance in analyzing data from the Wisconsin x-ray Microbeam database. We would also like to thank Sid Bacon, Andrew Faulkner, and Dominic Massaro for their helpful comments on an earlier draft of this paper. The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of the Army or Department of Defense.

[1]The overall effect of orthography on speech detection was relatively small (~0.5 dB). It is possible that the small size of the overall effect of orthography limited our ability to see significant differences in masking release for the different sentences. However, it should be pointed out that the sentence that yielded the smallest BCMP (sentence 2) yielded the largest masking protection when presented with orthography, suggesting that the mechanisms involved in AV_M and AV_O conditions may have been different.

[2]It may be possible to improve the amount of masking protection afforded by visual orthography by presenting the orthographic signal simultaneously and with the same overall duration as the target sentence. This would provide a temporal marker for the onset and offset of the target sentence, which according to Gordon (2000) helps bind the signal (acoustic target sentence) and cosignal (orthographic signal) and facilitates masking protection.

American National Standards Institute (**1989**). ''Specifications for audiometers,'' ANSI S3.6-1989, American National Standards Institute, New York.

Braida, L. D. (**1991**). ''Crossmodal integration in the identification of consonant segments,'' Q. J. Exp. Psych. **43**, 647–677.

Bruning, J. L., and Kintz, B. L. (**1968**). *Computational Handbook of Statistics* (Scott, Foresman, Glenview, IL).

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., Woodruff, P. W., Iversen, S. D., and David, A. S. (**1997**). ''Activation of auditory cortex during silent speechreading,'' Science **276**, 593–596.

Egan, J. P., Greenberg, G. Z., and Shulman, A. I. (**1961a**). ''Interval of time uncertainty in auditory detection,'' J. Acoust. Soc. Am. **33**, 771–778.

Egan, J. P., Schulman, A. I., and Greenberg, G. Z. (**1961b**). ''Memory for waveform and time uncertainty in auditory detection,'' J. Acoust. Soc. Am. **33**, 779–781.

Fairbanks, G. (**1950**). ''A physiologic correlative of vowel intensity,'' Speech Monogr. **17**, 390–395.

Frost, R. (**1991**). ''Phoentic recoding of print and its effect on the detection of concurrent speech in amplitude-modulated noise,'' Cognition **39**, 195–214.

Frost, R., Repp B. H., and Katz, L. (**1988**). ''Can speech perception be influenced by simultaneous presentation of print?'' J. Mem. Lang. **27**, 741–755.

Gordon, P. C. (**1997a**). ''Coherence masking protection in brief noise complexes: Effects of temporal patterns,'' J. Acoust. Soc. Am. **102**, 2276–2283.

Gordon, P. C. (**1997b**). ''Coherence masking protection in speech sounds: The role of formant synchrony,'' Percept. Psychophys. **59**, 232–242.

Gordon, P. C. (**2000**). ''Masking protection in the perception of auditory objects,'' Speech Commun. **30**, 197–206.

Grant, K. W., and Braida, L. D. (**1991**). ''Evaluating the Articulation Index for audiovisual input,'' J. Acoust. Soc. Am. **89**, 2952–2960.

Grant, K. W., Braida, L. D., and Renn, R. J. (**1991**). ''Single-band amplitude envelope cues as an aid to speechreading,'' Q. J. Exp. Psych. **43**, 621–645.

Grant, K. W., Braida, L. D., and Renn, R. J. (**1994**). ''Auditory supplements to speechreading: Combining amplitude envelope cues from different spectral regions of speech,'' J. Acoust. Soc. Am. **95**, 1065–1073.

Grant, K. W., and Seitz, P. F. (**1998a**). ''Measures of auditory-visual integration in nonsense syllables and sentences,'' J. Acoust. Soc. Am. **104**, 2438–2450.

Grant, K. W., and Seitz, P. F. (**1998b**). ''The use of visible speech cues (speechreading) for directing auditory attention: Reducing temporal and spectral uncertainty in auditory detection of spoken sentences,'' J. Acoust. Soc. Am. **103**, 3018(A).

Grant, K. W., and Walden, B. E. (**1996**). ''Evaluating the Articulation Index for auditory-visual consonant recognition,'' J. Acoust. Soc. Am. **100**, 2415–2424.

Grant, K. W., and Walden, B. E. (**1995**). ''Predicting auditory-visual speech recognition in hearing-impaired listeners,'' presented at the XIIIth International Congress of Phonetic Sciences, Stockholm, Sweden, August 13–19, 1995, Vol. 3, pp. 122–129.

Grant, K. W., Walden, B. E., and Seitz, P. F. (**1998**). ''Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration,'' J. Acoust. Soc. Am. **103**, 2677–2690.

Hall III, J. W., Haggard, M. P., and Fernandes, M. A. (**1984**). ''Detection in noise by spectrotemporal pattern analysis,'' J. Acoust. Soc. Am. **76**, 50–56.

IEEE (**1969**). ''IEEE recommended practice for speech quality measurements,'' Institute of Electrical and Electronic Engineers, New York.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Luce, R. D. (**1963**). ''Detection and recognition,'' in *Handbook of Mathematical Psychology*, edited by R. D. Luce, R. R. Bush, and E. Galanter (Wiley, New York).

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.

Massaro, D. W. (**1987**). *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry* (Lawrence Earlbaum, Hillsdale, NJ).

Massaro, D.W. (**1998**). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle* (MIT Press, Cambridge, MA).

McGrath, M., and Summerfield, Q. (**1985**). ''Intermodal timing relations and audio-visual speech recognition by normal-hearing adults,'' J. Acoust. Soc. Am. **77**, 678–685.

Meredith, M. A., Nemitz, J. W., and Stein, B. E. (**1987**). ''Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors,'' J. Neurosci. **10**, 3727–3742.

Miller, G. A., Heise, G. A., and Lichten, W. (**1951**). ''The intelligibility of speech as a function of the context of the speech materials,'' J. Exp. Psychol. **41**, 329–335.

Miller, G. A., and Nicely, P. E. (**1955**). ''An analysis of perceptual confusions among some English consonants,'' J. Acoust. Soc. Am. **27**, 338–352.

Repp, B. H., Frost, R., and Zsiga, E. (**1992**). ''Lexical mediation between sight and sound in speechreading,'' Q. J. Exp. Psych. **45**, 1–20.

Richards, V. M. (**1987**). ''Monaural envelope correlation perceptions,'' J. Acoust. Soc. Am. **82**, 1621–1630.

Sams, M., Aulanko, R., Hamalainen, M., Hari, R., Lounasmaa, O. V., Lu, S. T., and Simola, J. (**1991**). ''Seeing speech: Visual information from lip

movements modifies activity in the human auditory cortex,'' Neurosci. Lett. **127**, 141–145.
Seitz, P. F., and Grant, K. W. (**1999**). ''Modality effects on perceptual encoding and memory representations of spoken words,'' Percept. Psychophys. (submitted).
Spence, C., Ranson, J., and Driver, J. (**2000**). ''Cross-modal selective attention: On the difficulty of ignoring sounds at the locus of visual attention,'' Percept. Psychophys. **62**, 410–424.
Stein, B. E., and Meredith, M. A. (**1993**). *The Merging of the Senses* (MIT Press, Cambridge, MA).
Stevens, K. N., and House, A. S. (**1955**). ''Development of a quantitative description of vowel articulation,'' J. Acoust. Soc. Am. **27**, 484–493.
Stevens, K. N., and House, A. S. (**1961**). ''An acoustical theory of vowel production and some of its implications,'' J. Speech Hear. Res. **4**, 303–320.
Sumby, W. H., and Pollack, I. (**1954**). ''Visual contribution to speech intelligibility in noise,'' J. Acoust. Soc. Am. **26**, 212–215.
Summerfield, Q. (**1987**). ''Some preliminaries to a comprehensive account of audio-visual speech perception,'' in *Hearing by Eye: The Psychology of Lip-Reading*, edited by B. Dodd and R. Campbell (Lawrence Erlbaum, Hillsdale, NJ).
Walden, B. E., Prosek, R. A., and Worthington, D.W. (**1975**). ''Auditory and audiovisual feature transmission in hearing-impaired adults,'' J. Speech Hear. Res. **18**, 272–280.
Watson, C. S., and Nichols, T. L. (**1976**). ''Detectability of auditory signals presented without defined observation intervals,'' J. Acoust. Soc. Am. **59**, 655–668.
Westbury, J. R. (**1994**). *X-Ray Microbeam Speech Production Database User's Handbook* (University of Wisconsin, Madison, WI).
Yehia, H., Rubin, P., and Vatikiotis-Bateson, E. (**1998**). ''Quantitative association of vocal-tract and facial behavior,'' Speech Commun. **26**, 23–43.

Effects of frequency and level on auditory stream segregation

Marina M. Rose[a] and Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, United Kingdom

(Received 19 November 1999; revised 21 March 2000; accepted 30 May 2000)

This study examined the effect of center frequency and level on the perceptual grouping of rapid tone sequences. The sequence ABA-ABA-⋯ was used, where A and B represent sinusoidal tone bursts (10-ms rise/fall, 80-ms steady state, 20-ms interval between tones) and - represents a silent interval of 120 ms. In experiment 1, tone A was fixed in frequency at 62, 125, 250, 500, 1000, 2000, 4000, 6000, or 8000 Hz. Both tones had a level of approximately 40 dB SL. Tone B started with a frequency well above that of tone A, and its frequency was swept toward that of tone A so that the frequency separation between them decreased in an exponential manner. Subjects were required to indicate when they could no longer hear the tones A and B as two separate streams, but heard only a single stream with a ''gallop'' rhythm. This changeover point between percepts is called the fission boundary. The separation between tones A and B at the fission boundary was roughly independent of the frequency of tone A when expressed as the difference in number of equivalent rectangular bandwidths (ERBs) between A and B, but varied more with frequency when the difference was expressed in barks or cents. In experiment 2, the center frequency was fixed at 250, 1000, or 4000 Hz, and the level of the A and B tones was 40, 55, 70, or 85 dB SPL. The frequency separation of the A and B tones at the fission boundary tended to increase slightly with increasing level, in a manner consistent with the broadening of the auditory filter with increasing level. The results support the ''peripheral channeling'' explanation of stream segregation advanced by Hartmann and Johnson [Music Percept. **9**, 155–184 (1991)], and indicate that the perception of fusion or fission in alternating-tone sequences depends partly upon the degree of overlap of the excitation patterns evoked by the successive sounds in the cochlea, as assumed in the theory of Beauvois and Meddis [J. Acoust. Soc. Am. **99**, 2270–2280 (1996)]. [S0001-4966(00)01509-5]

PACS numbers: 43.66.Lj, 43.66.Mk [DWG]

INTRODUCTION

When we listen to rapid sequences of sounds, sounds with components falling in similar frequency ranges tend to be grouped together (i.e., perceived as if they come from a single source—called fusion or coherence), whereas sounds with components in different frequency ranges tend to be perceived as different streams (as coming from more than one source—called fission or segregation) (van Noorden, 1975; Bregman, 1990). For intermediate frequency separations between successive tones in a rapid sequence, the percept may be ambiguous: either fission or fusion may be heard, depending on the attentional set of the subject and the instructions given. For example, the subject may be instructed to try to hear a single stream or to try to hear two streams. When the percept is ambiguous, the tendency to hear fission increases with increasing exposure time to the tone sequence (Bregman, 1978). For the tone sequences typically used in experiments on streaming (with alternation rates in the range 2 to 10/s), the percept seems to stabilize after about 10 s (Anstis and Saida, 1985). If the frequency separation is larger than a critical value, called the temporal coherence boundary (van Noorden, 1975), two streams are always heard, at least after a few tones have been presented. If the frequency separation is less than a (different) critical value, called the fission boundary (van Noorden, 1975), then a single stream is heard no matter how long the sequence continues. The fission boundary is affected only slightly by the repetition period of the tones, while the temporal coherence boundary increases markedly with increasing tone-repetition period (van Noorden, 1975).

It has been proposed (Beauvois and Meddis, 1991, 1995, 1996; McCabe and Denham, 1997; Hartmann and Johnson, 1991) that the perceptual organization of such sequences of sounds depends upon the filtering that takes place in the cochlea. For example, the computer model of Beauvois and Meddis (1996) works on the principle that the percept of hearing a single stream as opposed to hearing more than one stream depends partly upon the overlap of the excitation patterns evoked by the sounds in the cochlea; a large degree of overlap leads to fusion while a small degree of overlap leads to fission. Experimental evidence supporting this theory has been found by Hartmann and Johnson (1991), whose results indicated that successive tones are integrated into the same perceptual stream if they excite the same peripheral channels. The overlap of the excitation patterns depends on the sharpness of the auditory filters (Moore and Glasberg, 1983); for a given frequency separation of two successive sounds,

[a]Present address: University Laboratory of Physiology, Parks Road, Oxford OX1 3PT, United Kingdom.

the sharper the auditory filters, the smaller will be the overlap of the excitation patterns. Beauvois and Meddis (1996) assumed that the bandwidths of the auditory filters, and the sharpness of the excitation patterns, could be characterized using the equations given by Glasberg and Moore (1990). These equations describe how the equivalent rectangular bandwidth (ERB) of the auditory filter varies with center frequency, and also define how the frequency scale can be transformed into a scale based on ERB number (analogous to the bark scale; see Zwicker, 1961; Zwicker and Terhardt, 1980). The ERB number, E, corresponding to a given center frequency, F (kHz), is

$$E = 21.4 \log_{10}(4.37F + 1). \quad (1)$$

The theory of Beauvois and Meddis (1996) predicts that the frequency difference between successive tones at the fission boundary should be constant when expressed as the difference in E value between the two tones, ΔE, regardless of the center frequency of the tones. It also predicts that the ΔE value at the fission boundary should increase slightly with increasing level, as the bandwidth of the auditory filter increases with increasing level (Glasberg and Moore, 1990). Beauvois and Meddis (1996) assessed the first prediction by using their computer model to simulate the data of Miller and Heise (1950), who examined the perception of a tone sequence alternating between two frequencies with a rate of ten tones per second. Miller and Heise determined the frequency separation at which ''the trill seems to break.'' They named this interval the ''trill threshold.'' It is not clear from their description whether they measured the temporal coherence boundary or the fission boundary; indeed different subjects may have adopted different criteria. The computer simulation of Beauvois and Meddis (1996) indicated that the trill threshold corresponds to the fission boundary, and that the fission boundary is roughly equal in size to the ERB at the frequency of the lower tone of the trill. However, considerable individual variability was apparent in the data of Miller and Heise (1950) and no information was given about within-subject reproducibility.

In a previous paper (Rose and Moore, 1997) we presented data on fission boundaries as a function of center frequency using a method with well-defined instructions, that had been determined to give stable and reproducible results. The data were reasonably consistent with the prediction of Beauvois and Meddis (1996) that the frequency separation between tones at the fission boundary would be roughly constant across center frequency when expressed as ΔE values. Fission boundaries were more nearly constant when expressed in this way than when expressed in Barks. However, the fission boundaries were also nearly constant when expressed in cents (hundredths of a semitone). For midrange frequencies, the ERB, bark, and cent scales are similar in form. Hence, the range of center frequencies used in our earlier experiment (250 to 2000 Hz) was not large enough to allow us to determine clearly whether the fission boundary was more nearly constant when expressed on a scale related to auditory frequency selectivity (ERBs or barks), or when expressed on a frequency-ratio scale (cents). Experiment 1 of the present paper measured fission boundaries over a very wide range of center frequencies (62 to 8000 Hz), to give a clearer answer to this question; differences between the ERB, bark, and cent scales are more marked at very low and very high center frequencies. Our second experiment tested the second prediction of the model of Beauvois and Meddis (1996), namely that the frequency separation at the fission boundary would increase somewhat with increasing level.

I. EXPERIMENT 1: MEASUREMENT OF THE FISSION BOUNDARY OVER A WIDE FREQUENCY RANGE

A. Procedure and stimuli

The method used here is similar to that in our earlier study (Rose and Moore, 1997), and is based on a sequence of tones obtained from the sequence ABAB... by omission of every other tone B, as in ABA-ABA-$\cdots$ (van Noorden, 1975). When the frequency ratio of A and B is very small, the subject perceives a single stream with a characteristic galloplike rhythm (fusion). When the frequency ratio is very large, two streams are heard (fission), the stream composed of the A tones having twice the tempo of the stream containing the B tones. At the start of a run, the frequency ratio of tones A and B was always large enough that fission was heard, at least after a few tones had been presented. Subjects were instructed to try to hear fission for as long as possible when the frequency ratio between successive tones was progressively decreased from a large initial value, and to indicate when they could no longer hear two streams. Consistent with the findings of van Noorden (1975), we found that this procedure gives stable and reproducible results (Rose and Moore, 1997).

The frequency of tone A was kept constant (F_{fix}) and the frequency of tone B (F_{var}) was variable. At the beginning of each trial, tone B started at a frequency well above that of tone A, and swept toward the frequency of tone A in an exponential manner, described by

$$F'_{\text{var}} = F_{\text{fix}}(F_{\text{var}}/F_{\text{fix}})^{(1/s)}, \quad (2)$$

where F'_{var}=frequency of tone B (Hz) in the $(i+1)$th ABA-group, F_{fix}=fixed frequency of tone A (Hz), F_{var}=frequency of tone B (Hz) in the ith ABA-group, $s=1.1$ (to give a reasonable rate of change for the exponential function).

With this formula, the size of change for the variable frequency depended on the ratio between the variable and the fixed frequency. At the beginning of each sweep, when the frequency ratio between tones A and B was large, the ratio between tone B in the ith ABA-group and tone B in the $(i+1)$th ABA-group was also large. When the frequency ratio between A and B became smaller, the ratio between tone B in the ith and tone B in the $(i+1)$th ABA-group also became smaller. This was done to determine the fission boundary more accurately. After several repeats of this sequence, the interval between A and B became so small that the subject could not hear two streams any more. At this point, the subject was asked to press the response button (see instructions), indicating that the fission boundary was reached. Depending on the exact starting value for the frequency of tone B and the subject's perception, this took between 20 and 40 ABA groups (10 to 20 s).

TABLE I. The lower and upper limits of the range of starting frequencies, F_{init}, for tone B for each frequency of tone A, F_{fix}.

	F_{fix} (Hz)								
	62	125	250	500	1000	2000	4000	6000	8000
Lower	187	375	750	1500	2000	4000	8000	12 000	16 000
Higher	250	500	1000	2000	3000	4400	10 000	14 000	18 000

Each tone had an 80-ms steady-state duration and 10-ms raised-cosine ramps. Between tones there was a 20-ms silence; hence, the tone-repetition time was 120 ms. After each ABA group there was a 120-ms pause (silent interval, indicated as -). F_{fix} was 62, 125, 250, 500, 1000, 2000, 4000, 6000, or 8000 Hz. The initial frequency of tone B, F_{init}, was chosen randomly from a previously defined starting range, which depended on F_{fix}. The range of F_{init} for each value of F_{fix} is given in Table I. The stimuli were presented monaurally, and the two ears were tested alternately.

All subjects were instructed to listen selectively. A typical oral instruction was as follows: ''You will hear two alternating tones in one ear. One tone is fixed in frequency, and the other one is varying. Please try to follow the varying one as long as you hear it as a separate pattern. When you cannot follow it separately any more, you will hear a distinct gallop-like rhythm. At this point, press the stop button. As soon as you release the stop button, a very similar sweep will start again. After five such sweeps, there will be a short rest, and the same procedure will be repeated with your other ear. Do you have any questions?''

One run consisted of five sweeps which all had the same fixed frequency for tone A but a randomized starting frequency for tone B. Only the last four out of the five sweeps contributed to the mean for one run. Usually the standard deviation of the mean frequency of tone B at the fission boundary was below 3% of the arithmetic mean value for a given run. Rarely, when the standard deviation exceeded 5% of the arithmetic mean value, the run was discarded and a new run obtained. After finishing one run, the same procedure was repeated with the other ear. At least four runs were obtained for each ear at each frequency, and the overall mean was calculated from the means of the last three. Testing for each condition was spread over at least two experimental sessions. Standard deviations of the mean frequency of tone B at the fission boundary were usually small, with values that were usually less than 3% and always less than 5% of the mean.

Stimuli were digitally generated using a Tucker-Davis System II, with a TDT DD1 16-bit digital-to-analog converter (50-kHz sampling rate). The stimulus level was controlled using a TDT PA4 programmable attenuator, and stimuli were then passed to a headphone buffer (TDT HB6), a final manual attenuator (Hatfield 2125), and one earpiece of a Sennheiser HD414 headphone. Sound levels were specified as levels close to the eardrum. The calibration of the Sennheiser HD414 earphone was based on average measurements for three subjects (not those of the present study), using a Rastronics Portarem 2000 system, covering the frequency range from 125 to 8000 Hz. The probe microphone was positioned within 4 mm of the eardrum as recommended by the manufacturer, using the on-screen display to monitor standing waves in the ear canal and to place the valley in the frequency response curve above 8 kHz (Siegel, 1994). For a fixed input voltage, the standard deviation of the sound level across subjects was typically about 1 dB for frequencies up to 1 kHz, about 2 dB for frequencies from 2 to 6 kHz, and about 3 dB from 6 to 8 kHz. The subjects were tested in an IAC double-walled sound-attenuating chamber.

B. Subjects and choice of test level

Seven subjects were used, aged 24 to 40 years. Absolute thresholds were estimated in dB SPL using a two-alternative forced-choice method, as described in Rose and Moore (1997). These values were converted to absolute thresholds in dB HL (hearing level, i.e., dB relative to the average threshold for young normally hearing subjects) using the values for the normal absolute threshold for monaural listening specified in Moore *et al.* (1997). All subjects had thresholds better than 20 dB HL for all audiometric frequencies from 125 to 8000 Hz, except for the left ear of subject MMR at 6 kHz, in which case the threshold was 35 dB HL.

In the great majority of cases, the fixed tone A was set at a level 40 dB above the absolute threshold, i.e., at 40 dB SL. A fixed SL rather than a fixed SPL was used as the former gave a comfortable loudness over a wide frequency range. The voltage applied to the earphone to generate tone B was the same as the voltage used to generate tone A, i.e., we did not compensate for small differences in loudness that might have resulted from the difference in frequency of the A and B tones. However, in the region of the fission boundary, the A and B tones were always judged to be nearly equal in loudness. In two cases, the level of tone A was not set to 40 dB SL. Subject JIA, who had the highest absolute threshold of 63 dB SPL in his right ear at 62 Hz (nominal input level, as no calibration for the earphone was available at this frequency), was not tested at 103 dB SPL, as would have been appropriate to give 40 dB SL, but was tested at a lower level of 100 dB SPL. This level was chosen to ensure a comfortable loudness. The same criterion applied to subject UUS, who complained that the sound was uncomfortably loud at a level of 93 dB SPL (corresponding to 40 dB SL). He was tested at a level of 90 dB SPL at 62 Hz in both ears.

C. Results

The absolute values of the fission boundaries varied across subjects, perhaps reflecting differences in criterion. This was also found in our earlier work (Rose and Moore, 1997). For a given center frequency, the standard deviation (sd) across subjects, in ΔE units, was typically around 0.15. However, the pattern of the results was similar across subjects, so we will concentrate on the trends apparent in the mean data.

To determine whether the fission boundary is related to the familiar musical scale of semitones or to an auditory scale which reflects the frequency selectivity of the auditory system, the differences in the frequencies of tones A and B at the fission boundary were expressed in three different ways.

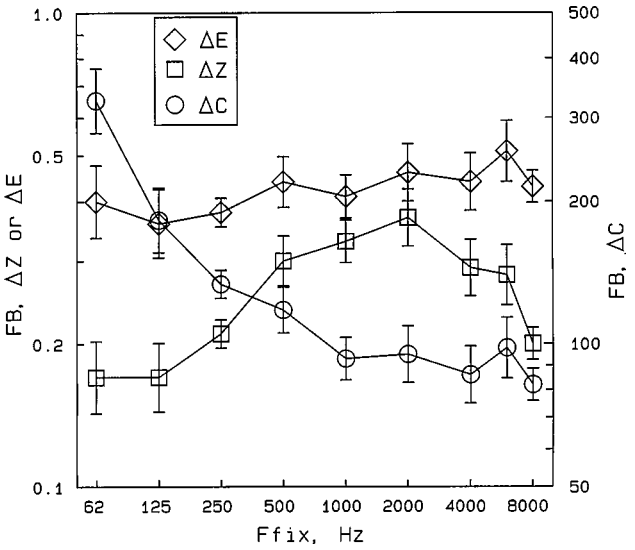


FIG. 1. Mean values across ears of the fission boundary (FB) for seven normally hearing subjects, expressed as ΔE (ERBs), ΔZ (barks) (both on the left ordinate), and ΔC (cents, on the right ordinate), plotted as a function of the frequency of tone A, F_{fix}. Error bars indicate ± one standard error.

(1) As the corresponding difference in cents, ΔC. The frequency ratio corresponding to one cent is $2^{(1/1200)} = 1.000\,577\,79$.
(2) As the difference in ERB number, ΔE.
(3) As the difference in barks, ΔZ. This is related to the classical critical bandwidth, and has also been called the ''critical band rate'' (Zwicker, 1961; Zwicker and Terhardt, 1980). The formula proposed by Traunmüller (1990) was used to convert frequency to number of barks,

$$Z=[26.81/(1+1960/f)]-0.53, \tag{3}$$

where f is frequency in Hz.

The mean fission boundaries for the seven subjects are shown in Fig. 1. The error bars indicate ± one standard error. Their size is largely determined by variability in the overall values of the fission boundary across subjects; some subjects consistently had larger fission boundaries than others, for all frequencies. The reproducibility of the means is markedly smaller than indicated by the error bars.

The value of ΔE did not vary markedly with center frequency. The value of ΔC increased at low center frequencies, while ΔZ showed a clear maximum around 2000 Hz. The extent to which the measures varied as a function of frequency was assessed by calculating the ratio of the standard deviation across frequency to the mean across frequency. This ratio was 0.10 for ΔE, 0.28 for ΔZ, and 0.58 for ΔC. To determine whether the relative variability of ΔE was significantly smaller than that of ΔZ or ΔC, the normalized standard deviations were converted to variances by squaring them, and the variance ratios were calculated for $\Delta Z/\Delta E$ and $\Delta C/\Delta E$. The resulting F-ratios were 8 and 34, respectively. With seven degrees of freedom for each variance, the F-ratio needs to be larger than 6.99 to achieve statistical significance at $p<0.01$. Thus the variance in ΔZ is significantly greater than that in ΔE, and the variance in ΔC is significantly greater than that in ΔE (even allowing for the fact that two comparisons are being made). The separation between the two tones at the fission boundary for a wide range of frequencies is therefore most nearly constant when it is expressed as ΔE. This is consistent with the theory of Beauvois and Meddis (1996), and is also consistent with the ERB scale being a more accurate representation of human frequency selectivity than the bark scale.

It should be noted that results might have differed if we had kept the SPL constant across frequency rather than the SL. If a fixed SPL had been used, it would have been necessary to use a rather high level, to ensure that the 62-Hz tone was clearly audible. The effects of level were explored in experiment 2.

II. EXPERIMENT 2: THE EFFECT OF OVERALL LEVEL ON THE FISSION BOUNDARY

The auditory filter bandwidth increases at high levels, at least for medium and high frequencies (Weber, 1977; Glasberg and Moore, 1990). If the fission boundary is related to peripheral filtering, one would expect to find larger ΔE values at high levels. The second experiment tested this prediction.

A. Method

Six normally hearing subjects (aged 26–48 years) were tested in the same way as described above. Three of the subjects had taken part in experiment 1, and three were new. All subjects had absolute thresholds better than 20 dB HL at the test frequencies. The fission boundary was measured for three center frequencies ($F_{\text{fix}}=250$, 1000, and 4000 Hz), and four levels at each frequency (40, 55, 70, and 85 dB SPL). Note that, in this experiment, the levels are specified in dB SPL, not dB SL.

B. Results

Figure 2 shows the mean fission boundaries across the six subjects, plotted as a function of sound level with center frequency as parameter. The fission boundaries are expressed as ΔE values. The error bars indicate ± one standard error. As before, their size is largely determined by variability in the overall values of the fission boundary across subjects; some subjects consistently had larger fission boundaries than others, for all frequencies and levels. However, the pattern of results was similar across subjects. The ΔE values increase slightly with increasing center frequency, and they also tend to increase with level, although the effect is small at 250 Hz, and the effect at 4000 Hz is not monotonic. The nonmonotonicity was present for both ears of three subjects and for one ear only for the other three subjects. A within-subjects analysis of variance of the ΔE values, with factors frequency and level, gave significant main effects of frequency, $F(2,10)=8.24$, $p=0.008$, and of level, $F(3,15)=11.46$, $p<0.001$. The interaction of level and frequency approached, but did not reach, significance at the 0.05 level; $F(6,30)=1.94$, $p=0.106$.

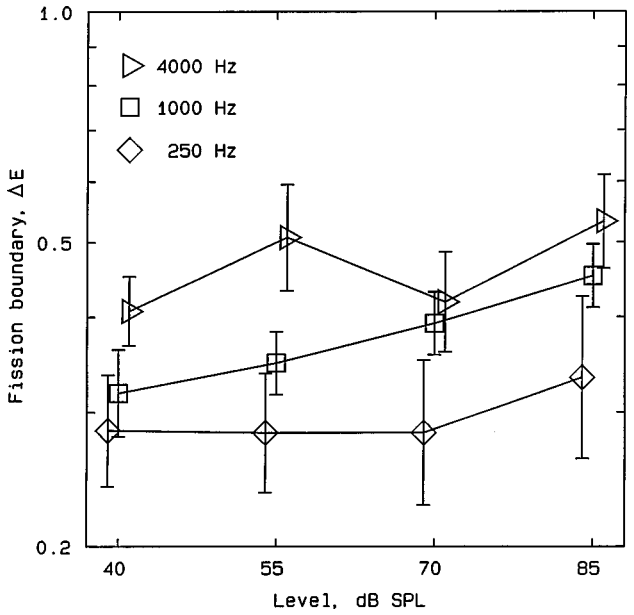


FIG. 2. Mean values of the fission boundary across six subjects plotted as a function of sound-pressure level, with center frequency (250, 1000, or 4000 Hz) as parameter. The fission boundaries are expressed as ΔE values. Error bars indicate $\pm$ one standard error.

The overall effect of center frequency is slightly larger than found in experiment 1; the ΔE values at the fission boundary for center frequencies of 250, 1000, and 4000 Hz, were, respectively, 0.38, 0.41, and 0.44 in experiment 1, and 0.30, 0.41, and 0.46 in experiment 2. This may reflect individual differences in the variation of the ERB of the auditory filter with center frequency; only three subjects were common to both experiments. In addition, in experiment 2, the sensation level of the stimuli, for a fixed sound-pressure level, was about 6 dB less at 250 Hz than at the two higher center frequencies, so the smaller fission boundaries at 250 Hz might partly reflect the lower sensation level at that frequency. Consistent with this, for the three subjects who took part in both experiments, the fission boundary in experiment 2 at 40 dB SPL was consistently lower than that measured in experiment 1 at 40 dB SL.

The increase in level from 40 to 85 dB SPL resulted in increases in the mean fission boundary by factors of 1.17, 1.42, and 1.31, for center frequencies of 250, 1000, and 4000 Hz, respectively. The factors for the two higher frequencies are similar to the factor by which the ERB of the auditory filter broadens with increasing level over the same range (Glasberg and Moore, 1990). The smaller effect at 250 Hz is consistent with evidence that the sharpness of the auditory filter changes only slightly with level for low center frequencies (Rosen and Stock, 1992; Moore and Glasberg, 1997; Moore *et al.*, 1997; Hicks and Bacon, 1999). The smaller change at low center frequencies may reflect a reduced contribution of the cochlear active mechanism (Cooper and Rhode, 1995; Yates, 1995; Rhode and Cooper, 1996). Overall, the variation in the fission boundaries with level is consistent with the way that the auditory filter bandwidth varies with level, lending support to the hypothesis that the fission boundaries are related to the sharpness of the auditory filters.

III. DISCUSSION AND CONCLUSIONS

Our first experiment measured the frequency separation between tones A and B at the fission boundary over a wide range of frequencies for tone A. The results showed that, when expressed as the difference in ERB number between the frequencies of the A and B tones, ΔE, the fission boundary varied only slightly with the frequency of tone A. The variation with frequency was greater when the fission boundary was expressed as the difference in barks, ΔZ, or the frequency ratio in cents, ΔC. These results are consistent with the assumption of the theory of Beauvois and Meddis (1996) and the experimental results of Hartmann and Johnson (1991) indicating that the perception of fusion or fission in alternating-tone sequences depends partly upon the overlap of the excitation patterns evoked by the successive sounds in the cochlea. The results also suggest that the ERB scale provides a better representation of frequency resolution than the bark scale.

Our second experiment showed that the value of ΔE at the fission boundary tended to increase with increasing level, in a way similar to the broadening of the auditory filter with increasing level. Again, this is consistent with the theory of Beauvois and Meddis (1996) and the experimental results of Hartmann and Johnson (1991).

It seems clear, however, that factors other than frequency selectivity can affect auditory stream segregation. For example, perceptual fission between successive complex tones filtered into the same frequency region can be produced by differences in periodicity (Vliegen and Oxenham, 1999; Vliegen *et al.*, 1999). Also, in our earlier paper (Rose and Moore, 1997), we showed that subjects with cochlear hearing loss sometimes had near-normal fission boundaries, in spite of their auditory filters being broader than normal, although other hearing-impaired subjects did show larger-than-normal fission boundaries. We suggested that the fission boundary may depend on the frequency-to-place mapping within the cochlea; the fission boundary may correspond to the point where the peaks in the excitation patterns evoked by successive tones are separated by a constant distance along the basilar membrane. This distance might be more important than the degree of overlap of the excitation patterns evoked by the successive tones. This could explain why the ΔE magnitudes for normal-hearing subjects are roughly constant across frequency; the ERB scale corresponds reasonably well to a scale of distance along the basilar membrane in normally hearing subjects (Moore, 1986). The mean value of ΔE at the fission boundary of about 0.42 corresponds to a distance of about 0.37 mm along the basilar membrane. However, if the distance between peaks on the basilar membrane were the only factor involved, we would not expect the fission boundary to change with overall level [there is some evidence that the position of the peak on the basilar membrane changes with level (Ruggero *et al.*, 1997), but the amount of this shift would be similar for tones whose frequencies differed only slightly, as was the case for tones A and B at the fission boundary]. In fact, as shown in experiment 2, fission boundaries increase slightly but significantly with increasing level.

In conclusion, our results show that the fission boundary

is roughly constant over a wide range of frequencies when expressed as the difference in ERB number, ΔE. The fission boundary increases somewhat with increasing level. The results are broadly consistent with the idea that auditory stream segregation depends partly on the overlap of the excitation patterns of the successive tones on the basilar membrane and/or on the distance between the peaks of the excitation patterns on the basilar membrane. However, these are probably not the only factors involved.

ACKNOWLEDGMENTS

This work was supported by a studentship from Defeating Deafness (The Hearing Research Trust) awarded to author M.M.R. and by the Medical Research Council (UK). We thank Thomas Baer and Brian Glasberg for assistance with programming and José Alcántara for calibration of the earphones. José Alcántara, Thomas Baer, and Brian Glasberg provided helpful comments on an earlier version of this paper. We also thank the reviewers, Mike Beauvois and Nick Hill, for helpful comments.

Anstis, S., and Saida, S. (**1985**). ''Adaptation to auditory streaming of frequency-modulated tones,'' J. Exp. Psychol. **11**, 257–271.

Beauvois, M. W., and Meddis, R. (**1991**). ''A computer model of auditory stream segregation,'' Q. J. Exp. Psychol. **43A**, 517–541.

Beauvois, M. W., and Meddis, R. (**1995**). ''Computer simulation of Gestalt auditory grouping by frequency proximity,'' in *Neural Computation and Psychology*, edited by L. S. Smith and P. J. B. Hancock (Springer-Verlag, Berlin).

Beauvois, M. W., and Meddis, R. (**1996**). ''Computer simulation of auditory stream segregation in alternating-tone sequences,'' J. Acoust. Soc. Am. **99**, 2270–2280.

Bregman, A. S. (**1978**). ''Auditory streaming is cumulative,'' J. Exp. Psychol. **4**, 380–387.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA).

Cooper, N. P., and Rhode, W. S. (**1995**). ''Nonlinear mechanics at the apex of the guinea-pig cochlea,'' Hear. Res. **82**, 225–243.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.

Hartmann, W. M., and Johnson, D. (**1991**). ''Stream segregation and peripheral channeling,'' Music Percept. **9**, 155–184.

Hicks, M. L., and Bacon, S. P. (**1999**). ''Psychophysical measures of auditory nonlinearities as a function of frequency in individuals with normal hearing,'' J. Acoust. Soc. Am. **105**, 326–338.

McCabe, S. L., and Denham, M. J. (**1997**). ''A model of auditory streaming,'' J. Acoust. Soc. Am. **101**, 1611–1621.

Miller, G. A., and Heise, G. A. (**1950**). ''The trill threshold,'' J. Acoust. Soc. Am. **22**, 637–638.

Moore, B. C. J. (**1986**). ''Parallels between frequency selectivity measured psychophysically and in cochlear mechanics,'' Scand. Audiol. Suppl. **25**, 139–152.

Moore, B. C. J., and Glasberg, B. R. (**1983**). ''Suggested formulae for calculating auditory-filter bandwidths and excitation patterns,'' J. Acoust. Soc. Am. **74**, 750–753.

Moore, B. C. J., and Glasberg, B. R. (**1997**). ''A model of loudness perception applied to cochlear hearing loss,'' Aud. Neurosci. **3**, 289–311.

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.

Rhode, W. S., and Cooper, N. P. (**1996**). ''Nonlinear mechanics in the apical turn of the chinchilla cochlea *in vivo*,'' Aud. Neurosci. **3**, 101–121.

Rose, M. M., and Moore, B. C. J. (**1997**). ''Perceptual grouping of tone sequences by normally hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **102**, 1768–1778.

Rosen, S., and Stock, D. (**1992**). ''Auditory filter bandwidths as a function of level at low frequencies (125 Hz-1 kHz),'' J. Acoust. Soc. Am. **92**, 773–781.

Ruggero, M. A., 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.

Siegel, J. H. (**1994**). ''Ear-canal standing waves and high-frequency sound calibration using otoacoustic emission probes,'' J. Acoust. Soc. Am. **95**, 2589–2597.

Traunmüller, H. (**1990**). ''Analytical expressions for the tonotopic sensory scale,'' J. Acoust. Soc. Am. **88**, 97–100.

van Noorden, L. P. A. S. (**1975**). ''Temporal coherence in the perception of tone sequences,'' Ph.D. Thesis, Eindhoven University of Technology.

Vliegen, J., Moore, B. C. J., and Oxenham, A. J. (**1999**). ''The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task,'' J. Acoust. Soc. Am. **106**, 938–945.

Vliegen, J., and Oxenham, A. J. (**1999**). ''Sequential stream segregation in the absence of spectral cues,'' J. Acoust. Soc. Am. **105**, 339–346.

Weber, D. L. (**1977**). ''Growth of masking and the auditory filter,'' J. Acoust. Soc. Am. **62**, 424–429.

Yates, G. K. (**1995**). ''Cochlear structure and function,'' in *Hearing*, edited by B. C. J. Moore (Academic, San Diego).

Zwicker, E. (**1961**). ''Subdivision of the audible frequency range into critical bands (Frequenzgruppen),'' J. Acoust. Soc. Am. **33**, 248.

Zwicker, E., and Terhardt, E. (**1980**). ''Analytical expressions for critical band rate and critical bandwidth as a function of frequency,'' J. Acoust. Soc. Am. **68**, 1523–1525.

Neural system identification model of human sound localization

Craig Jin[a)]
Auditory Neuroscience Laboratory, Department of Physiology, and Computer and Engineering Laboratory, Department of Electrical and Information Engineering, The University of Sydney, New South Wales 2006, Australia

Markus Schenkel[b)]
Computer Engineering Laboratory, Department of Electrical and Information Engineering, The University of Sydney, New South Wales 2006, Australia

Simon Carlile
Auditory Neuroscience Laboratory, Department of Physiology, and Institute for Biomedical Research, The University of Sydney, New South Wales 2006, Australia

(Received 19 May 1999; revised 23 February 2000; accepted 9 June 2000)

This paper examines the role of biological constraints in the human auditory localization process. A psychophysical and neural system modeling approach was undertaken in which performance comparisons between competing models and a human subject explore the relevant biologically plausible ''realism constraints.'' The directional acoustical cues, upon which sound localization is based, were derived from the human subject's head-related transfer functions (HRTFs). Sound stimuli were generated by convolving bandpass noise with the HRTFs and were presented to both the subject and the model. The input stimuli to the model were processed using the Auditory Image Model of cochlear processing. The cochlear data were then analyzed by a time-delay neural network which integrated temporal and spectral information to determine the spatial location of the sound source. The combined cochlear model and neural network provided a system model of the sound localization process. Aspects of humanlike localization performance were qualitatively achieved for broadband and bandpass stimuli when the model architecture incorporated frequency division (i.e., the progressive integration of information across the different frequency channels) and was trained using variable bandwidth and center-frequency sounds. Results indicate that both issues are relevant to human sound localization performance. © *2000 Acoustical Society of America.* [S0001-4966(00)02709-0]

PACS numbers: 43.66.Qp, 43.66.Ba, 43.66.Pn [DWG]

I. INTRODUCTION

A. Human sound localization

The ability to accurately estimate the position of a sound source based upon the available acoustical cues has obvious evolutionary advantages in terms of avoiding predators and finding prey. Indeed, humans are amazingly accurate in their ability to localize broadband sounds. Cues to a sound's location include the interaural differences in time of arrival (ITDs), the interaural differences in sound level (ILDs), and the monaural spectral cues (recent reviews: Middlebrookes and Green, 1991; Wightman and Kistler, 1993; King and Carlile, 1995). There has been a considerable amount of psychoacoustical research into the auditory processes involved in human sound localization (for a recent review with emphasis on virtual space acoustics see Carlile, 1996a). Furthermore, numerous models of the human sound localization process, as well as that for other animals, have been proposed [Searle *et al.*, 1976; Middlebrooks, 1992; Neti *et al.*, 1992a (for the cat); Backman and Karjalainen, 1993; Nandy *et al.*, 1993; Rosen *et al.*, 1994 (for the barn owl); Chau and Duda, 1995; Datum *et al.*, 1996; Duda, 1997; Janko *et al.*, 1997; Macpherson, 1997; Hofman and Opstal, 1998]. However, there still remains a large gap between the psychophysical and the model explanations (but see, Macpherson, 1997; Janko *et al.*, 1997; Hofman and Van Opstal, 1998). Principal congruence between the two approaches exists for localization performance under restricted conditions, such as for narrow-band sounds where spectral integration is not required, or for restricted regions of space (see Blauert, 1997). Unfortunately, there is no existing computational model that accounts well for human sound localization performance for a wide range of sounds (e.g., varying in bandwidth and center frequency). Furthermore, the biological constraints pertinent to the sound localization process have generally not been explored by these models. Such constraints include the spectral resolution of the auditory system in terms of the number of critical band filters or frequency channels and the role of tonotopic processing and frequency division (progressive integration of information across frequency channels). In addition, the performance requirements of such a system are substantial and involve, for example, the accommodation of spectrally complex sounds, the robustness to irregularity in the sound source spectrum (Macpherson, 1996), and the channel-based structure of spatial coding as evidenced by auditory spatial aftereffects (Hyams and Carlile, 1996). The

[a)]Electronic mail: craigj@physiol.usyd.edu.au
[b)]Current address: Supercomputing Systems, Technoparkstrasse 1, Zürich, Switzerland, 8005.

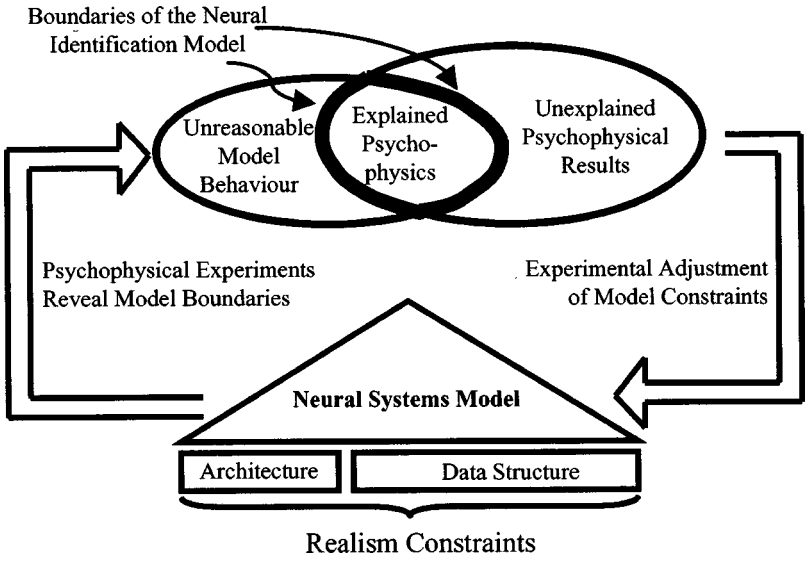


FIG. 1. The neural system modeling paradigm.

crux of the matter is the notion that ''meaningful realism,'' if built into a model, provides for a better understanding of the underlying processes.

This paper attempts to bridge part of this gap between the modeling and psychophysics. It describes the development of a time-delay neural network model that integrates both spectral and temporal cues for auditory sound localization and compares the results with the corresponding human psychophysical evidence. Realistic constraints have been incorporated into the modeling process by imposing frequency division on the network structure and using training sounds with random center frequency and bandwidth. It is the primary purpose of this paper to examine the relevance of these two issues to human sound localization performance. Additionally, the relationship of these models to a matched filtering algorithm is examined, as is the degradation of localization performance with varying sound level.

With respect to the first issue of frequency division, previous studies have generally neglected its role in sound localization and also its relationship with the integration of spectral information. Nonetheless, the auditory system generally integrates information across frequency to derive the location of a sound source (described below) and it still remains an open question as to just how this integration is accomplished. What is meant by frequency division in this work is that the first computational layer of the network model is *not* fully connected to the input cochlear frequency channels. Instead, the network structure is constrained such that a *progressive* integration of information across the different frequency channels is enforced by the network structure as information flows through the different computational layers from the input to the output. Other localization models have certainly used a tonotopic arrangement for their input, but not for the connections of their network model. Typically (e.g., Neti *et al.*, 1992) all of the neurons in the first computational layer are fully connected to *all* of the input cochlear frequency channels.

The second issue takes a look at the challenging problem of training a network model to localize sounds with varying bandwidth and center frequency. The issue touches upon the question of how frequency content influences the integration of spectral information for sound localization. Furthermore, results suggest that networks with frequency division may have a performance advantage in coping with varying sound conditions.

The model described here is primarily a ''black-box'' or system identification model. Some of the general advantages of the system identification approach using neural networks are well described by David Zipser (1992), who first introduced the concept of ''realism constraints'' into the process of neural system identification. As he states (p. 861), ''The neural system identification paradigm provides a systematic way to generate realistic models starting with a high-level description of a hypothesized computation and some architectural and physiological constraints.'' System identification is concerned with replicating the input-output behavior of the system being modeled and neural system identification attempts to model the nervous system in a similar fashion using supervised artificial neural networks. Figure 1 shows a simple diagram illustrating the neural system identification paradigm as it has been applied to the model discussed in this paper. A number of different network architectures and structures for the training data were examined and the results were compared with empirical localization data before making further adjustments to the model. Furthermore, the ''black-box'' approach was also used to indirectly address two complicated issues related to the sound localization problem: (i) identifying spectral features related to source location and (ii) developing a formal encoding of spatial location (described below).

Modeling the head as a simple sphere, it has been shown that two prominent cues to a sound's location are the interaural time and level cues. The inherent spatial ambiguity derived from such interaural time and level cues has become known as the ''cone of confusion'' (Wallach, 1939; Blauert, 1997). A further set of cues is provided by the pinna and concha which preferentially amplify or attenuate different frequency components of the sound spectrum depending on the spatial location of the sound source. Psychoacoustical work has demonstrated the importance of these spectral cues

in resolving the ambiguity of the interaural time and level cues (Batteau, 1967; Musicant and Butler, 1984; Oldfield and Parker, 1984), in localization within the vertical dimension (Roffer and Butler, 1968; Gardner and Gardner, 1973; Searle *et al.*, 1975) and in generating the percept of externalized sounds (Durlach *et al.*, 1992). It is generally accepted that the relevant acoustical cues to a sound's location are described by the head-related transfer function (HRTF) which is typically represented by a finite-length impulse response (FIR) filter (Wightman and Kistler, 1989a; Møller, 1992; Pralong and Carlile, 1994) measured in an anechoic space (recent review: Pralong and Carlile, 1996). Sounds electronically filtered with the HRTF should be localizable when played over earphones which bypass the acoustical filtering of the outer ear. This has been demonstrated (Wightman and Kistler, 1989b; Pralong and Carlile, 1996) and the illusion of free-field sounds using headphones is known as virtual auditory space.

B. Outer ear filtering and spatial location

HRTFs allow for model explorations of localization processing that combine spectral cues along with the interaural time and level cues (Middlebrooks, 1992; Macpherson, 1994). As HRTFs are highly individualized (Wenzel *et al.*, 1988), the modeling results can be compared with human localization performance on an individual basis so as to extract general principles of localization processing via the specifics of individual behavior. Thus a major criterion differentiating more recent sound localization models is their ability to incorporate the individualized spectral filtering of the outer ear. Lyon (1983) and Lindemann (1986) provide early models of the localization process that neglect outer ear filtering but do incorporate interaural time and level cues. Their models derive from the Jeffress (1948) type of interaural cross-correlation model (recent review: Kuwada *et al.*, 1997). Both the Lyon and Lindemann models emphasize the connection between sound localization and sound separation and rely upon temporal correlation applied to the signals at the two ears. Gaik (1993) amended Lindemann's model to incorporate differences in level across frequency channels by adding weighting elements derived from HRTFs. Unfortunately, Gaik's model remains constrained to lateralization; no estimate of source elevation is made. Experimentation with the small desktop Koala robots (van Schaik, 1998; Shamma *et al.*, 1998) has highlighted the difficulties associated with navigating toward a sound source in the absence of an estimate of source elevation. Basically it was found that as the source was moved off the horizontal plane, the robots underestimated the lateral position of the source and moved in several iterative and halting steps.

As HRTF recordings became more common, models of the sound localization process began to estimate source elevation by incorporating information relating to the spectral filtering of the outer ear. These models can mainly be classified into two categories. One class of algorithms uses a mathematically well-defined distance measure to directly match the input sound with a database of HRTFs that are associated with specific directions in space (Middlebrooks, 1992; Nandy *et al.*, 1993; Chau and Duda, 1995; Duda, 1997; Macpherson, 1997; Hofman and Opstal, 1998). Another class of algorithms employs an optimization process (often using neural networks) to determine the matching between an input sound and a given location in space (Neti *et al.*, 1992; Backman and Karjalainen, 1993; Rosen *et al.*, 1994; Datum *et al.*, 1996; Janko *et al.*, 1997). These algorithms are similar to each other and provide for an encoding of space based on outer ear filtering.

The difficulties associated with incorporating spectral filtering into the localization process occur at the second or algorithmic level of David Marr's (1982) hierarchy of explanation for perceptual mechanisms. These difficulties arise because the feature detection paradigm, despite its striking and overwhelming historical success in accounting for sensory phenomena, is difficult to apply when the computational quantities and algorithm are not easily separated. In other words, identifying the exact spectral features used in sound localization is inherently difficult. Generally it is assumed that the most salient spectral cues would be used by the auditory system. Zakarauskas and Cynader (1993) have directly addressed this issue. They propose two possible spectral features that are mathematically specified by first and second order zero-mean difference operators as applied to the spectral signal (see also Carlile, 1990). More recently, Macpherson (1997) has investigated the use of the gradient of the spectral signal. Zakarauskas and Cynader (1993) suggested zero-mean difference operations because they remain invariant to level manipulations. Other methods, which also maintain level invariance, either identify local features in the signal or apply some sort of normalization to the signal.

Possible spectral features that have been suggested as useful for sound localization are spectral notches (Bloom, 1977; May and Huang, 1997) and covert peaks (Musicant and Butler, 1984). For the cat, there is evidence that notch detecting neurons have been found in the dorsal cochlear nucleus (Young *et al.*, 1992) and neural modeling using the cat's external ear transfer functions by Neti *et al.* (1992) revealed that some neurons in the network's hidden layer behave similarly to such notch detecting neurons. Unfortunately the situation may not be so simple for humans. Recent psychoacoustical work by Macpherson (1994, 1997) using irregularly shaped broadband sounds indicates that local features such as notches and peaks are too simple to account for human sound localization behavior. The determination of a specific algorithm to process spectral features was overstepped in this work by employing a time-delay neural network. The neural network optimizes internal network weights that correspond indirectly to ''spectral features.'' Unlike previous network models, this network incorporates both spectral and temporal information within its architecture.

The issue of the neural encoding for auditory space similarly blurs the boundary between a computational quantity and its corresponding algorithm. In other words, the contribution of the different acoustical cues toward the neural encoding of space depends intimately upon the nature of the spatial encoding itself. It is common to assume a description of space using either an azimuth-elevation angle coordinate system or a lateral-polar angle coordinate system. While

such coordinate systems seem natural, they have the subtle property of being discontinuous.[1] In this work, a direct representation of space in terms of a particular coordinate system was avoided by coding for location directly. That is to say, a local encoding of space was used in which neurons representing locations nearer to the sound source have a higher probability of firing (see Sec. II E 4). This derives from the well-established paradigm that the nervous system uses overlapping receptive fields to encode properties of the physical world. Therefore the time-delay neural network was trained to map spectral and temporal ''features'' of its input neural activity pattern to different locations in space using an encoding that peaked at the output neuron representing the target location of the input sound and then decayed away as one moves to neurons representing more distant locations. Significantly, the only restrictive assumptions placed upon the model by this realism constraint of overlapping receptive fields involve local properties of the encoding of space and not a specific global coordinate mapping.

II. METHODS

A. Overview

The sound localization performance of a human subject was tested in the free-field. The subject was one of the authors, a male of approximately 30 years of age with normal hearing in both ears, as determined by an audiometric examination, and is referred to as subject X. The sound stimuli consisted of three different bandpassed sounds: (1) a low-passed sound with frequencies 300–2000 Hz; (2) a high-passed sound with frequencies 2000–14000 Hz; and (3) a broadband sound with frequencies 300–14000 Hz. These frequency bands respectively cover conditions in which either temporal cues, spectral cues, or both dominate the localization process (the evidence for this fact derives from the ''intensity-theory'' and ''phase-theory'' of spatial coding, related to the duplex theory of sound localization; see Carlile, 1996b; Carlile *et al.*, 1999). Such sounds should hopefully allow for the integration of both spectral and temporal information during localization and provide for a rigorous examination of subsequent modeling. The subject performed five localization trials for each sound condition; 76 test locations, evenly distributed on the sphere, were used during each trial. As the responses to the low-pass stimuli showed an unbalanced attraction to the frontal hemisphere, the test conditions were arranged such that the low-pass and broadband stimuli were randomly mixed in equal proportions, while the trials with high-pass stimuli were performed in straight sets. Therefore in order to obtain a single trial result for the low-pass or broadband conditions, the subject needed to perform two tests with randomly mixed sounds. The trial results for an individual condition were then obtained by unscrambling the two tests.

Numerous studies have been conducted in which subjects are presented with spectrally restricted sound stimuli, as above, in an effort to clarify the role of the spectral cues in the localization process (e.g., Carlile *et al.*, 1999; Blauert, 1997; Butler, 1986; King and Oldfield, 1997; Middlebrooks and Green, 1992). These studies have shown that accurate sound localization requires a relatively large signal bandwidth (King and Oldfield, 1997) and that subjects tend to make *systematic* localization errors on spectrally restricted sounds (Carlile *et al.*, 1999). Furthermore, it is generally believed that the localization errors are related to the spectral filtering characteristics of the auditory periphery (Blauert, 1997; Middlebrooks and Green, 1992). Using sound stimuli similar to those described above, Carlile *et al.* (1999) have demonstrated that the mislocalizations of different subjects varied, but that for a given subject these localization errors were systematic. It is thus a basic premise of this work that the subject's systematic mislocalizations (demonstrated below) should be related in a meaningful way to the acoustical properties of his auditory periphery.

In order to incorporate the acoustical filtering of the outer ear into the modeling process, measurements were made of the subject's HRTFs at approximately 400 directions evenly distributed on the sphere. Directional transfer functions (DTFs) were then calculated from the HRTFs as in Middlebrooks and Green (1990). These DTFs were then used in the modeling process to simulate the directional aspects of the acoustical filtering of the outer ear. To establish that the DTFs appropriately indicate the direction of a sound source and were therefore an accurate input to the model, the subject repeated the identical localization task as above, but in a virtual sound environment. In this sound environment, the sound stimuli were filtered with the DTFs and played over in-ear tubephones so that the subject's cues to the location of the virtual sound source came entirely from the DTFs.

The sound localization process was then modeled using two basic system components: (1) a modified version of the physiological Auditory Image Model (Patterson and Allerhand, 1995; Giguère and Woodland, 1994) which simulates the spectro-temporal characteristics of peripheral auditory processing and (2) the computational architecture of a time-delay neural network (TDNN). A TDNN was chosen because of its ability to process and weight both spectral and temporal cues in a biologically plausible manner. Previous work has used neural networks without time delays, but with a number of inputs devoted to an ITD cross-correlation cue (e.g., see Backman and Karjalainen, 1993; Chung *et al.*, 2000) and other inputs devoted to the separate frequency channels corresponding to the spectral cues. Another technique, used by Macpherson (1997), is to artificially but equally weight the ITD cue versus the spectral cues.

B. Measuring sound localization performance

A detailed presentation of the methods used to train and test the subject in free-field sound localization as well as generate and present sound stimuli can be found in Carlile *et al.* (1997a). A short summary of the basic techniques is presented here.

1. The testing environment

The human localization experiments were carried out in a darkened anechoic chamber. Free-field sound stimuli were presented from a loudspeaker carried on a semicircular robotic arm. This arrangement allowed for placement of the speaker at almost any location on the surface of an imaginary

sphere, one meter in radius, centered on the subject's head. The perceived location of the sound source was indicated by the subject turning squarely to face the sound and then pointing his nose in the direction of the perceived source. The subject's head orientation and position were monitored using an electromagnetic sensor system (Polhemus, Inc.) consisting of a transmitter and receiver, the latter fastened to a headband worn by the subject. Between localization trials, the subject was required to align himself with the calibrated start position (azimuth 0°, elevation 0°) so that each stimulus was oriented correctly. This task was aided by a grid of colored light emitting diodes indicating instantaneous alignment error. A hand-held response button was used by the subject to both initiate a trial and indicate his completion of the localization task.

2. Stimulus generation

The noise stimuli were generated using D/A conversion at 80 kHz and delivered to a power amplifier (Quad 306) via a programmable attenuator (TDT: PA4). The loudspeaker (VIFA No. D26TG-35), mounted on the hoop-positioning system, had a response characteristic that was spectrally flat within 10 dB between 2 and 16 kHz and spectrally flat within 10 dB. Between 2 kHz and 300 Hz, the speaker had a high-pass roll-off of approximately 8 dB per octave. The sound stimuli were presented at approximately 70 dB SPL measured at the center of the stimulus positioning system. A ''fresh'' white Gaussian noise was appropriately bandpassed for each trial. The duration of the stimulus was restricted to 150 ms (with 10-ms raised-cosine onset and offset ramps) to ensure that head movement was not initiated during the presentation (Thurlow and Mergener, 1970; Carlile *et al.*, 1997).

C. Measurement of the directional transfer functions

Measurements of the subject's outer ear filter functions were carried out in the anechoic chamber. The measurements were made for both ears simultaneously using a ''blocked ear'' technique with a small Knowles electret microphone (EA–1954) placed in each ear (Møller *et al.*, 1995). Three hundred ninety-three measurements were made at locations evenly distributed on the sphere. In order to improve the signal-to-noise ratio, digitally constructed Golay codes with a 1024 impulse length were used as the recording stimulus (see Zhou *et al.*, 1992). The response of the microphone was bandpass filtered from 200 Hz to 16 kHz, digitized at 80 kHz, and averaged over 16 repetitions of the stimulus.

The transfer function of the outer ear was obtained by deconvolving the response of the microphone in the free-field from the response recorded at the entrance to the blocked ear (see Mehrgardt and Mellert, 1977; Wightman and Kistler, 1989b; Pralong and Carlile, 1994). The free-field calibration of the probe microphones was recorded without the subject in the anechoic chamber and at the position corresponding to the center of the subject's head. Following the deconvolution, the directional transfer functions (DTFs) were calculated as in Middlebrooks and Green (1990).

D. Evaluation of the directional transfer functions

A detailed presentation of the methods used to evaluate a subject's measured DTF recordings can be found in Pralong *et al.* (1996a). The DTFs should accurately capture, for a particular individual, all of the relevant acoustical cues to the direction of a sound source. Therefore sounds which are electronically filtered with the appropriate DTF for a particular direction in space should appear to come from that direction in space when played over headphones. To assess the quality of the DTF measurements, the subject repeated the previous localization test in virtual auditory space. The virtual sound stimuli were constructed by convolving the measured DTFs with the original sound source and then playing this sound to the subject using tubephones (ER-2, Etymotic Research). The tubephones are designed to produce an approximately flat frequency response, within 3 dB, at the human ear drum over the frequency range 200 Hz to 16 kHz. The ear-canal resonance, although removed by computing the DTF, is not reintroduced by the tubephones. The testing paradigm, itself, was identical to that in the free-field (Sec. II B).

E. A network model of sound localization

1. Overview of the modeling process

The sound localization model consisted of two basic system components: (1) a modified version of the physiological Auditory Image Model which simulates the spectro-temporal characteristics of peripheral auditory processing, and (2) the computational architecture of a time-delay neural network. The modeling process can be broken down into four stages as shown in Fig. 2. In the first stage a sound stimulus was generated with specific bandpass characteristics. The sound stimulus was then filtered with the subject's right and left ear DTFs to render an auditory stimulus originating from a particular location in space. The psychophysical data indicate that this stimulus is a good approximation to the real sound that would be heard by the subject in the free-field (see Fig. 7 below; Zahorik *et al.*, 1996; Hartmann and Wittenberg, 1996). The auditory stimulus was then processed by the Auditory Image Model (AIM) to generate a neural activity profile that simulates the output of the inner hair cells in the organ of Corti. This neural activity profile indicates the spiking probability of auditory nerve fibers. Finally, in the fourth and last stage, a time-delay neural network computed the spatial direction of the sound input based on the distribution of neural activity calculated by AIM.

2. Peripheral auditory processing of the sound input

The sound stimuli presented to the model were filtered with the subject's DTFs in the same manner as that used for evaluating the subject's sound localization ability in a virtual sound space (see Sec. II D). Following filtering, the physiological version of AIM, as opposed to the functional version, was used to process the stimulus because of its greater biological plausibility and its ability to include sound level dependencies. The standard physiological version includes its own external ear filtering. Therefore, with the aid of Christian Giguère (an author of the original software code), a

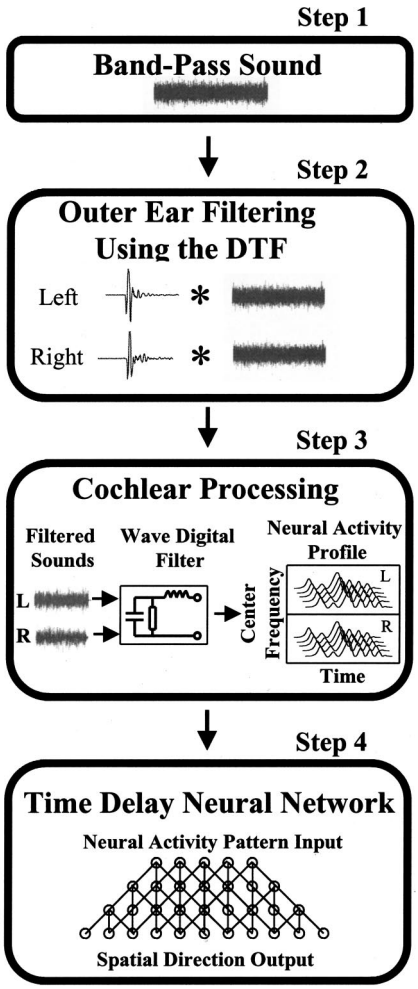


FIG. 2. The sound localization model.

software switch was implemented to disable this filtering under these conditions. The AIM model simulated the following: (1) the transmission through the middle ear; (2) the basilar membrane motion and cochlear hydrodynamics; (3) the fast motile mechanism of the outer hair cells; and (4) the neural transduction process of the inner hair cells. The distribution of cochlear filters across frequency was chosen such that the minimum center frequency was 300 Hz and the maximum center frequency was 14 kHz, with the density of filters being 5 filters per critical band, giving a total of 153 filters essentially equally spaced on a logarithmic scale (roughly 5 or more filters per critical band is required by the physiological version of AIM). A sampling frequency of 80 kHz was chosen to match the sampling frequency of the subject's HRTFs.

The output of the AIM model was downsampled from 80 to 6.7 kHz and the number of output channels reduced from 153 to 31, resulting in approximately one channel per critical band. Thus the auditory preprocessing was performed with high spectral and temporal resolution and reduction occurred only at the level of the neural activity profile. This downsampling amounts to making implicit assumptions (''realism constraints'') about the temporal and spectral resolution of the underlying physiology. On the assumption that the auditory nerve contains phase information up to 3 kHz, a 6.7-kHz sample rate should be suitable for capturing ongoing ITD information. Higher temporal resolution, however, would be required to effectively render the onset time information. As for the reduction in spectral resolution, the 31 cochlear channels seemed sufficient to produce reasonable localization performance with a manageable amount of computation. This spectral resolution is compared with that of other published models in Table I. To our knowledge, no systematic modeling has been performed to justify any particular density of frequency channels, although psychophysical evidence seems to indicate that human sound localization is surprisingly robust to reductions in spectral resolution (Leung *et al.*, 1998; Kulkarni and Colburn, 1998).

TABLE I. Comparison of different models and their spectral resolution.

Model	Begin frequency (kHz)	End frequency (kHz)	Frequency channels
Neti *et al.*, 1992 (binaural)	2	14	44
Neti *et al.*, 1992 (monaural)	2	14	88
Middlebrooks, 1992	3	16	40
Zakarauskas *et al.*, 1993	1	22	104
Chau and Duda, 1995	2	22	40
Janko *et al.*, 1997	0.38	17	23
Current model	0.3	14	31

3. The time-delay neural network

As a time-delay neural network was used to model the auditory system's sound localization process, a brief review of TDNNs is given (for a more complete description see Schenkel, 1995; Wan, 1993). In traditional artificial neural network models, the summing nodes of the neurons correspond to a simple static weighted sum and the connections via the summing nodes provide a spatial representation of the signal flow through the network. A fundamental aspect of the TDNN is that it essentially adds temporal filtering to the summing nodes of the neurons, thus incorporating a temporal dimension into the network computation. The temporal filtering is implemented as a discrete finite impulse response (FIR) filter,

$$s(k)=\sum_{n=0}^{N} w(n)x(k-n),$$

where k is a discrete time index, N is the length of the FIR filter, $x(k)$ is the input, and $s(k)$ is the filter output. This operation is shown in Fig. 3 using the standard representation of a tapped delay-line for the FIR filter. The output $y(k)$ from the node is then calculated by performing a soft thresholding operation using, in this case, the hyberbolic tangent,

$$y(k)=\tanh(s(k)).$$

Each connection node in the TDNN is formed using a FIR filter, such as described above. Therefore, a TDNN can be described as a multi-layer feed-forward neural network in which the outputs of a layer are buffered in an internal memory for several time steps and then connected to the next layer (see Fig. 4). The training of a TDNN involves adapting the FIR filter coefficients at the connection nodes to mini-

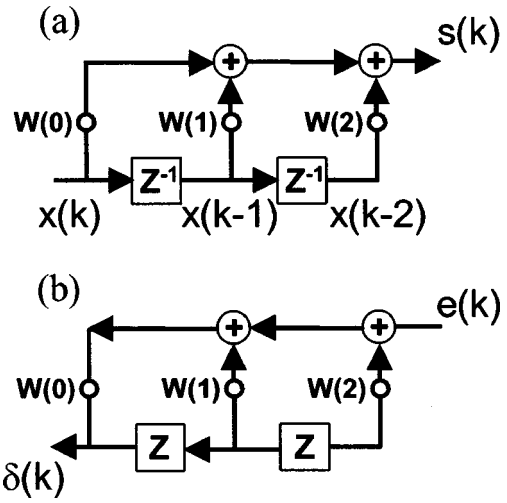


FIG. 3. (a) The summing node of a neuron in a time-delay neural network is represented as an FIR filter model with a tapped delay line. Z^{-1} represents a time domain unit delay operator. (b) Temporal backpropagation of error, e, through a single FIR filter to obtain the backpropagation δ-term. (Adapted from Wan, 1993.)

mize the computational error. An algorithm known as temporal backpropagation can be used to train the TDNN. The temporal backpropagation algorithm is akin to the traditional backpropagation algorithm and one formulation of the algorithm (see Wan, 1993) describes the backpropagated delta terms (as is commonly described in the regular backpropagation algorithm) as being calculated by filtering the error ''backwards'' through the FIR filter, with the reverse FIR filter described as shown in Fig. 3(b).

The layers of a TDNN, as described above, perform successively higher-level feature extraction. TDNNs have previously been applied to speech recognition and are well suited to sequential signal processing tasks (Waibel *et al.*, 1989; Lang and Hinton, 1988). The input fields of the neurons are restricted to a time window of constant duration and generally process the input signal sequentially, e.g., the network

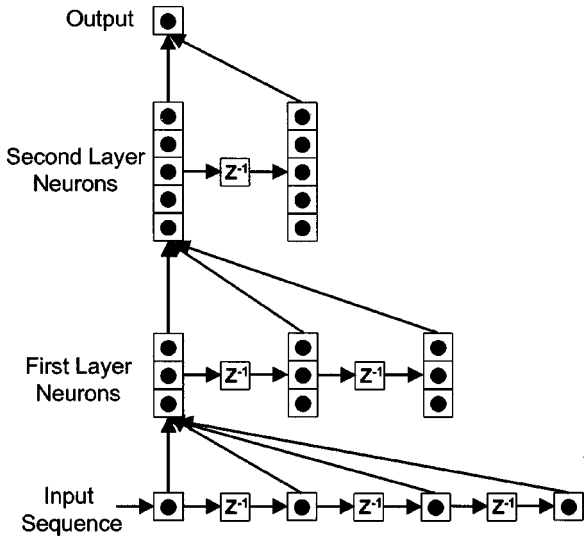


FIG. 4. A time-delay neural network. All outputs in a given layer are buffered a number of time steps and then connected to the next layer. (Adapted from Wan, 1993.)

''moves across'' the input. There are thus far fewer weights in this architecture than there would be in a fully connected network that processed the full extent of the input signal. In this case, the input to the TDNN was the neural activity pattern produced by the Auditory Image Model, consisting of the temporal signals in the different cochlear channels for the left ear and similarly for the right ear. Any given input neuron would then detect a particular local feature in the input signal (essentially the spiking probability of the different auditory nerve fibers). The receptive field of the neurons in the first computational layer was typically restricted to a time window of 2.25 ms.[2] This time window is of the same order of magnitude as the spectral integration time of 5 ms measured psychophysically using FM sweeps by Hofman and Van Opstal (1998). As the neurons move across the time axis, they detect the presence or absence of a particular local feature. By using several neurons at each time-step, the network detects numerous different features. This operation corresponds in essence to a convolution of the weight kernels of the neurons with the input. The time component of the input representation is gradually eliminated by subsampling the convolution at each layer. To partially compensate for the loss of information, the network usually and gradually increases the number of features. Such a structure is referred to as a bi-pyramidal network architecture and progressively converts time information into feature information. Through this structure, the activity of neurons at progressively higher levels depends on increasingly larger parts of the input. In this work, the last layer of the TDNN consisted of as many neurons as there were different classes of spatial directions. Thus, output scores were produced for each of 393 directions in space, evenly distributed on a sphere, with approximately 10 degrees of angular spread between directions. These scores can be interpreted as the relative probabilities that the given sound came from that particular direction in space.

4. A neural encoding of space

The spatial location of a sound source was encoded by the network as a distribution of neural activity in its output layer. Following Neti *et al.* (1992), this encoding was modeled on empirical data obtained from the superior colliculus of a cat, as demonstrated by Middlebrooks and Knudsen (1984). Such an encoding entails a distributed response with the peak occurring at the output neuron representing the target location of the input sound. The output response would then decay away in the form of a two-dimensional Gaussian as one moves to neurons farther away from the target location. It has been implicitly assumed that the output response has isotropic resolution on the sphere. The neural activity of a given output neuron was calculated, during training of the network, according to the following formula:

$$\gamma_i = e^{(\ln(0.9)\,((\alpha_i)^2/\kappa^2))},$$

where γ_i is the output activity for the ith neuron, α_i is the spherical angle between the sound input location and the location represented by the ith neuron and κ is a constant spherical angle determining the inherent precision in the output. As given by the formula, the output neural activity is scaled between 0 and 1 and will be within 90% of full value

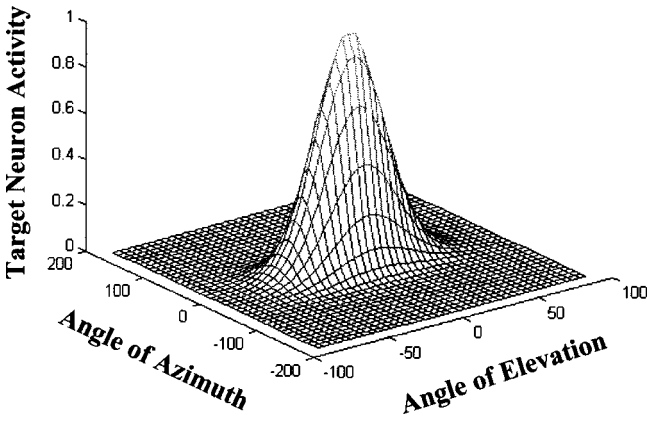


FIG. 5. Target neural activity for a sound located at zero degrees of azimuth and elevation.

when α_i is within the range specified by κ. In this work, κ was set at 10 degrees of spherical angle. Figure 5 shows an example target neural activity for an input sound located directly in front. Work by Chung *et al.* (2000) indicates that the value of κ is not particularly critical. Using a network without time delays, they showed that variations in the dispersion of the double Gaussian encoding between 10 and 60 degrees of spherical angle did not significantly affect the accuracy of their network's localization performance, in terms of the mean azimuth and elevation estimates. However, their modeling involves only the localization of broadband sounds, the localization of which would be most insensitive to parameter modifications.

III. NETWORK ARCHITECTURES

The major auditory brainstem nuclei demonstrate substantial frequency division within their structure (see Irvine, 1992). In other words, the systematic frequency organization of the primary auditory nerve fibers that innervate the cochlea seems to carry forward to the nuclei within the brainstem's auditory pathway. This arrangement is described as a tonotopic organization. Despite this fact and to our knowledge, no previous network model for sound localization incorporates such frequency division within its architecture. Therefore, in this work, three different architectures were examined with varying amounts of frequency division imposed upon the network structure. The different architectures are described below. Importantly, a network architecture was considered *different* only if there exists a fundamental difference in its structure. So, for example, two networks with a slightly different number of neurons in a hidden layer were *not* considered different.[3]

A. MLP architecture

As a basis for developing a TDNN architecture, a simple multi-layer perceptron (MLP) architecture without time delays was examined. The input to the MLP was calculated as the temporal average of the input data for the TDNNs. In other words, the output of each channel of the AIM is averaged over the duration of the stimulus so that the MLP gets one value per channel per ear. The number of neurons in each layer of the MLP is given in Table II and each layer of the network was fully connected to the previous layer (i.e., there was no frequency division within its structure).

TABLE II. The MLP network.

Layer	Neurons
Input	62
Hidden 1	40
Hidden 2	20
Output	393

B. TDNN architecture A

A TDNN architecture was created with a similar structure to the MLP, such that the relative number of neurons in different layers of the TDNN network were the same as that for the MLP. The description of a TDNN architecture is more complex than that for an MLP. In order to fully describe a computational layer of the TDNN, four characteristic numbers must be specified: (1) the number of neurons; (2) the kernel length, a number which determines the size of the current layer's time-window in terms of the number of time steps of the previous layer; (3) the kernel width, a number which specifies how many neurons in the previous layer with which there are actual connections; and (4) the undersampling factor, a number describing the multiplicative factor by which the current layer's time-step interval is increased from the previous layer's. Using this nomenclature, the architecture of the first TDNN, henceforth referred to as architecture A, is summarized in Table III and each layer of the network was fully connected to the previous layer.

C. TDNN architecture B

The brainstem nuclei involved in processing ILDs and ITDs are physiologically distinct with the separate streams represented in the lateral superior olive and the medial superior olive (Irvine, 1992). Therefore it was decided to impose a similar network structure on the second TDNN architecture, architecture B. This TDNN essentially consisted of the combination of two temporary and separate networks that were trained individually on low-pass (300 Hz to 2 kHz) and high-pass (2 kHz to 14 kHz) sounds. The assumption was that the network trained on low-pass sounds would emphasize the ITD information contained within the low-frequency cochlear channels, while the network trained on high-pass sounds would emphasize the ILD information contained within the high-frequency cochlear channels. The first hidden layers of both these networks were then combined to produce a hidden layer with twice as many neurons, which was then used as the first hidden layer of the TDNN with

TABLE III. The TDNN with architecture A.

Layer	Neurons	Kernel length	Kernel width	Undersampling
Input	62	...	...	...
Hidden 1	40	15	62	2
Hidden 2	20	10	40	2
Output	393	6	20	1

TABLE IV. The TDNN with architecture B.

Layer	Neurons	Kernel length	Kernel width	Undersampling
Input	62	⋯	⋯	⋯
Hidden 1	60	15	62	2
Hidden 2	20	15	60	2
Output	393	3	20	1

architecture B. Finally, this TDNN was trained identically to the TDNN with architecture A, except that the weights of the first hidden layer (which were trained previously) were held fixed, i.e., they were not updated during training. The detailed structure of the architecture is described in Table IV with the 60 neurons in the first hidden layer resulting from the combination of two layers of 30 neurons. As for architecture A, each layer of the network was fully connected to the previous layer.

D. TDNN architecture C

A third TDNN network architecture, architecture C, was developed with more substantial frequency division within its structure than the previous networks. This network's connections were constrained by frequency in a tonotopic-like arrangement (see Fig. 6). The 31 input cochlear frequency channels were split into ten overlapping groups consisting generally of six contiguous frequency channels. So, for example, the first group contained the frequency channels 1–6, while the second group contained the frequency channels 4–9 and so forth. Each input group of the first hidden layer contained 5 neurons. The kernel widths of these neurons were set, not to the total number of frequency channels in the input layer, but only to the six contiguous frequency channels defining the group. Information across the different groups of frequency channels was progressively integrated in the higher layers of the network. Thus, for example, the second hidden layer contained six groups of neurons connected to four or more different neuronal groups in the first layer. The size of the neuronal groups in the second layer was adjusted so that those spanning a greater frequency range contained more neurons. The exact connection arrangement can be seen in Fig. 6 and the architecture of the different layers is summarized in Table V.

TABLE V. The TDNN with architecture C.

Layer	Neurons	Kernel length	Kernel width	Undersampling
Input	62	⋯	⋯	⋯
Hidden 1	50	15	6	2
Hidden 2	28	10	4,5,6	2
Output	393	4	28	1

IV. TRAINING NETWORKS TO LOCALIZE SOUNDS

A. Learning algorithm

In terms of a neural system modeling paradigm, supervised training of a neural network is an optimization procedure, which to a first approximation should not depend on the learning algorithm. In other words, if a highly restrictive and specialized algorithm is required to achieve convergence upon a reasonable solution, then the model is likely to be poor and uninformative. In this work, the error backpropagation algorithm was used with a summed squared error measure (the most well-known derivation is given by Rumelhart *et al.*, 1986). Typically a training cycle was initiated with a learning rate of 0.05 and was decremented in stages with the training continued until a plateau was reached in the learning curve. The final average error per output neuron for a network solution depended on the structure of the training data with a typical value being 0.1 per neuron.

B. Structure of the training data

More relevant perhaps than the training algorithm, from a modeling point of view, is the learning task itself and the actual structure of the training data. In this work, one objective was to better approximate listening conditions in which a multitude of sounds are possible. Such conditions are, from an ecological point of view, arguably more plausible than listening conditions in which, for example, only broadband sounds are presented. In an effort to better approximate such listening conditions, a training method that uses sounds with varying bandwidth and center frequency was used. Such a training method presents a challenging problem because the network must learn to localize a sound based on features that are not certain, but only probable given the sound environment. For example, the large spectral contrast at the band-edge of a bandpass sound is not truly indicative of the sound's location. However, if a spectral notch in an HRTF

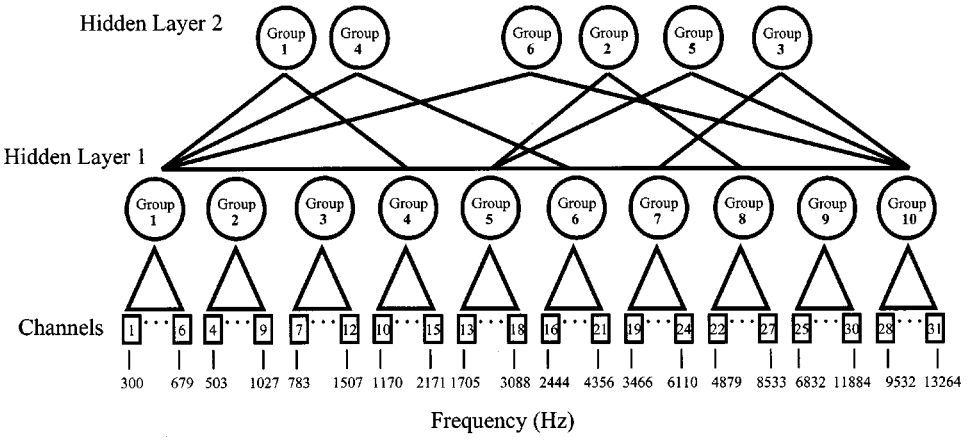


FIG. 6. Tonotopic arrangement of neurons for architecture C. Each group contains a number of neurons, e.g., five for the first hidden layer. See text for more details.

TABLE VI. A comparison of the total weights, VC dimension, and the number of patterns used for the different networks.

Network	Total weights	VC dimension	Patterns
MLP	11 593	2.4×10^5	55 020
TDNN A	95 518	2.0×10^6	220 080
TDNN B	99 759	2.1×10^6	220 080
TDNN C	61 416	1.3×10^6	220 080

for a given location occurs at the frequency of the band-edge, then large spectral contrast at that frequency can *sometimes* be indicative of the sound's location. The notion of ''probable'' spectral features is well described by Rogers and Butler (1992) and Butler and Musicant (1993).

In order to better approximate listening conditions with varying sound stimuli, two different sound arrangements were used for training the networks. The first arrangement consisted of three types of stimuli embodying the test conditions: low-pass, high-pass, and broadband noise. These were presented with equal one-third probability. In a second arrangement, the center frequency and bandwidth of the noise were chosen randomly. That is to say, the bandwidth was chosen uniformly and randomly from the number of allowable cochlear channels (i.e., a number between 2 and 31 channels inclusive). The center frequency was then chosen uniformly and randomly from those cochlear channels that could accommodate the given bandwidth. In this way, the band-edge frequencies were always chosen to be within the frequency range 300–14000 Hz. For example, suppose that a bandwidth of 15 cochlear channels was chosen. The center frequency would then be chosen from the cochlear channels between 8 and 24 inclusive. Thus the noise bandwidths were based on the same logarithmic frequency scale as the cochlear channels.

Additionally, one had to ensure that an appropriate number of training patterns was used. A list of the total number of weights and the Vapnick-Chervonenkis (VC) dimension for the different networks is given in Table VI. The VC dimension is a scalar value providing a measure of the intrinsic storage capacity of a network, i.e., the number of random patterns that can be stored and was calculated as described in Baum and Haussler (1989). As shown, the number of connection weights, or free parameters of the model, is large and varies between approximately 10 000 and 100 000. Therefore an even larger number of patterns was required during training; approximately 200 000 patterns were used (see Table VI). The training stimuli generally consisted of 2800 directional sounds and the estimated equivalent number of training patterns, shown in Table VI, was calculated based on the double Gaussian output encoding and the convolutional structure of the TDNNs. In actuality, the most practical criterion for determining an appropriate number of training patterns was to empirically observe the point at which the network's localization performance on unseen test sounds was equivalent to that for similar training sounds.

C. Frequency selective training

In association with the varying frequency structure of the training data and the tonotopic network architecture, the backpropagation training algorithm automatically provides, in this case, a frequency selective learning algorithm. It is a feature of the learning rule that a neuron's weights are only updated when there is input activity in the corresponding connections. As the input cochlear frequency channels are connected to separate groups of neurons in the first computational layer, some bandpass sounds will not activate all of the neuronal groups. So, for example, a training sound containing only low frequencies will not train the high-frequency neurons and vice versa.

V. METHODS OF ANALYSIS

A. Quadrature view

The extent of the qualitative differences in the localization performance of the competing models was sufficient to clearly distinguish between them. The distribution of localization responses are presented on four different spherical plots, referred to as a quadrature view. The four plots illustrate the front, back, left, and right hemispheres of space looking in toward the subject. Each perspective shows a hemisphere of space as seen from the outside looking in (e.g., see Fig. 7). Thus the localization response for any sound target located in the frontal hemisphere is shown on the ''front'' plot. Similarly, the response for any target location toward the rear of the subject is shown on the ''back'' plot and so on. The target locations are indicated by crosses and, depending upon the number of localization responses for a given target, either the response or the centroid of a number of responses are denoted by a filled circle. A solid line is drawn connecting the target and response locations along a great circle.

B. Spherical statistics

Methods of spherical statistics were used to analyze the experimental psychophysical data. The localization responses of the subject for a given bandpass condition and presentation method (e.g., low-pass, high-pass; free-field, virtual space) were pooled across the five trials for that condition. For a specific target location, the mean direction of the localization judgments was calculated in terms of the ''centroid,'' a measure of central tendency in spherical statistics (see Fisher *et al.*, 1987). In addition, the standard deviation about the centroid was modeled using either a Fisher or Kent distribution, depending upon whether the data points were symmetrically distributed (Fisher) or best fit by two principal components (Kent). A detailed explanation of these methods is given in Leong and Carlile (1998). Thus when the subject's localization performance is shown in quadrature view, the standard deviation about the centroid is depicted by an ellipse with the length of the major and minor axes indicating the amount of variation.

VI. RESULTS

A. Localization performance of the subject

The subject's localization performance in the free-field as compared with that in the virtual auditory space is demonstrated in Fig. 7. Human free-field localization is shown in

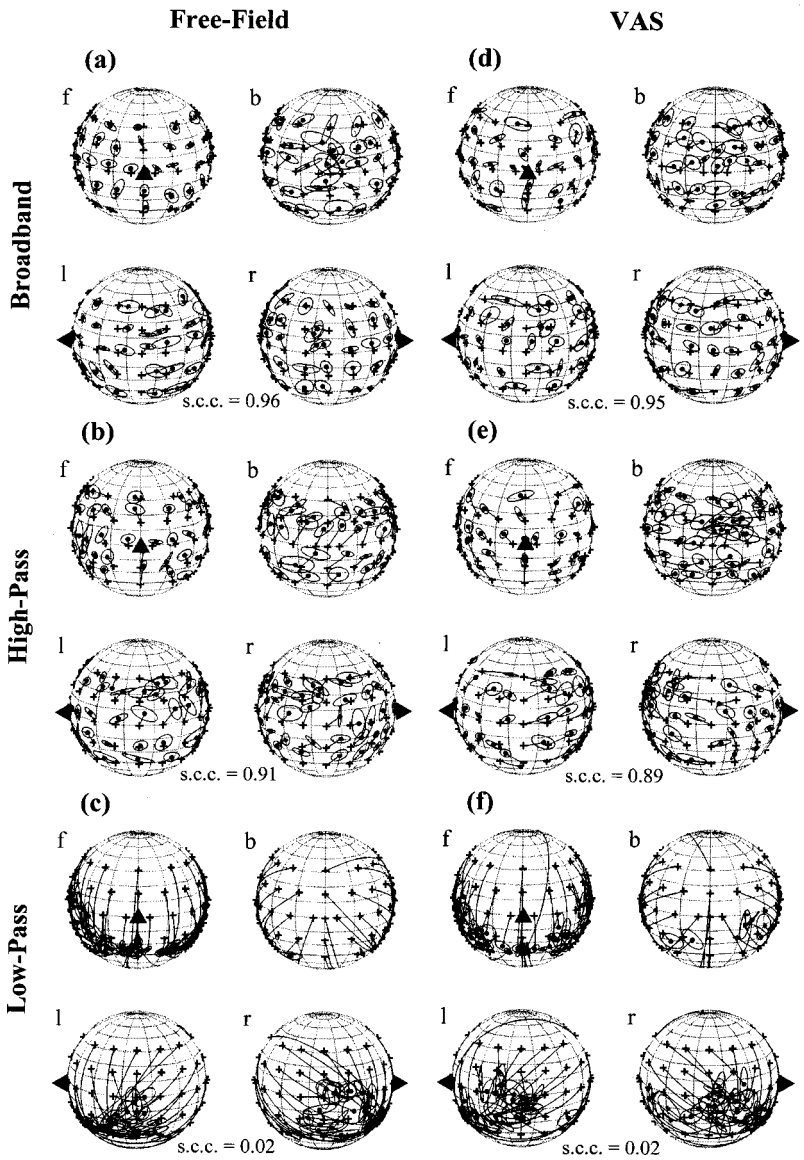


FIG. 7. The localization responses for subject X in both the free-field and virtual space are plotted in quadrature view showing the front (f), back (b), left (l), and right (r) hemispheres of space. To assist in identifying the spatial orientation, a solid black triangle indicates the nose position. The free-field responses are shown on the left and the virtual auditory space responses are on the right. A cross denotes the target location, a filled circle denotes the centroid of the perceived locations and a solid line joins the two. The distribution of localization errors is shown by an ellipse, whose major and minor axes indicate standard deviation. The spherical correlation coefficient (s.c.c.) is indicated at the bottom of the quadrature plots. (a) Free-field broadband localization. (b) Free-field high-pass localization. (c) Free-field low-pass localization. (d) Virtual source broadband localization. (e) Virtual source high-pass localization. (f) Virtual source low-pass localization.

the first column, Figs. 7(a)–(c), while the human virtual auditory space localization is shown in the second column, Figs. 7(d)–(f). The three different bandpass conditions are illustrated in three rows and consist of broadband [Figs. 7(a), 7(d)], high-pass [Figs. 7(b), (e)], and low-pass [Figs. 7(c), (f)] noise. As discussed earlier (see Sec. II), the contrast between the subject's free-field localization and virtual auditory space localization should portray the combined influence of the following differing conditions: (1) The free-field sounds were spectrally flat to within 10 dB between 2 and 16 kHz and there was a rather steep high-pass roll-off (8 dB/octave) between 2 kHz and 300 Hz, while the virtual sound sources were spectrally flat within 3 dB between 300 Hz and 16 kHz. (2) The acoustical cues provided naturally during the free-field localization are embodied in the HRTFs, whereas the acoustical cues provided via electronic filtering during the virtual auditory space localization are embodied in the DTFs. Despite these differing sound conditions, the localization data demonstrate that the performance between the two sound environments was convincingly similar. This suggests that the DTFs were an accurate measure of the psychophysically relevant components of the HRTFs in the free-field condition. The data show that the subject performed well in both the broadband and high-pass sound conditions and rather poorly in the low-pass condition. In both the broadband and high-pass sound conditions [Figs. 7(a), (d); 7(b), (e)], the front-back errors were in the minority of responses and were removed when calculating the standard deviation of the responses about the centroid. Front-back errors were defined as localization responses in which the perceived location crossed the vertical plane through the interaural axis which separates the anterior and posterior hemispheres of space.[4] For the low-pass condition, the subject demonstrated gross mislocalizations, with the responses clustering toward the lower and frontal hemispheres. The distribution of errors for the low-pass condition was again based on the majority response such that front-back differences were removed; the gross mislocalizations correspond mainly to the traditional cone of confusion errors (see Carlile *et al.*, 1999).

As a quantitative measure of the similarity between the two sound environments, the spherical correlation coefficient

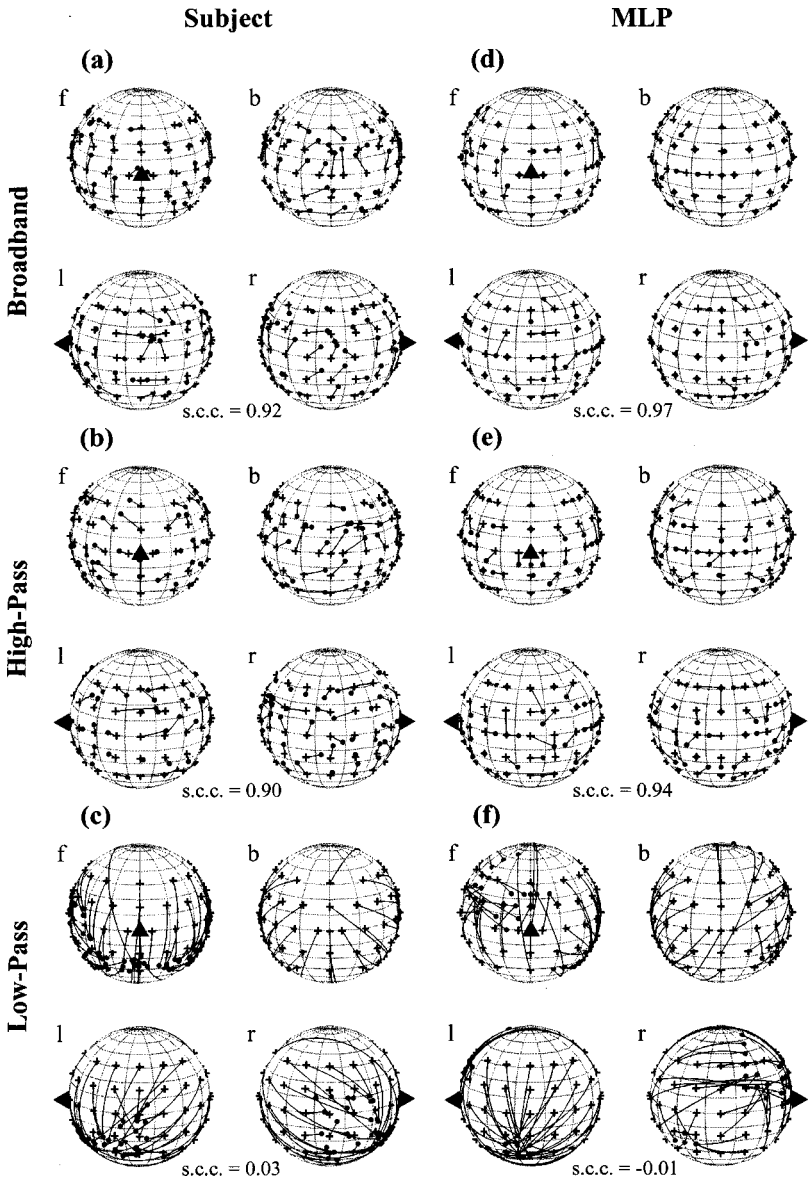


FIG. 8. The localization responses of the MLP are compared with that for subject X in the free-field. The responses are plotted in quadrature view with the MLP localization on the right and the subject's localization on the left. A cross denotes the target location, a filled circle denotes the perceived location and a solid line joins the two. (a) Free-field broadband localization. (b) Free-field high-pass localization. (c) Free-field low-pass localization. (d) MLP broadband localization. (e) MLP high-pass localization. (f) MLP low-pass localization.

(see Fisher *et al.*, 1987) between the target and perceived locations was computed for each experiment. These data consisted of five trials with 76 locations, yielding a total of 380 localizations. A coefficient of 0.96 was obtained for the free-field environment with 2% of the responses being front-back errors. In the virtual sound environment, a similarly high correlation coefficient of 0.95 was obtained with 5% of the responses being front-back errors. The high degree of correlation between the target and perceived locations in the virtual sound environment again suggests that the synthesized virtual sound sources were highly effective and, most importantly, are suitable for modeling the localization process. Certainly, however, the increased number of front-back errors that occurred in the virtual sound space suggests that it is not an exact reproduction of the free-field. In this case, it is not known whether these differences were caused by signal differences or behavioral and performance issues. Nonetheless, these quantitative measures indicate that both sound environments conveyed much the same acoustical information with respect to sound location.

B. Localization performance of the MLP

The localization performance of the multi-layer perceptron is shown in Fig. 8. This particular network was trained with equal probability on the three bandpass conditions. The localization responses of the network are plotted side by side with the localization responses of the subject in the free-field. This is meant as an aid to the reader in assessing the model's performance. The localization data are shown in quadrature view and arranged such that the subject's responses are in the left column [Figs. 8(a)–(c)] and the model's responses are on the right [Figs. 8(d)–(f)]. The localization responses for the three bandpass conditions are illustrated in different rows. Only single trial results are shown for the models as a qualitative examination of the accuracy of the localization responses was sufficient, rather than the more detailed and quantitative analysis of the variance of replicate responses. It was for this reason that no effort was made to smoothly interpolate or otherwise refine the output of the network to have a precision better than

TABLE VII. Comparison of the localization performance results for the subject and the competing models.

	Sound condition	Spherical correlation coefficient	Mean spherical angle error	Percentage front-back error
Subject X	Broadband	0.92	11	0
	High-pass	0.90	12	1
	Low-pass	0.03	50	37
MLP	Broadband	0.97	5	0
	High-pass	0.94	9	0
	Low-pass	−0.01	53	42
TDNN	Broadband	0.89	9	1
architecture A	High-pass	0.36	35	26
	Low-pass	0.00	54	59
TDNN	Broadband	0.93	10	0
architecture B	High-pass	0.92	10	1
	Low-pass	0.25	34	20
TDNN	Broadband	0.94	9	0
architecture C	High-pass	0.92	10	0
Fixed training	Low-pass	0.09	40	30
TDNN	Broadband	0.96	7	0
architecture C	High-pass	0.85	15	1
random training	Low-pass	0.18	37	26

approximately ten degrees of spherical angle, which would have been required for any meaningful interpretation of response variation. The MLP illustrated in Fig. 8 demonstrates good localization performance for both the broadband and high-pass sound conditions and much poorer performance for the low-pass sound condition. The performance of the MLP is quantitatively compared with the human subject in Table VII using the spherical correlation coefficient, the average spherical angle error and the percentage of front-back confusions. There is evident similarity, but with some differences, between the subject's localization performance and that of the MLP; notably, there are some structural similarities in the low-pass sound mislocalizations. The network's low-pass sound localization in the left hemisphere of space is sometimes pulled, similarly to that of the subject, toward the lower and frontal hemisphere of space. However, unlike the subject's localization performance, the MLP demonstrates significant asymmetry between the left and right hemispheres of space for the low-pass stimulus. It is unlikely that the response asymmetry was caused by an error in the model because: (1) the same calculations are performed independently of the stimulus location; (2) the model demonstrates symmetry in its localization performance for the broadband and high-pass sounds. It is also unlikely that there was an error in the sound input because the subject localized the same sounds as the MLP. It can, of course, be argued that the subject's acoustical filtering at the left and right ear (the HRTFs) is asymmetric, but to a first approximation it is likely that the acoustical filtering is essentially bilaterally symmetric. Thus this asymmetry shown in the *low-pass* condition may indicate that the training of the network was insufficient to develop meaningful structure in the network weights corresponding to the low-frequency cochlear channels. There is also likely to be some redundancy in the information across the different cochlear frequency channels, which may complicate the sufficient development of the network weights corresponding to the low-frequency cochlear channels. This gives some indication that training methods using sounds with varying frequency content may be an important consideration when examining neural network models of sound localization.

These results, in general, indicate that the methodology used for preprocessing the input sound stimuli was reasonable and that the error backpropagation algorithm could reasonably extract information relating to the source location and thus paved the way for developing the TDNN, a model with a substantially more realistic ability to extract temporal as well as spectral information from the sound stimulus.

C. Localization performance of the TDNN with architecture A

The localization responses of the TDNN with architecture A, the architecture most similar to the MLP, is shown in Fig. 9. A quantitative comparison of this network's performance with that of the human subject is also given in Table VII. The training for this network was identical to that for the MLP. Despite the similarities between the two networks, the TDNN failed to localize sounds across the different bandpass conditions. This TDNN shows poor localization performance in the high-pass condition, in contrast to the human subject, and a considerable clustering of responses in the low-pass condition toward the left and right interaural axes, again in contrast to the human subject where the responses cluster at lower frontal elevations. This result is representative of several networks similar in structure to architecture A, but with a variable number of hidden neurons and minor differences in training regimes. This TDNN did, however, show reasonable performance in the broadband sound condition. The poorer localization performance of the TDNN as compared with the MLP could be attributed to the incorporation of temporal information within the frequency channels. As the training methods for the two networks were identical, one must examine the architectures and the input data to find the cause behind the differences. As described in Sec. III, the architectures were the same except for the convolutional structure imposed by the TDNN. The TDNN's weights are convolved with the input signal over time and thus can incorporate temporal information from the input data into its processing. Given that there is temporal information in the signal, this information could interfere with the processing of the spectral information which results in differences in the localization performance of the two networks. Error in the backpropagation algorithm can be eliminated because of the reasonable performance by the TDNN in localizing broadband sounds. Thus it appears most likely that there has been a failure of the TDNN to robustly incorporate temporal and spectral information into the localization process across the different bandpass conditions. As will be seen in the following sections, this inadequacy to incorporate both temporal and spectral information is removed, to a great extent, by dividing the input frequency channels into different groups and then progressively interconnecting the neurons in the higher layers across frequency.

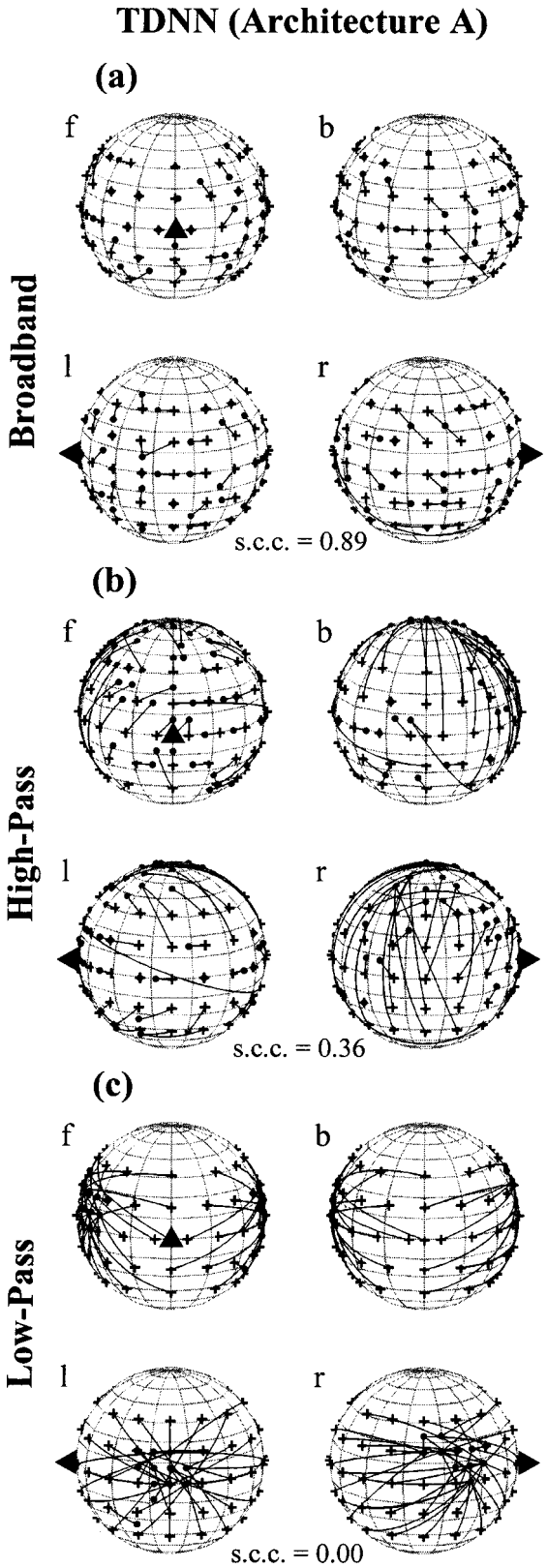


FIG. 9. The localization responses of the TDNN with architecture A is shown in quadrature view. Other details as in Fig. 8. (a) Broadband localization. (b) High-pass localization. (c) Low-pass localization.

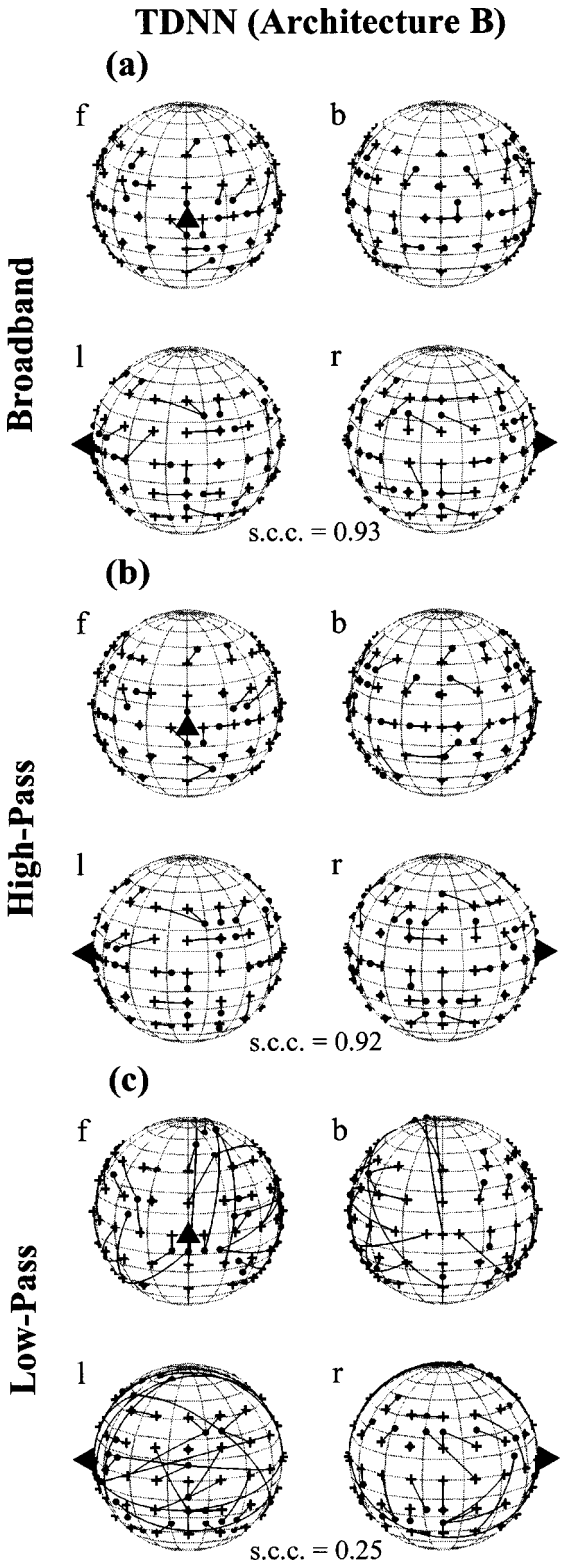


FIG. 10. The localization responses of the TDNN with architecture B is shown in quadrature view. The first hidden layer of this network is combined from two separate networks that were trained on high-pass and low-pass sounds, respectively. Other details as in Fig. 8. (a) Broadband localization. (b) High-pass localization. (c) Low-pass localization.

D. Localization performance of the TDNN with architecture B

The localization performance of the TDNN with architecture B is shown in Fig. 10. The localization is good in both the broadband and high-pass conditions and poorer in the low-pass condition. Compared to the subject's localization, there was less aggregation of responses to a particular region in the low-pass condition. A quantitative comparison

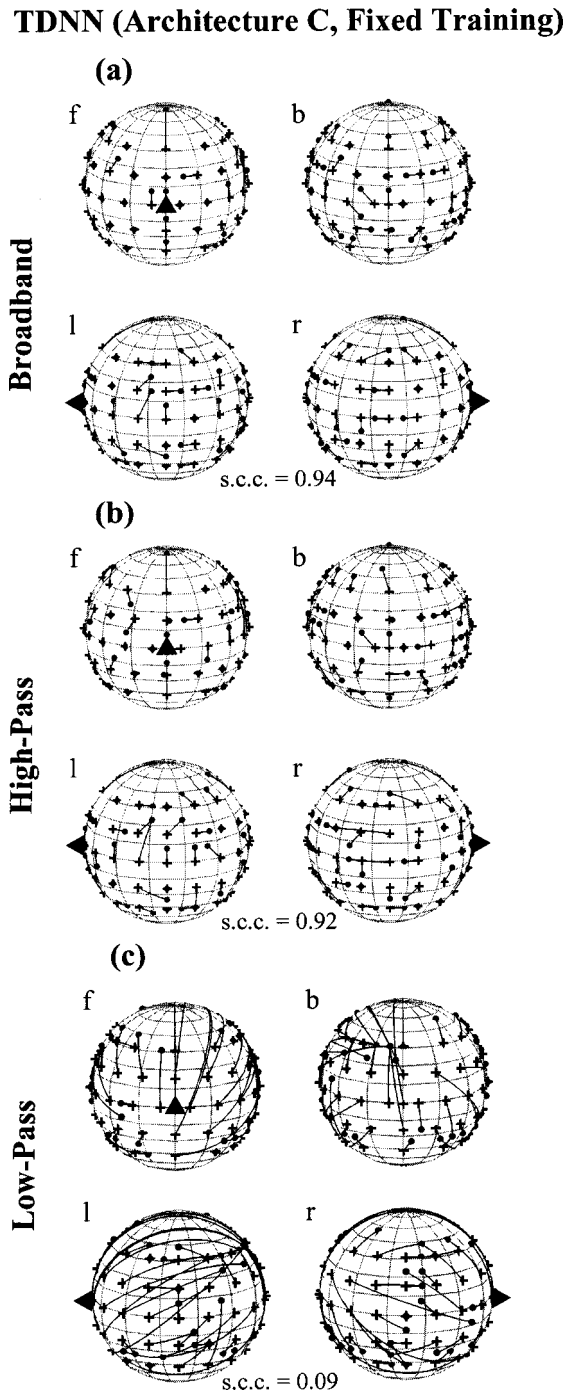


FIG. 11. The localization responses of the TDNN with architecture C is shown in quadrature view. This network has a tonotopic structure and training was performed equally on low-pass, high-pass, and broadband sounds. Other details as in Fig. 8. (a) Broadband localization. (b) High-pass localization. (c) Low-pass localization.

of the network's performance with that of the human subject is also given in Table VII. Most importantly, compared to the TDNN with architecture A, this TDNN was able to robustly incorporate temporal information into the localization process across the different bandpass conditions. Thus the frequency division imposed upon this network's computational architecture improved the network's localization performance.

E. Localization performance of the TDNN with architecture C

The TDNN with architecture C had a more realistic tonotopic structure (see Sec. III D) and was trained separately on the two different sound arrangements described in Sec. IV B. The results obtained when this TDNN was trained equally on the broadband, high-pass, and low-pass sounds are shown in Fig. 11. The localization performance was good for both the broadband and high-pass sound conditions and poorer for the low-pass condition, which produced many front-back errors. More interesting, perhaps, is the network's localization performance when trained on sounds with random center frequencies and bandwidths. Not only did such training better approximate listening conditions with variable sounds, but it also provided a more substantial investigation into the network's localization abilities. As discussed above, when randomly bandpassed sounds were presented during training, the learning task becomes considerably more complicated. The network was exposed to more than just three bandpass sound conditions during training. Nonetheless, this model produced localization performance across conditions that compared reasonably well with the localization of the human subject (see Fig. 12). Compared to the responses of the other TDNN models, this TDNN's low-pass localization performance shows a closer resemblance to that of the subject; many of the localization responses were pulled downward and toward the front. A quantitative description of the performance of both of the networks described in this section are given in Table VII.

The localization performance of both the MLP and the TDNN with architecture C have demonstrated similarities with the human subject's performance across the three different bandpass conditions (as discussed in previous sections). Nonetheless, the TDNN possesses the added advantage of incorporating temporal information along with the spectral information during the localization process. This additional and substantial realism constraint supplied the motivating factor for using the TDNN with architecture C in the continuing experiments that follow. These experiments probed the localization process using sounds with only restricted high frequencies and varying level.

F. Restricted high-frequency sound localization

Because the training data for TDNN (architecture C) were not fixed to any particular frequency region, it was reasonable to expect the network's performance to generalize, for example, to a restricted high-frequency condition. Therefore the subject's sound localization ability was further tested in the free-field using sounds bandpassed between 7637 and 13264 Hz.[5] The subject's localization results for a single test with 76 trial locations are shown in Fig. 13(a) and can be contrasted with that for the TDNN, shown in Fig. 13(b). The localization performance of the TDNN again shows qualitative similarities to that of the subject, with both demonstrating a clear aggregation of responses to restricted spatial regions in the lower hemisphere. A quantitative comparison of the TDNN's localization performance with that of the human subject is given in Table VII. The localization

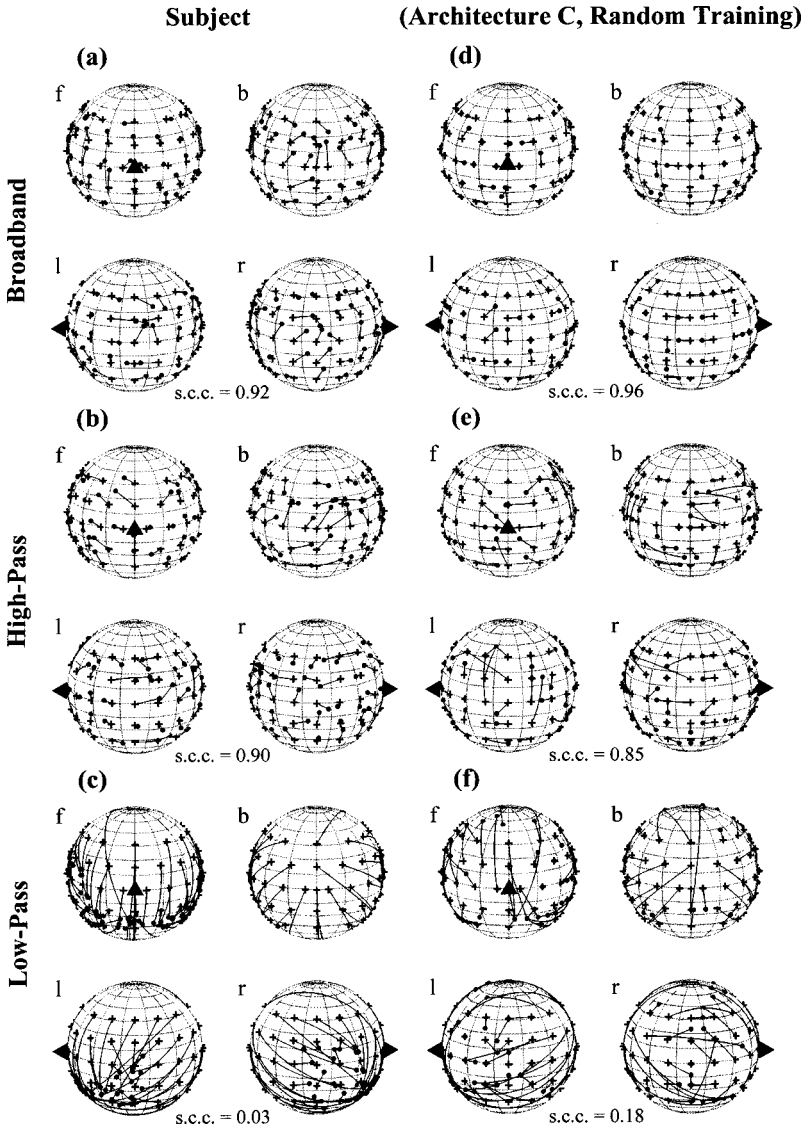


FIG. 12. The localization responses of the TDNN with architecture C are compared with that for subject X in the free-field and shown in quadrature view. The TDNN had a tonotopic architecture and was trained using sounds with random center frequency and bandwidth. Other details as in Fig. 8. (a) Free-field broadband localization. (b) Free-field high-pass localization. (c) Free-field low-pass localization. (d) TDNN broadband localization. (e) TDNN high-pass localization. (f) TDNN low-pass localization.

performance of both the subject and the TDNN, as shown here, are contrasted below with that of a matched filtering model.

G. Sound localization at varying sound levels

The sound level of the sounds used to train the networks were set and controlled via a parameter in the Auditory Image Model. The training of the TDNNs was performed in such a fashion, that for any particular location in space, the sound level did not vary by more than 1 dB SPL during repeated presentations of the sound. It was of interest to examine the robustness of the model's localization performance under the influence of varying sound level. Thus, the localization performance of the TDNN (architecture C) trained on sounds with random center frequencies and bandwidth was examined, using a broadband sound source, across a range of sound levels varying from 60 to 80 dB SPL. The sound level used for training was approximately 67 dB SPL. It was observed that the transfer function of the inner hair cell model was essentially linear for this range of sound levels (within AIM, a medium spontaneous-rate auditory nerve fiber was used with a threshold shift of 0 dB, see Giguère and Woodland, 1994). The performance of the network was measured in terms of the spherical correlation coefficient between the target and response locations and is shown in Fig. 14. The spherical correlation coefficient peaked, of course, at the training sound level of 67 dB SPL and remained above 0.8 between 60 and 75 dB SPL. Angular errors corresponding with the correlation of 0.8 were typically between 10 and 15 degrees of spherical angle. In other words, a graceful degradation in localization performance was demonstrated over a range in sound level of 15 dB.

The same TDNN, as above, was also tested on broadband sounds, 10 dB louder in one ear than the other. The results of these tests are shown in Fig. 15 and clearly illustrate that the localization responses were pulled toward the side with the louder sound [Figs. 15(a), (b)]. The localization performance of the model was much improved when the sound level to both ears was increased [Fig. 15(c)]. While the magnitude of the effects shown here are certainly not human-like (see Wightman and Kistler, 1997), such behavior suggests that interaural level difference cues were a prominent and constant feature of the data that conferred some measure of robustness (in the model) to sound level variations.

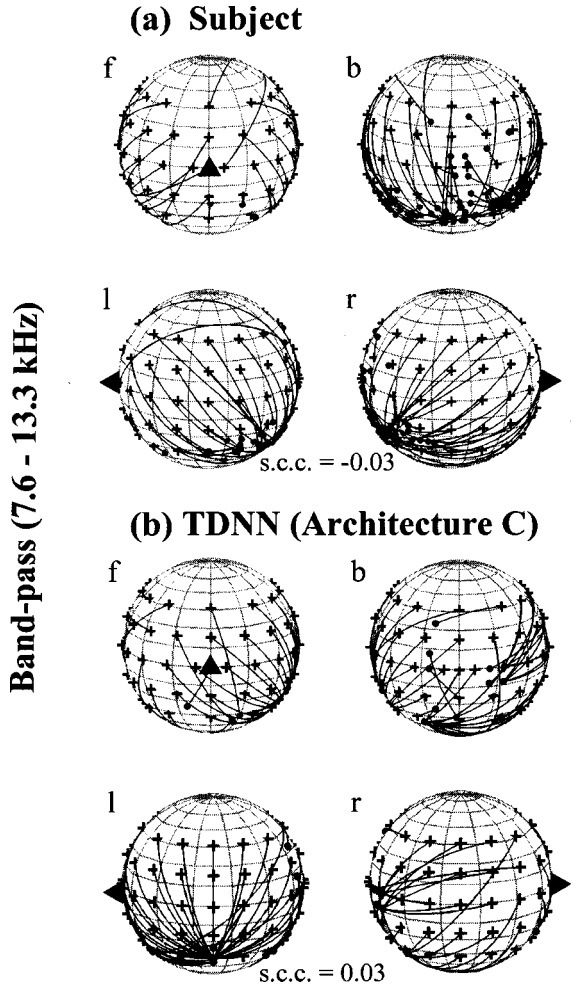


FIG. 13. The localization performance of both subject X and the TDNN with architecture C (trained using sounds with random center-frequency and bandwidth) are shown for the restricted high-frequency condition consisting of sounds bandpassed 7637–13264 Hz. Other details as in Fig. 8. (a) Subject's free-field localization. (b) TDNN localization.

H. TDNN encoding of temporal information

It has been argued above that the TDNNs encode temporal information as well as spectral information. Direct evidence for this fact is indicated in Fig. 16 where the weights for the first hidden layer of the TDNN with architecture C (trained on sounds with random center frequencies and bandwidth) are displayed. The weights for the low-frequency neurons clearly demonstrate a spectro-temporal structure while the weights for the high-frequency neurons indicate a constant spectral structure across time. This is in accordance with what one would expect given both the information provided via the auditory nerve response and the basic acoustics of sound diffraction around the head (Lord Rayleigh, 1907).

VII. DISCUSSION

A. Tonotopic processing from a computational viewpoint

As Secs. VI C–VI E clearly demonstrate, a network structure with frequency division, one that divides the input frequency channels into different groups and then progressively interconnects the neurons in the higher layers across

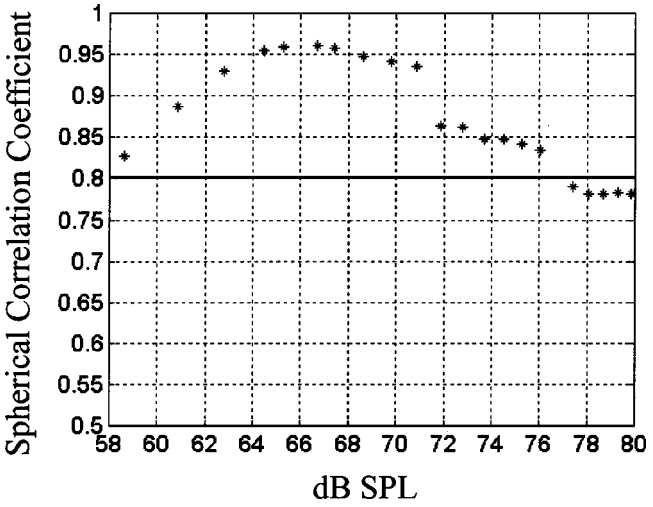


FIG. 14. The change in the localization performance of the TDNN with architecture C (trained using sounds with random center-frequency and bandwidth) is shown as a function of the input sound level. The overall localization performance was reasonable (i.e., s.c.c.$>$0.8) over a range of 15 dB.

frequency, was more robust in its localization performance across sounds with variable center frequency and bandwidth than a simple fully connected network. There are a few likely explanations for this observation. One line of reasoning argues that it may be easier for the tonotopic network to prevent a narrow band of frequency channels from dominating the localization computation across the entire set of sound stimuli. It was observed, in data not shown, that the low-frequency neurons required more training cycles to fully develop their weight structure than the high-frequency neurons. This observation is in accordance with the fact that these neurons demonstrate both a spectral and temporal structure in their connection weights (as shown in Fig. 16). The mismatch in training rates provides at least one source of conflict between the low- and high-frequency neurons. In the case of the TDNN with architecture A, in which each neuron within the first layer received input across the entire frequency range, the information within the different frequency channels may have destructively interfered with one another. The result was poor localization performance (Fig. 9).

Similar to the line of reasoning above, it can be argued that the frequency division imposed upon the network structure more evenly spreads the computation among the neurons in the different frequency groups and thus produces a more uniform computational structure. Evidence in line with this argument is the fact that for any single bandpass condition, it was found that the TDNN did not require frequency division within its architecture to produce quality solutions when trained *only* on these bandpassed sounds. In other words, it is the combination of frequency division and varying sound stimuli that encouraged the network to develop meaningful connections for all frequencies. The notion of a more uniform computational structure, as just described, is similar to two other ideas in the neural network literature. One is the notion of fault-tolerance in neural networks, an idea well explained by Neti *et al.* (1992) in their neural network models of the sound localization of the cat. Neti *et al.*

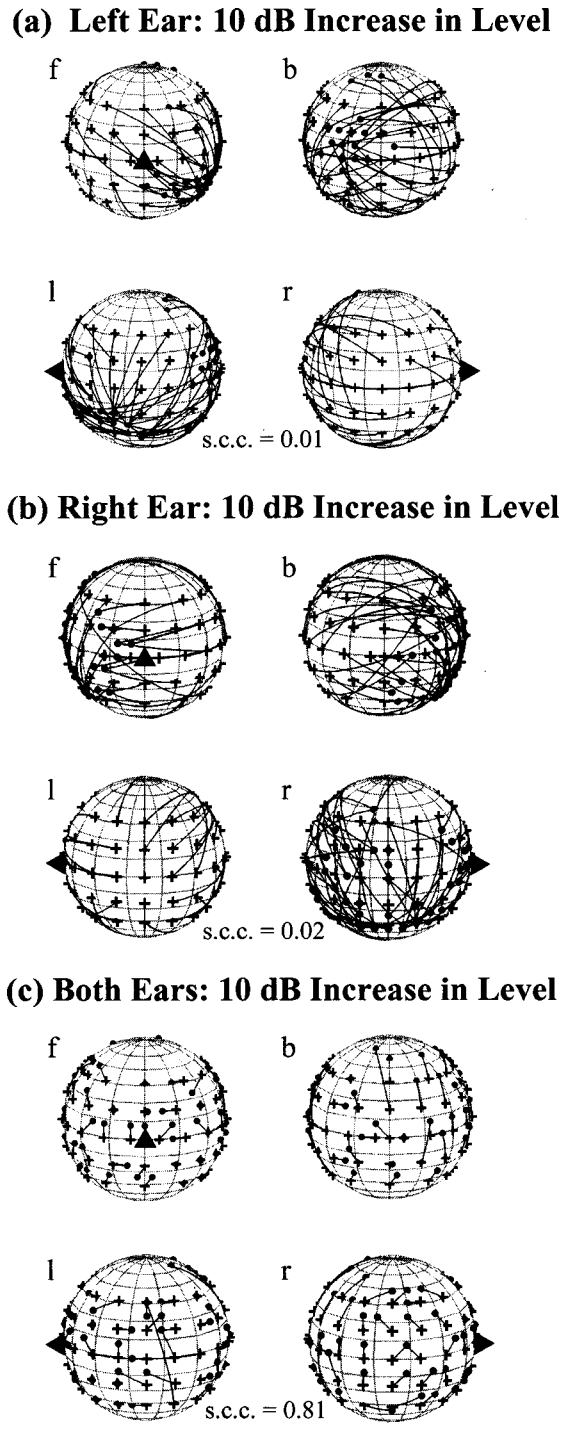


FIG. 15. The localization performance of the TDNN with architecture C (trained using sounds with random center-frequency and bandwidth) is shown (a) for sounds 10 dB louder on the left only, (b) 10 dB louder on the right only, and (c) 10 dB louder at both ears.

do not argue specifically for the connection between a tonotopic structure and fault-tolerance. However, they did develop a novel learning algorithm that minimizes a network's error subject to the constraint that the increase in the error caused by removing a neuron from a computational layer of the network is uniformly small for all neurons within that hidden layer. To the extent that there is some overlap between spreading the computation across frequency and spreading the computation across neurons, one may encourage the other. It has also been argued (Shinn-Cunningham, 1998) that competitive learning with structural inhibition may allow the TDNNs to train appropriately across different bandpass conditions without requiring a tonotopic architecture. Such an argument assumes that competition would assist in distributing the computation across different neurons and supposedly, therefore, across frequency. Nonetheless, the network structure with frequency division, as demonstrated here, provided sufficient means for robustly training the TDNNs across the different bandpass conditions and thus played an important role in the computational process.

B. Matched filtering and sound localization

A number of previous sound localization models have used a relatively straightforward matched filter or template matching analysis (e.g., Middlebrooks, 1992; Nandy *et al.*, 1993; Macpherson, 1997). In such cases, the ITD and spectrum of a given input sound is commonly cross-correlated with the ITD and spectrum of an entire database of sounds for which the location is known. The location with the highest correlation is then chosen as the optimal source location. In other words, matched filtering analysis can be seen as an optimal detection algorithm that, in the case of sound localization, assumes the system possesses a database of sounds, similar to the incoming sound, with which it has previously stored perfect knowledge of the source location with perfect resolution. The system then simply chooses the ''best match.''

Matched filtering analysis[6] was performed on the sounds bandpassed 7637–13264 Hz and the results are shown in Fig. 17. The matched filtering clearly localizes the sounds much better than the subject or the TDNN model as shown previously in Figs. 13(a)–(b) (a quantitative description of the performance is given in Table VIII). The matched filtering model used the same number of cochlear channels as the TDNNs and therefore contained the same inherent spectral resolution. This spectral resolution (31 cochlear channels) is certainly less than the spectral resolution of the human cochlea. This shows that although there was sufficient information to localize the sounds from the point of view of matched filtering, neither the human nor TDNN demonstrated such ability in their localization performance. This has important implications for understanding the TDNN localization responses shown in Fig. 13(b). Mathematically, a noncompetitive neural network, such as the TDNN, can learn to perform matched filtering. This can be seen from the formulas for the cross-correlation analysis and the output of a neuron in the neural network, respectively,

$$\mathrm{corr}_j = \sum_{i=1}^{n} \mathrm{input}_i \mathrm{template}_i^j,$$

$$\mathrm{neuron}_j = f\left(\sum_{i=1}^{n} \mathrm{input}_i \mathrm{weight}_i^j + \mathrm{bias}_j \right),$$

where corr_j is the cross-correlation result corresponding to the location labeled by j, input_i is the ith spectral component of the input sound, $\mathrm{template}_i^j$ is the ith spectral component of the sound template corresponding to location j, neuron_j is

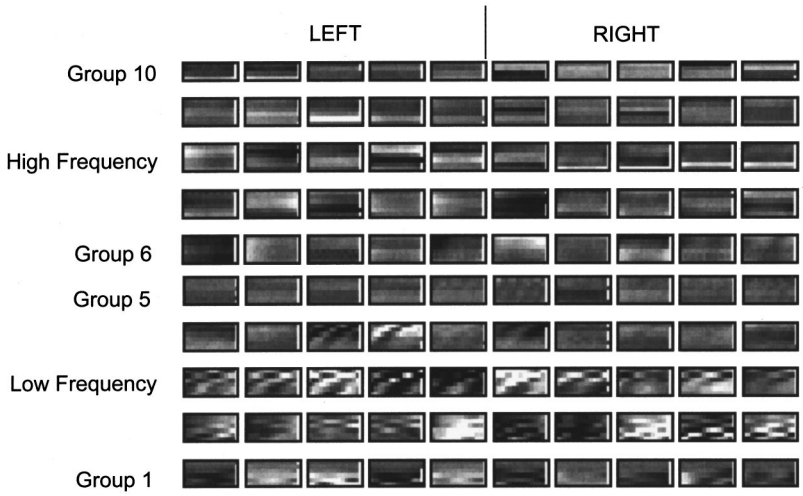


FIG. 16. The weights of the neurons in the first hidden layer are shown for the TDNN with architecture C (trained using sounds with random center-frequency and bandwidth). The weights of a given neuron are plotted in a block using gray-scale coloring. Time runs horizontally along the block and different frequency channels run vertically.

the output value of the neuron coding location j, weight_i^j is the network weight connecting the ith spectral component of the input sound with neuron_j, bias_j is the constant network bias term for neuron_j, and f is the network squashing function. In order for the TDNN to localize similarly to the matched filtering model, the network weights corresponding to a given location need to assume the form of the filter template for that location. As all of the training sounds were flat-spectrum, the TDNN received no ambiguity as far as the source spectrum was concerned. Thus it is likely that the difference in the distribution of localization responses in Fig. 13(b), as compared with that in Fig. 17, has been encouraged by using training sounds with random center frequency and bandwidth. The results suggest that training with widely varying sound stimuli makes it difficult for the TDNN to achieve results comparable with matched filtering and that this may also apply to human sound localization performance. If so, this would provide a partial explanation as to why the human localization performance is not optimal from a matched filtering standpoint.

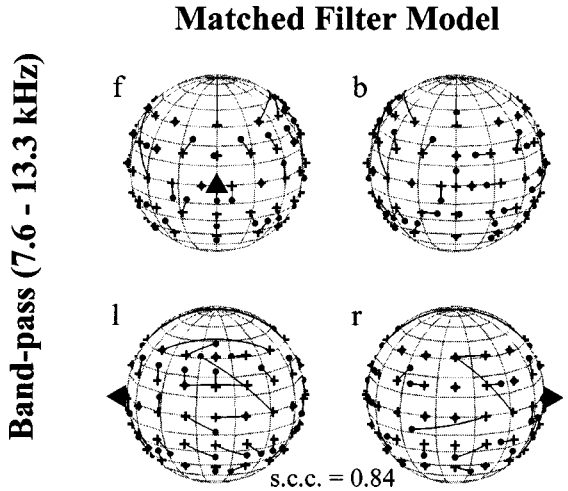


FIG. 17. The localization performance of the matched filtering algorithm is shown for the restricted high-frequency condition consisting of sounds bandpassed 7637–13264 Hz. The localization accuracy surpasses that of subject X and the TDNN shown in Fig. 13.

VIII. CONCLUSIONS

This study portrays a neural system identification model that was readily used to explore the computational process being performed by the human auditory system as it localizes sounds in space. During the development of the model, three physiological constraints were imposed upon the modeling process: (1) a TDNN model was used to incorporate the important role of spectral and temporal processing in the auditory nervous system, (2) a tonotopic structure was added to the network, (3) the training sounds contained randomly varying center frequencies and bandwidths. Not only did these constraints allow for a biologically plausible model, but more importantly they provided for an increased understanding of the role that these biological constraints play in the localization process. It was found that frequency division provided an essential ingredient for obtaining meaningful results across the different bandpass sound conditions. In particular, it allowed for the meaningful incorporation of temporal information into the low-frequency neurons of the network. In addition, it was argued that the tonotopic network structure more evenly distributed the computation over the hidden layer neurons, each of whose input was restricted in frequency. In this fashion, no particular frequency band could dominate the localization computation. As for the network training with variable sounds, it was assumed that this was a more biologically plausible listening condition than one in which the sounds were always broadband, though it makes the learning task more difficult for the network. The degradation in localization performance caused by training

TABLE VIII. Comparison of the localization performance results for the subject, the TDNN with architecture C (trained using sounds with random center frequency and bandwidth), and the matched filtering model using a sound stimulus bandpassed 7637–13264 Hz.

	Spherical correlation coefficient	Mean spherical angle error	Percentage front-back error
Subject X	−0.03	63	55
TDNN	0.03	68	72
Matched filtering	0.84	13	8

with variable sounds was compared with human performance. The similarities obtained give some indication that this variability may partially explain why human sound localization is not optimal in terms of a simple matched filtering approach. In addition, the graceful degradation in localization performance with sound level variations was demonstrated. Finally, the modeling results obtained with the tonotopic network give some indication that the localization of arbitrarily bandpassed sounds can be related to external ear acoustics. However, given the discrepancies between the model and human localization performance, it may well be profitable to examine each cue to location individually in terms of its role or weighting in the localization judgment. This information may then inform future localization models with regard to the synthesis procedure that the auditory system must be using to combine information across the different cues to produce a coherent judgment of sound location.

ACKNOWLEDGMENTS

The authors would like to thank André van Schaik for his most valuable comments and insights and Marwan Jabri for his support. This research was supported by the Dora Lush Biomedical Postgraduate Scholarship to C.J. and the Australian Research Council (ARC Grant Nos. A49530190, A49700049, A79905421). The authors would also like to thank Pierre Zakarauskas and two anonymous referees for their valuable comments on an earlier version of this manuscript.

[1]Any encoding of directions on a sphere that employs only two spatial variables must always be discontinuous because the surface of a sphere is not homeomorphic to the plane (see Guillemin and Pollack, 1974).

[2]The time-delay networks progressively integrate more temporal information in the higher layers (in this case, up to about 8 ms).

[3]It has been suggested by Zipser (1992) that empirical evidence indicates that certain similarity classes exist in neural network solutions, although this has not been mathematically proven. Thus, it may be that for problems that are sufficiently ''smooth'' and networks that do not have too many hidden units, much the same computation is being performed.

[4]Target locations with an azimuthal angle within five degrees of the interaural axis were excluded from the front-back error analysis.

[5]These particular frequencies relate to the center frequencies of the model's cochlear filters.

[6]The matched filtering analysis used templates derived from flat-spectrum noise bandpassed between 7637 and 13264 Hz, while some previous template-matching models have used templates derived from flat-spectrum, broadband noises. In other words, in this work the spectral analysis was constrained to the frequency region containing signal energy. There is no evidence suggesting that the auditory system performs spectral shape analysis in frequency regions with no signal energy. It is rather more likely that the large spectral contrast at the low-frequency band-edge may influence human sound localization performance, but will not bias the matched filtering model because of the within-band spectral assumption.

Backman, J., and Karjalainen, M. (**1993**). ''Neural network models of sound localization based on directional filtering by the pinna,'' in *Proceedings IEEE International Conference Acoustics Speech and Signal Processing* (IEEE, New York), pp. I-125–I-128.

Batteau, D. (**1967**). ''The role of the pinna in human localization,'' Proc. R. Soc. London, Ser. B **168**, 158–180.

Baum, E. B., and Haussler, D. (**1989**). ''What size net gives valid generalization,'' Neural Comput. **1**, 151–160.

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization*, revised edition (MIT Press, Cambridge, MA).

Bloom, P. (**1977**). ''Determination of monaural sensitivity changes due to the pinna by use of minimum-audible-field measurements in the lateral vertical plane,'' J. Acoust. Soc. Am. **61**(3), 820–828.

Butler, R. (**1986**). ''The bandwidth effect on monaural and binaural localization,'' Hear. Res. **21**, 67–73.

Butler, R. A., and Musicant, A. D. (**1993**). ''Binaural localization: Influence of stimulus frequency and the linkage to covert peak areas,'' Hear. Res. **67**, 220–229.

Carlile, S. (**1990**). ''The auditory periphery of the ferret II: The spectral transformations of the external ear and their implications for sound localization,'' J. Acoust. Soc. Am. **88**(5), 2196–2204.

Carlile, S. (**1996a**). ''Auditory space,'' in *Virtual Auditory Space: Generation and Applications*, edited by S. Carlile (Chapman and Hall, New York), Chap. 1, pp. 1–25.

Carlile, S. (**1996b**). *Virtual Auditory Space: Generation and Applications* (Chapman and Hall, New York).

Carlile, S., Delaney, S., and Corderoy, A. (**1999**). ''The localization of spectrally restricted sounds by human listeners,'' Hear. Res. **128**, 175–189.

Carlile, S., Leong, P., and Hyams, S. (**1997**). ''The nature and distribution of errors in the localization of sounds by humans,'' Hear. Res. **114**, 179–196.

Chau, W., and Duda, R. O. (**1995**). ''Combined monaural and binaural localization of sound sources,'' in *29th Asilomar Conference on Signals, Systems, and Computers* (IEEE, New York).

Chung, W., Carlile, S., and Leong, P. (**2000**). ''A performance adequate computational model for auditory localization,'' J. Acoust. Soc. Am. **107**(1), 432–445.

Datum, M. S., Palmieri, F., and Moiseff, A. (**1996**). ''An artificial neural network for sound localization using binaural cues,'' J. Acoust. Soc. Am. **100**, 372–383.

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 (Lawrence Erlbaum, Mahwah, NJ), Chap. 3, pp. 49–75.

Durlach, N., Rigopulos, A., Pang, X., Woods, W., Kulkarni, A., Colburn, H., and Wenzel, E. (**1992**). ''On the externalization of auditory images,'' Presence **1**, 251–257.

Fisher, I. N., Lewis, T., and Embleton, B. J. J. (**1987**). *Statistical Analysis of Spherical Data* (Cambridge University Press, Cambridge).

Gaik, W. (**1993**). ''Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling,'' J. Acoust. Soc. Am. **94**(1), 98–110.

Gardner, M. B., and Gardner, R. S. (**1973**). ''Problems of localization in the median plane: Effect of pinnae cavity occlusion,'' J. Acoust. Soc. Am. **53**, 400–408.

Giguère, C., and Woodland, P. C. (**1994**). ''A computational model of the auditory periphery for speech and hearing research. I. Ascending path,'' J. Acoust. Soc. Am. **95**, 331–342.

Guillemin, V., and Pollack, A. (**1974**). *Differential Topology* (Prentice-Hall, Englewood Cliffs, NJ).

Hartmann, W. M., and Wittenberg, A. (**1996**). ''On the externalization of sound images,'' J. Acoust. Soc. Am. **99**(6), 3678–3688.

Hofman, P. M., and Van Opstal, J. A. V. (**1998**). ''Spectro-temporal factors in two-dimensional human sound localization,'' J. Acoust. Soc. Am. **103**, 2634–2648.

Hyams, S., and Carlile, S. (**1996**). ''After-effects in auditory localization: Evidence for channel based processing,'' Proc. Aust. Neurosci. Soc. **7**, 225.

Irvine, D. R. F. (**1992**). ''Physiology of the auditory brainstem,'' in *The Mammalian Auditory Pathway: Neurophysiology*, edited by A. N. Popper and R. R. Fay (Springer-Verlag, New York), Chap. 4, pp. 153–231.

Janko, J. A., Anderson, T. R., and Gilkey, R. H. (**1997**). ''Using neural networks to evaluate the viability of monaural and interaural cues for sound localization,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Lawrence Erlbaum, Mahwah), Chap. 26, pp. 557–570.

Jeffress, L. A. (**1948**). ''A place theory of sound localization,'' J. Comp. Physiol. Psychol. **41**, 35–39.

King, A. J., and Carlile, S. (**1995**). ''Neural coding for auditory space,'' in *The Cognitive Neurosciences*, edited by M. S. Gazzaniga (MIT Press, Cambridge), Chap. 17, pp. 279–293.

King, R. B., and Oldfield, S. R. (**1997**). ''The impact of signal bandwidth on auditory localization: Implications for the design of three-dimensional audio displays,'' Human Fact. **39**, 287–295.

Kulkarni, A., and Colburn, H. S. (**1998**). ''Role of spectral detail in sound-

source localization,'' Nature (London) **396**(6713), 747–749.
Kuwada, S., Batra, R., and Fitzpatrick, D. C. (**1997**). ''Neural processing of binaural temporal cues,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Lawrence Erlbaum, Mahway), Chap. 20, pp. 399–425.
Lang, K. J., and Hinton, G. E. (**1988**). ''A time-delay neural network architecture for speech recognition,'' Technical Report, Carnegie-Mellon University, Pittsburgh, PA.
Leong, P. H. W., and Carlile, S. (**1998**). ''Methods for spherical data analysis and visualisation,'' J. Neurosci. Methods **80**, 191–200.
Leung, J., Jin, C., and Carlile, S. (**1998**). ''Generation of smoothly rendered virtual auditory space,'' Proc. Aust. Neurosci. Soc. **9**, 167.
Lindemann, W. (**1986**). ''Extension of a binaural cross-correlation model by means of contralateral inhibition, I: Simulation of lateralization of stationary signals,'' J. Acoust. Soc. Am. **80**, 1608–1622.
Lyon, R. F. (**1983**). ''A computational model of binaural localization and separation,'' in *Proc. IEEE International Conference Acoustics Speech and Signal Processing* (IEEE, New York).
Macpherson, E. A. (**1994**). ''On the role of head-related transfer function spectral notches in the judgement of sound source elevation,'' in *Auditory Display: Sonification, Audification, and Auditory Interface, SFI Studies in the Sciences of Complexity*, edited by G. Kramer and S. Smith (Addison-Wesley, Reading, MA), pp. 187–194.
Macpherson, E. A. (**1996**). ''Effects of source spectrum irregularity and uncertainty on sound localization,'' J. Acoust. Soc. Am. **99**, 2515.
Macpherson, E. A. (**1997**). ''A comparison of spectral correlation and local feature-matching models of pinna cue processing,'' J. Acoust. Soc. Am. **101**, 3105.
Marr, D. (**1982**). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (Freeman, San Francisco, CA).
May, B. J., and Huang, A. Y. (**1997**). ''Spectral cues for sound localization in cats: A model for discharge rate representations in the auditory nerve,'' J. Acoust. Soc. Am. **101**(5), 2705–2719.
Mehrgardt, S., and Mellert, V. (**1977**). ''Transformation characteristics of the external human ear,'' J. Acoust. Soc. Am. **61**(6), 1567–1576.
Middlebrooks, J. (**1992**). ''Narrow-band sound localization related to external ear acoustics,'' J. Acoust. Soc. Am. **92**(5), 2607–2624.
Middlebrooks, J., and Green, D. (**1990**). ''Directional dependence of interaural envelope delays,'' J. Acoust. Soc. Am. **87**(5), 2149–2162.
Middlebrooks, J., and Green, D. (**1991**). ''Sound localization by human listeners,'' Annu. Rev. Psychol. **42**, 135–159.
Middlebrooks, J., and Green, D. (**1992**). ''Observations on a principal components analysis of head-related transfer functions,'' J. Acoust. Soc. Am. **92**(1), 597–599.
Middlebrooks, J., and Knudsen, E. (**1984**). ''A neural code for auditory space in the cat's superior colliculus,'' J. Neurosci. **4**(10), 2621–2634.
Møller, H. (**1992**). ''Fundamentals of binaural technology,'' Appl. Acoust. **36**, 171–218.
Møller, H., Sørensen, M. F., Hammershøi, D., and Jensen, C. B. (**1995**). ''Head-related transfer functions of human subjects,'' J. Audio Eng. Soc. **43**(5), 300–321.
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.
Nandy, D., Rao, K. R., and Ben-Arie, J. (**1993**). ''Multiple auditory template matching using expansion,'' in *Proceedings of the 36th Midwest Symposium on Circuits and Systems* (IEEE, New York), pp. 899–902.
Neti, C., Young, E., and Schneider, M. (**1992**). ''Neural network models of sound localization based on directional filtering by the pinna,'' J. Acoust. Soc. Am. **92**(6), 3140–3156.
Oldfield, S. R., and Parker, S. P. A. (**1984**). ''Acuity of sound localization: A topography of auditory space II: Pinna cues absent,'' Percep. **13**, 601–617.
Patterson, R. D., and Allerhand, M. H. (**1995**). ''Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform,'' J. Acoust. Soc. Am. **98**, 1890–1894.
Pralong, D., and Carlile, S. (**1994**). ''Measuring the human head-related transfer functions: A novel method for the construction and calibration of a miniature 'in-ear' recording system,'' J. Acoust. Soc. Am. **95**(6), 3435–3444.
Pralong, D., and Carlile, S. (**1996**). ''Generation and validation of virtual auditory space,'' in *Virtual Auditory Space: Generation and Applications*, edited by S. Carlile (Chapman and Hall, New York), Chap. 4, pp. 109–151.
Rayleigh, L. J. (**1907**). ''On our perception of sound direction,'' Philos. Mag. **13**, 214–232.
Roffler, S. K., and Butler, R. A. (**1968**). ''Factors that influence the localization of sound in the vertical plane,'' J. Acoust. Soc. Am. **43**(6), 1255–1259.
Rogers, M. E., and Butler, R. A. (**1992**). ''The linkage between stimulus frequency and covert peak areas as it relates to monaural localization,'' Percept. Psychophys. **52**(5), 526–546.
Rosen, D., Rumelhart, D., and Knudsen, E. (**1994**). ''A connectionist model of the owl's localization system,'' in *Advances in Neural Information Processing Systems VI*, edited by J. D. Cowan, G. Tesauro, and J. Alspector (Morgan Kaufmann, San Mateo, CA), pp. 606–613.
Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (**1986**). ''Learning internal representations by error backpropagation,'' in *Parallel Distributed Processing*, edited by D. E. Rumelhart and J. L. McClelland (MIT Press, Cambridge, MA), Vol. 1, Chap. 8, pp. 318–362.
van Schaik, F. A. (**1998**). Personal communication.
Schenkel, M. (**1995**). ''Handwriting recognition using neural networks and hidden Markov models,'' Ph.D. thesis, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland.
Searle, C., Braida, L., Cuddy, D., and Davis, M. (**1975**). ''Binaural pinna disparity: Another auditory localization cue,'' J. Acoust. Soc. Am. **57**(2), 448–455.
Searle, C., Braida, L., Davis, M., and Colburn, H. S. (**1976**). ''Model for auditory localization,'' J. Acoust. Soc. Am. **60**(5), 1164–1175.
Shamma, S., Cohen, A., Horiuchi, T., Indiveri, G., Douglas, R., Koch, C., and Sejnowski, T. (**1998**). ''Report to the national science foundation: Workshop on neuromorphic engineering Telluride, Colorado,'' World Wide Web address: http://www.ini.unizh.ch:80/telluride98/report98/.
Shinn-Cunningham, B. G. (**1998**). Personal communication.
Thurlow, W., and Mergener, J. R. (**1970**). ''Effect of stimulus duration on localization of direction of noise stimuli,'' J. Speech Hear. Res. **13**, 826–838.
Waibel, A., Hanazawa, T., Hinton, G. E., Shikano, K., and Lang, K. (**1989**). ''Phoneme recognition using time-delay neural networks,'' IEEE Trans. Acoust., Speech, Signal Process. **37**, 328–339.
Wallach, H. (**1939**). ''On sound localization,'' J. Acoust. Soc. Am. **10**, 270–274.
Wan, E. A. (**1993**). ''Neural network with embedded tapped-delay lines,'' in *Predicting the Future and Understanding the Past, SFI Studies in the Sciences of Complexity*, edited by A. Weingard and N. Gershenfeld (Addison-Wesley, Reading, MA), Vol. XVII.
Wenzel, E., Wightman, F., and Kistler, D. (**1988**). ''Acoustic origins of individual differences in sound localization behavior,'' J. Acoust. Soc. Am. Suppl. 1 **84**, S79.
Wightman, F., and Kistler, D. (**1993**). ''Sound localization,'' in *Human Psychophysics*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 155–192.
Wightman, F. L., and Kistler, D. J. (**1989a**). ''Headphone simulation of free field listening I: Stimulus synthesis,'' J. Acoust. Soc. Am. **85**(2), 858–867.
Wightman, F. L., and Kistler, D. J. (**1989b**). ''Headphone simulation of free field listening II: Psychophysical validation,'' J. Acoust. Soc. Am. **85**(2), 868–878.
Wightman, F. L., and Kistler, D. J. (**1997**). ''Monaural sound localization revisited,'' J. Acoust. Soc. Am. **101**, 1050–1063.
Young, E., Spirou, G., Rice, J., and Voigt, H. (**1992**). ''Neural organization and responses to complex stimuli in the dorsal cochlear nucleus,'' Philos. Trans. R. Soc. London, Ser. B **336**(1278), 407–413.
Zahorik, P., Wightman, F. L., and Kistler, D. J. (**1996**). ''The fidelity of virtual auditory displays,'' J. Acoust. Soc. Am. **99**, 2596.
Zakarauskas, P., and Cynader, M. S. (**1993**). ''A computational theory of spectral cue localization,'' J. Acoust. Soc. Am. **94**(3), 1323–1331.
Zhou, B., Green, D., and Middlebrooks, J. (**1992**). ''Characterization of external ear impulse responses using golay codes,'' J. Acoust. Soc. Am. **92** Pt. 1, 1169–1171.
Zipser, D. (**1992**). ''Identification models of the nervous system,'' Neuroscience (N.Y.) **47**, 853–862.

Complex tactile waveform discrimination

Sliman J. Bensmaïa and Mark Hollins
Department of Psychology, University of North Carolina at Chapel Hill, CB# 3270 Davie Hall, Chapel Hill, North Carolina 27599-3270

(Received 22 July 1999; accepted for publication 16 June 2000)

Complex vibrotactile waveforms consisting of two superimposed sinusoids at varying phases were presented to the fingertip, and observers made ''same-different'' judgments. It was found that the low-frequency (10 Hz+30 Hz) waveforms were discriminable from one another while discrimination of the high-frequency (100 Hz+300 Hz) vibrations was poor. High-frequency adaptation did not impair discrimination of the low-frequency waveforms, suggesting that the RA channel mediated discrimination. Low-frequency adaptation impaired discrimination of the high-frequency stimuli, suggesting that the RA channel likewise mediated the modest level of performance observed in the absence of an adapting stimulus. The results indicate that this channel encodes complex waveforms temporally. A simple model for low-frequency waveform discrimination is proposed. The results obtained with the high-frequency complex waveforms are compatible with the hypothesis that the PC channel integrates stimulus energy over time. © *2000 Acoustical Society of America.* [S0001-4966(00)04309-5]

PACS numbers: 43.66.Wv, 43.64.Vm [RVS]

INTRODUCTION

We have recently found evidence that the vibrotactile system may play an important role in the perception and discrimination of textures (Hollins *et al.*, 1998; Hollins and Risner, 2000). We propose that, as the skin moves across textured surfaces (or vice versa), vibrations produced in the skin are transduced by the vibrotactile channels, and that these vibrations relay textural information. Vibration may play a particularly important role in the perception of textural features that are too fine for the spatial resolution of the slowly adapting type I (SA I) mechanoreceptors, which have been shown to mediate the perception of relatively coarse surfaces (Connor *et al.*, 1990; Connor and Johnson, 1992). In addition, we have recorded vibrations produced in the skin during tactile stimulation by various moving surfaces (Hollins *et al.*, 1996). These vibrations were complex in nature. Therefore, ascertaining the ability of each channel to mediate the discrimination of complex waveforms may shed light on the respective roles of the vibrotactile channels in texture perception and in somesthesis as a whole.

The rapidly adapting (RA) channel is particularly well suited for encoding stimulus waveform as has been argued elsewhere (Hollins *et al.*, 1996). First, frequency discrimination is apparently based on temporal coding at low frequencies, the response to which is mediated primarily by the RA system (Talbot *et al.*, 1968; Bolanowski *et al.*, 1988). Second, some evidence suggests that the RA system uses as its main index of magnitude an encoded representation of peak amplitude of the stimulus waveform (Hollins *et al.*, 1996). Therefore, the RA system seems to preserve both the temporal structure and amplitude of a stimulus. These are desirable properties for a system charged with faithfully encoding stimulus waveform. One would therefore expect the RA channel to be a good mediator of complex waveform discrimination. One might further predict that discriminability of RA-mediated low-frequency waveforms will increase as a function of the difference in their peak amplitudes since amplitude seems to be a critical parameter encoded by the RA channel.

The Pacinian (PC) channel tells a different story. Several researchers have proposed that the PC channel functions as a linear integrator of stimulus energy, similarly to a critical band filter in audition (Verrillo and Gescheider, 1975; Marks, 1979; Gescheider *et al.*, 1994; Makous *et al.*, 1995). It has been reliably demonstrated that, within limits, PC thresholds decrease at a rate of approximately 3.0 dB per doubling of stimulus area or stimulus duration, suggesting that this channel integrates stimulus energy over space and time (Verrillo, 1968; Gescheider, 1976; Gescheider *et al.*, 1994). Using a magnitude estimation paradigm, Marks (1979) found that perceived intensity of superimposed sinusoids within the sensitivity spectrum of the PC channel (200 and 250 Hz) was a power function of stimulus energy rather than the summed loudness of the two frequency components. Furthermore, Makous *et al.* (1995), using a simultaneous masking paradigm, showed that the masker-induced threshold shift within the PC channel was determined by the ratio between the signal power and the noise (masker) power passing through the channel, as predicted by the critical band filter hypothesis. Based on the hypothesis that the PC channel integrates stimulus energy over time, one would predict that stimuli activating the PC channel would be discriminable to the extent that they differed in power.

Results obtained by Horch (1991) suggest a different coding mechanism for the PC channel. He presented participants with pairs of superimposed sinusoids differing only in the phase relationship between the high- and the low-frequency components, and instructed them to identify which member of each pair was complex. In this condition, they both were. The result of interest is that some pairs of complex stimuli were reliably discriminable. Horch proposed that these stimuli were discriminable on account of differ-

ences in peak acceleration, and that peak acceleration was the critical stimulus for activating Pacinian corpuscles. According to this hypothesis, then, one would expect that high-frequency waveforms should be discriminable to the extent that their peak accelerations differ.

The objective of the present study was to investigate the ability of the vibrotactile system to discriminate among complex mechanical waveforms, consisting of two superimposed sinusoids differing only in the phase of the high-frequency component with respect to the low-frequency component. In experiment 1, we examined (in separate sessions) the discriminability of waveforms within different parts of the frequency spectrum, in order to ascertain the ability of RA and PC channels to encode waveform. In experiment 2, we explored the possibility that the lack of discrimination obtained in the high-frequency condition of experiment 1 was specific to our initial choice of parameters. In experiment 3, we attempted to isolate the contribution of each channel to complex waveform discrimination using an adaptation paradigm.

EXPERIMENT 1

The first experiment was designed to ascertain the ability of the vibrotactile channels to mediate the discrimination of complex mechanical waveforms, each consisting of a pair of superimposed sinusoids. The stimuli fell into two frequency regions that spanned the sensitivity spectra of the two vibrotactile channels. The low-frequency pair, 10 Hz+30 Hz, was selected to activate preferentially the RA channel; the high-frequency pair, 100 Hz+300 Hz, was selected to activate preferentially the PC channel. The amplitudes of the components of the complex stimuli were set to be equal in sensation level (decibels above absolute threshold). We scaled amplitudes in this way in order to make the individual components' contributions to subjective intensity roughly comparable. High-amplitude (30 dB SL) stimuli were employed to ensure that the waveforms were transduced adequately by the mechanoreceptors.

Materials and methods

Apparatus

The waveforms were generated by a Keithley Metrabyte arbitrary waveform generator (PCIP-AWFG), controlled by a Dell Optiplex GXa computer. The waveforms were fed to an electromechanical transducer (model CS-510 tactile stimulator, Cantek Metratron Corp., Canonsburg, PA), which delivered the stimuli to the skin via a vibrating probe. The probe, made of Delrin polyacetal thermoplastic, was 5 mm in diameter and was surrounded by an annulus of the same material to minimize the spread of vibration along the surface of the skin. A 1-mm gap separated the probe contactor from the surround. Feedback from the vibrator, generated by a linear-variable differential transformer (LVDT), was sent back to the computer via an analog-to-digital converter, and translated into position. A digital thermometer (AT 42, Autogenic Systems, Inc., Wood Dale, IL), connected to the computer, was used to measure the participant's temperature during the experiment.

Participants

The participants were five graduate students, two females and three males. One was the first author (S.J.B); the other four were paid for their participation. They ranged in age from 22 to 24 and all were right-handed. The same participants took part in all three of the experiments. One of the participants consistently performed below chance-levels on all the tasks and complained that the waveforms felt different on all trials. Her data were therefore excluded from the analyses.

Threshold measurements

The absolute threshold of each participant was measured at each of the relevant frequencies, namely 10, 30, 100, and 300 Hz, using a two-interval forced-choice tracking protocol. Each run consisted of 50 trials. On a given trial, the participant was to select in which of two intervals the stimulus was presented. If he or she selected incorrectly, the amplitude of the stimulus was incremented; after three correct responses, not necessarily consecutive, the amplitude of the stimulus was decremented. The increment/decrement was 1 dB for the first 20 trials and .5 dB during the final 30 trials. The threshold estimate derived from a given run was the geometric mean of the amplitudes (i.e., the antilog of the arithmetic mean of dB values) of the stimuli presented on the first 25 consecutive trials during which the participant remained within a 2-dB window. If the participant failed to remain within a 2-dB window for 25 consecutive trials in a given run, the run was discarded. The overall threshold estimate for each frequency was obtained by computing the geometric mean of five such threshold estimates. Each participant was allowed two practice runs per frequency; the data for these runs were not included in the final threshold calculations.

Complex stimuli

The complex stimuli were generated by summing two sinusoids of unequal frequency. A frequency ratio of 3:1 for the two component sinusoids was utilized because it can yield complex waves that differ substantially in shape when the phase of the high-frequency component is varied.

The phase difference between the two components was varied for each frequency pair; the four phase differences used were 0, $\pi/2$, π, and $3\pi/2$ radians (or 0, 90, 180, 270 degrees). The formula for each complex wave was then

$$x_t = A_1 \sin(2\pi f_1 t) + A_2 \sin(2\pi f_2 t + \phi),$$

where A_1 and A_2 are the amplitudes of the low-frequency and high-frequency components, respectively; f_1 and f_2 are the frequencies of the low-frequency and high-frequency components, respectively ($f_2 = 3f_1$); ϕ is the phase (in radians) of the high-frequency component with respect to the low-frequency component; and t is a time increment. The amplitudes of the sinusoidal components of the complex waves were set at 30 dB SL, i.e., 30 dB above each subject's absolute threshold at that frequency.

The vibrotactile stimulator used in these experiments had previously been used mainly to deliver simple sinusoidal vibrations. When complex waves were fed into the stimula-

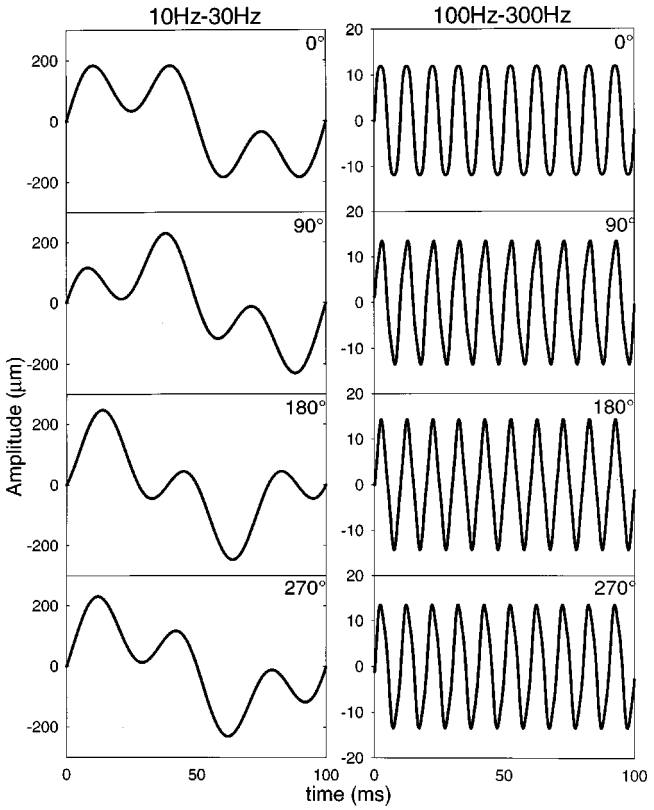


FIG. 1. 100-ms sample of the waveforms (10+30 Hz and 100+300 Hz) presented in experiment 1. The waveforms presented to S.J.B. are shown.

tor, it yielded outputs that deviated from the desired complex waveforms. The deviations were substantial for the high-frequency waves. To ensure an adequate stimulus set, it was necessary to fashion an analog input to the stimulator that yielded the desired mechanical complex wave from the stimulator. We wrote a computer program that adjusted the analog input to yield an adequate output. The program first computed the difference, point by point, between one cycle of the actual output given an initial analog input, and one cycle of the desired output. The program proceeded to scale the amplitude of the difference array by multiplying each of its components by the ratio of the peak amplitude of the analog input to the peak amplitude of the actual output; the resulting array was then added to the input signal array. The transformed input array was then fed back into the stimulator, and the process began anew. This process iterated until the output no longer improved. Finally, the input signal was rearranged so that the waveform was initiated at a zero-crossing with a positive slope. If several such points existed in a given waveform, the zero-crossing just preceding the largest positive displacement was selected. This selection process was effected in order to minimize the likelihood that cues at stimulus onset—such as onset transients—might facilitate the discrimination judgments. Figure 1 shows two sets of typical waveforms, corresponding to the two frequency conditions. Once an input signal was developed that yielded a satisfactory output, it was stored in a file; this file was called up by the computerized experimental protocol every time the associated mechanical waveform was to be delivered by the vibrotactile stimulator.

Procedure

The participant sat in a reclining dentist's chair and extended his or her left forearm and hand, palm facing upward, on an evacuation bag resting on the experimental table. The chair was adjusted until the participant was comfortable. Next, the thermometer probe was taped to the proximal phalanx of the participant's index finger. The distal index fingerpad was then placed under the surround ring, itself suspended above the table by aluminum scaffolding. Next, the contactor of the vibrotactile stimulator was lowered until it just touched the skin; initial contact was determined by monitoring the force display of the stimulator: an increase of .2 g or more in the force exerted on the contactor indicated that it had just touched the skin. The contactor was then lowered another 500 μm to ensure that the skin maintained contact with the contactor throughout vibratory stimulation.

A calibration routine was then initiated, in which the analog inputs to the stimulator (stored in files as described earlier) were adjusted so that the mechanical waveforms they yielded were of the desired amplitudes. Calibration was achieved by feeding the input into the vibrotactile stimulator, computing the peak-to-peak amplitude of the output, and then multiplying the input signal by the ratio of the desired output to the actual output. This calibration was necessary in order to compensate for small run-to-run variations in the force exerted on the contactor by the participant's finger, which would have otherwise had an impact on the amplitude of the mechanical vibration. The calibration process iterated until the amplitude of the vibration was within 1% of the desired amplitude for each of the four complex stimuli to be used in that run. Once calibration was complete, circumaural earphones were placed on the participant; low-pass filtered noise ($<$400 Hz) was presented through the earphones in order to eliminate auditory cues.

Each trial began with a warning light that remained on for 2 sec. The light was followed by a 1-sec interval to allow the participant to ready him- or herself for the stimulus. The two 1-sec test stimuli were then presented in succession, separated by a 1-sec interstimulus interval. Separate lights signaled the presentation of the two stimuli. Immediately after the second stimulus, a fourth light prompted the participant to enter his or her response. The participant was to press the left button on a response box if he or she thought the same vibration was presented in both intervals, or the right button if he or she felt different vibrations had been presented in the two stimulus intervals. Subsequent to the participant's response, feedback lights indicated whether it was correct or incorrect. The response was followed by a 13-sec intertrial period, without light or stimulus, which ended when the warning light came back on to indicate the beginning of the following trial. The 16-sec interval between the participant's response and the onset of the first stimulus was deemed necessary in order to avoid adaptation effects (cf. Goble and Hollins, 1993).

In a given run, four different stimuli were presented in random pairs. There were ten possible pairs: 0°–0°, 0°–90°, 0°–180°, 0°–270°, 90°–90°, 90°–180°, 90°–270°, 180°–180°, 180°–270°, 270°–270°. The only parameter that varied from stimulus to stimulus within a run was the phase of the

high-frequency component with respect to the low-frequency component. The stimulus pair presented on each trial was selected at random.

Each run consisted of 55 trials and lasted approximately 18.5 min. During the course of a run, skin temperature was maintained, to the extent possible, between 30 and 33 °C: If the participant's temperature dipped below 30 °C, a heating pad was placed on his or her arm; if the participant's temperature exceeded 33 °C, an air-conditioning unit was switched on.

Each session lasted approximately an hour and a half and consisted of three runs with a 15-min break between runs. Each participant was allowed three practice runs per condition.

Data analysis

We collected discrimination data until there were at least 100 observations for each of the ten phase-pairs (0°–0°, 0°–90°, 0°–180°, 0°–270°, 90°–90°, etc) in each frequency condition for each participant. Using a chi-squared test, we first ascertained whether the false alarm rates corresponding to the four ''same'' pairs (0°–0°, 90°–90°, 180°–180°, 270°–270°) within each frequency condition for each participant were statistically equivalent. The false alarm rates were the rates at which the participants responded ''different'' when in fact the phase parameter of both stimuli on a given trial was the same. Among the 20 comparisons tested in experiments 1, 2, and 3—each corresponding to one subject in one condition—(8 in experiment 1, 4 in experiment 2, and 8 in experiment 3), only in one case was the null hypothesis rejected with $p<0.5$. Since this is the expected rate of rejections given the α-level chosen, the false alarm rates were pooled within each frequency condition for each participant in all three experiments.

Of the two d' measures associated with the ''same-different'' paradigm, the one assuming a differencing rule (Sorkin, 1962) is generally preferred over the one assuming independent observations (Noreen, 1981) when more than two stimuli are being compared in a given run (Macmillan and Creelman, 1991; Dai *et al.*, 1996). Using the pooled false-alarm rates, a d' for each ''different'' pair (i.e., each pair in which the two waveforms differed in phase) was determined using the ''differencing rule'' tables in the appendix of Macmillan and Creelman (1991).

Results and discussion

Low-frequency condition (10+30 Hz)

Figure 2 shows the mean d's obtained in the low-frequency condition. The d's are generally high, indicating good discriminability of the low-frequency waveforms. One notable feature of the data collected in the low-frequency condition (10+30 Hz) was the substantial agreement among the participants as to which pairs are the most/least discriminable: 90°–270° was the most discriminable and 180°–270° the least discriminable pair in all cases. This is surprising in view of the fact that the 90°–270° pair is the only pair in which both waveforms have equal peak-to-peak amplitudes, and are actually mirror images of one another; one might

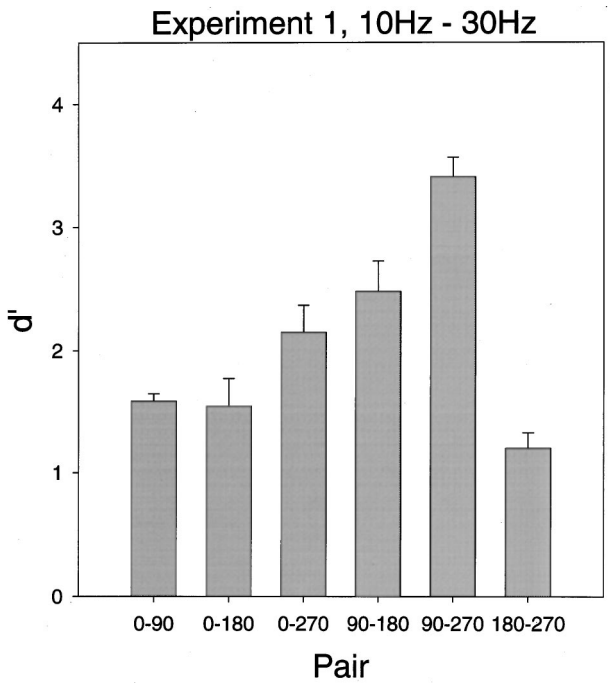


FIG. 2. Mean d' values for experiment 1, low-frequency condition (10 +30 Hz).

expect that these two waveforms would be the least discriminable. Also, the fact that the 0°–180° pair was among the least discriminable is unexpected because it was the pair in which the difference in amplitude between the two waveforms was greatest. All the waveforms were at least moderately discriminable from one another except perhaps for the 180°–270° pair. It seems, therefore, that the vibrotactile system has the capability to discriminate reliably among low-frequency stimuli such as those presented in this condition.

High-frequency condition (100+300 Hz)

The d's derived from the high-frequency (100 +300 Hz) data exhibit a different result than that observed in the low-frequency condition (Fig. 3): they are all low, and there is little agreement as to which waveforms are the most or least discriminable.

The fact that the participants were virtually unable to discriminate among the waveforms is compatible with the hypothesis that the Pacinian channel integrates stimulus energy and functions like a critical band in audition (Verrillo and Gescheider, 1975; Marks, 1979; Makous *et al.*, 1995). According to the critical band hypothesis, the waveforms in the high-frequency condition should be indiscriminable because they all have the same power.

EXPERIMENT 2

It was found that the high-frequency stimuli in experiment 1 were hardly distinguishable from one another. As mentioned above, this lack of discriminability could be attributed to the fact that the Pacinian system does not encode the form of the vibration but rather integrates its energy. An alternative explanation is that the waveforms in the high-frequency condition were simply more similar and, conse-

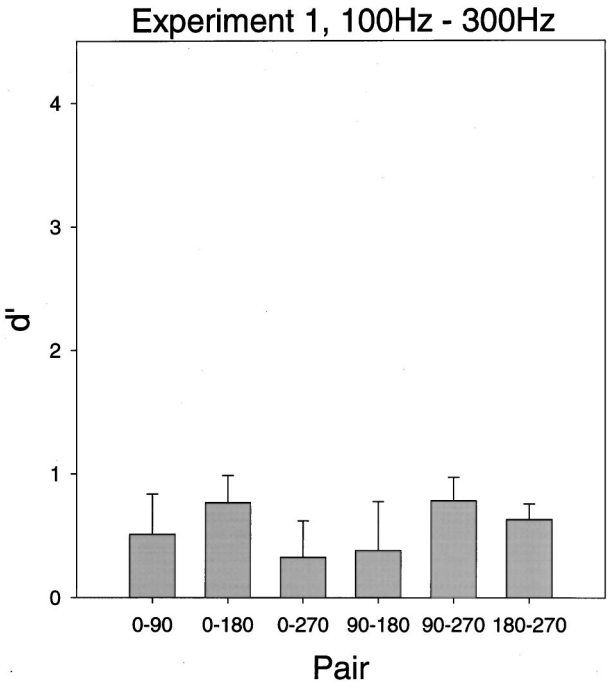


FIG. 3. Mean d' values for experiment 1, high-frequency condition (100 + 300 Hz).

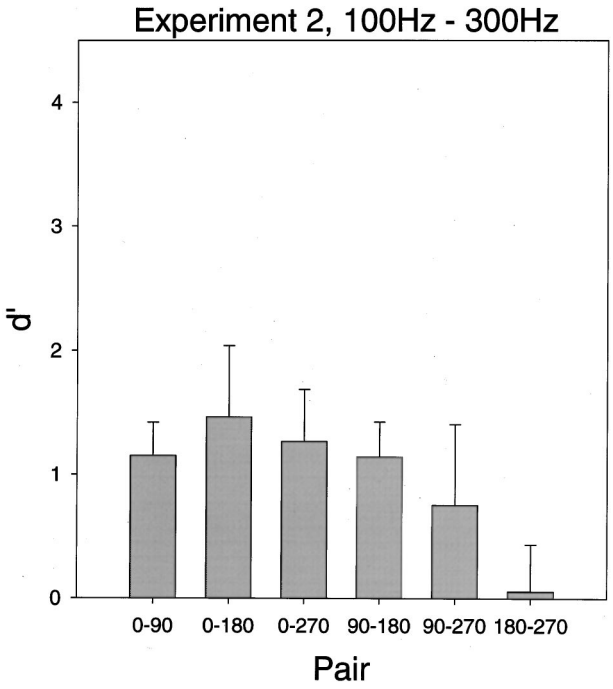


FIG. 4. Mean d' values for experiment 2, high-frequency condition (100 + 300 Hz).

quently, less discriminable. The amplitudes of the sinusoidal components of the waveforms in experiment 1 were scaled in decibels above detection threshold. Since the threshold-frequency curve in the low-frequency range is flat relative to that in the high-frequency range, the amplitudes of the two sinusoidal components in the low-frequency condition were more comparable than those of the two sinusoidal components in the high-frequency condition. In fact, the 300-Hz components of the high-frequency waveforms were almost one order of magnitude lower in amplitude than those of the corresponding 100-Hz components, and caused only slight differences in the waveforms (Fig. 1). The contribution of the 300-Hz component to the shape of these waveforms may therefore have been negligible, leaving the high-frequency waveforms virtually indiscriminable.

In experiment 2, the shapes of the high-frequency waveforms were equated with those of the low-frequency waveforms of experiment 1 to compare the abilities of the RA and PC systems to discriminate among waveforms identical in shape. In order to equate the shapes of the waveforms, the amplitude of the 300-Hz component of the high-frequency waves was increased such that the amplitude ratio of the two components was equal to that of the low-frequency waves in experiment 1. The resulting waves were then identical in shape to the low-frequency waves in experiment 1.

Materials and methods

Stimuli

The stimuli in experiment 2 were produced according to the same procedure as that applied in experiment 1. Only one frequency pair was presented, namely 100+ 300 Hz. The amplitude of the low-frequency component was 30 dB SL, as in experiment 1. The amplitude of the high-frequency component was such that the amplitude ratio of the two sinusoidal components was equal to the amplitude ratio of the 10 + 30 Hz stimuli in experiment 1. This ratio differed from one subject to another because of their different threshold functions; it ranged from 1.2 to 1.53. Thus the shapes of the waveforms in experiment 2 were identical to the shapes of the 10+ 30 Hz stimuli in experiment 1; the former differed from the latter only in frequency and peak-to-peak amplitude.

Procedure

The procedure was the same as that in experiment 1.

Results and discussion

Figure 4 shows the mean d's as a function of waveform pair. As in the high-frequency condition in experiment 1, there is considerable variability in d's across participants. The means range from 0.75 to 1.46 for five of the six waveform pairs, indicating a modicum of discriminability. These data raise the possibility that the PC channel is capable, to a certain extent, of mediating discrimination among waveforms. An alternative interpretation is that the modest amount of discrimination that was obtained in this task was mediated by the RA channel, since the amplitudes of the stimuli, higher in experiment 2 than in experiment 1, may have exceeded RA threshold. It is difficult to ascertain whether the stimuli in experiment 2 did exceed RA threshold because channel thresholds for complex stimuli are unknown.

EXPERIMENT 3

In order to further elucidate the roles of the RA and PC channels in complex waveform discrimination, it was essential to minimize the involvement of one channel in the dis-

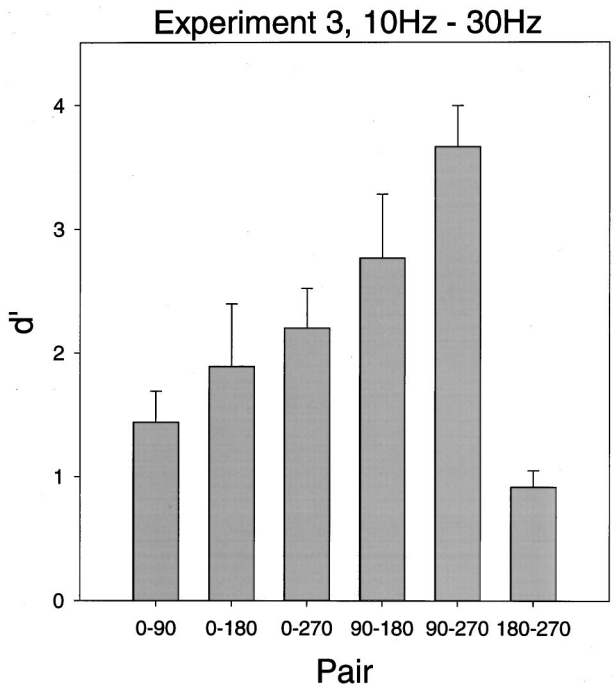


FIG. 5. Mean d' values for experiment 3, low-frequency condition (10 +30 Hz) with a 30-dB SL, 300-Hz adapting stimulus presented for 6 min initially, then for 16 sec in between trials.

crimination task in order to isolate, at least partially, the contribution of the other. It is a well-demonstrated fact that extended vibratory stimulation (adaptation) at a given frequency results in a selective decrease in the absolute sensitivity of the channel that mediates the perception of that vibratory stimulus (Gescheider and Wright, 1968, 1969; Verrillo and Gescheider, 1977; Gescheider and Verrillo, 1979; Hollins *et al.*, 1990, 1991). The objective of experiment 3 was to determine to what extent selective adaptation of the RA channel and the PC channel would affect discrimination.

Materials and methods

Adapted threshold measurements

Pilot testing was conducted in order to establish the parameters of adapting stimuli that would substantially increase detection thresholds (decrease absolute sensitivity) of each vibrotactile channel. Thresholds were measured with the two-interval forced-choice tracking protocol described earlier: Each run was preceded by 6 min of exposure to the adapting stimulus which was also presented during intertrial periods to maintain adaptation. There was a 1-sec interval without vibration between the offset of the adapting stimulus and the onset of the first stimulus interval on each trial.

The amount of adaptation of a channel has been shown to be proportional to the intensity (in dB) of the adapting stimulus relative to absolute threshold of the channel at the adapting stimulus frequency (Verrillo and Gescheider, 1977; Goble and Hollins, 1993). It was therefore essential to choose adapting frequencies at which the difference between RA and PC thresholds was maximal in order to selectively adapt one channel while minimally affecting the other. From among the frequencies for which absolute threshold had been measured in experiment 1, 10 and 300 Hz were chosen as the adapting frequencies because they are the two frequencies at which the difference between RA and PC threshold is greatest according to the threshold-frequency functions in Bolanowski *et al.* (1988).

The criterion amount of adaptation was a threshold shift of 10 dB with respect to unadapted threshold. The amplitude of the adapting stimulus at each frequency was adjusted until one was found that yielded at least a 10-dB shift in detection threshold at that same frequency in all of the participants. Thus the detection threshold of a 10-Hz test stimulus following adaptation using a 10-Hz adapting stimulus was measured, as was the detection threshold of a 300-Hz test stimulus following adaptation by a 300-Hz adapting stimulus. It was found that an adapting amplitude of 30 dB SL was sufficient for both frequencies.

Waveform discrimination

The waveform discrimination procedure was essentially the same as that in experiments 1 and 2, with the addition of the adapting stimulus. This stimulus was presented for 6 min at the outset of each run and for 16 sec between trials, with a 1-sec interval between the offset of the adapting stimulus and the onset of the first test stimulus. The test stimuli consisted of the low-frequency (10+30 Hz) complex waves from experiment 1, and the high-frequency (100+300 Hz) stimuli from experiment 2.

On runs with the low-frequency test stimuli, a 300-Hz, 30 dB SL adapter was used; when discrimination of the high-frequency test stimuli was measured, a 10-Hz, 30 dB SL adapter was presented.

Results and discussion

Low-frequency waveform discrimination with 300-Hz adaptation

The distribution of mean d's (Fig. 5) in the adaptation condition look very similar to those obtained in experiment 1, in which no adapting stimulus was applied. There was a slight overall improvement in performance with the high-frequency adapter, compared to performance without adaptation. This finding supports the interpretation that discrimination of these low-frequency stimuli is mediated by the RA system, which is largely insensitive to the 300-Hz adapting stimulus. One can conclude that adaptation of the Pacinian system did not impair, and may have slightly enhanced, discrimination of the low-frequency stimuli. This improvement in performance could not be attributed to a practice effect because the participants were allowed several practice sessions before archival data for each condition were collected, and each condition was run over the span of several months during which the participants showed no consistent improvement.

High-frequency waveform discrimination with 10-Hz adaptation

The performance of all the participants was considerably poorer in this condition (Fig. 6) than in experiment 2, in

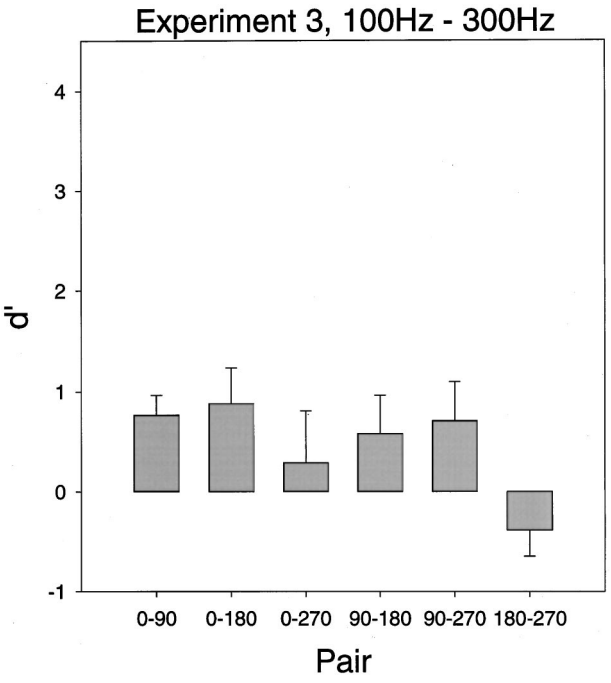


FIG. 6. Mean d' values for experiment 3, high-frequency condition (100 +300 Hz) with a 30-dB SL, 10-Hz adapting stimulus presented for 6 min initially, then for 16 sec in between trials.

which the same stimuli were presented without an adapting stimulus. This result implies that the PC system does not support effective waveform discrimination.

Adaptation of the PC channel resulted in little change in the discriminability of the low-frequency waveforms while adaptation of the RA channel resulted in a marked decrease in the discriminability of the high-frequency waveforms for all participants. It seems, therefore, that waveform discrimination in the low-frequency conditions and the modest amount of discrimination obtained in the high-frequency conditions were mediated by the RA system.

GENERAL DISCUSSION

The PC channel

We found in experiment 3 that, after the application of a 10-Hz, 30-dB SL adapting stimulus, the discriminability of high-frequency stimuli was virtually nil following adaptation of the RA system. This finding is compatible with the hypothesis that the PC channel integrates stimulus energy over time and functions as a critical band filter (Verrillo and Gescheider, 1975; Marks, 1979; Makous *et al.*, 1995): Since all the stimuli within each condition were of equal power, one would predict from this hypothesis that they would be indiscriminable if perceived only by way of the PC channel.

Horch (1991) found that certain complex waveforms could be reliably discriminated from one another and hypothesized that this discriminability was due to differences in peak acceleration. To test the applicability of this proposal to our own data, we computed the difference in peak acceleration for each pair of high-frequency waveforms from experiment 3 and standardized the resulting differences within participant. Then, we computed the correlation between the resulting standardized differences in peak acceleration for the waveform pairs and their associated d's (from the high-frequency condition of experiment 3), also standardized within participant; the Spearman correlation was small and nonsignificant (Spearman's $\rho = 0.15$, $P > 0.4$). Our results, therefore, cast doubt on the notion that differences in peak acceleration play a major role in high-frequency waveform discrimination. An alternative explanation of Horch's data stems from the fact that the waveforms that were found to be discriminable differed in power (by a factor of 1.7) while those that were indiscriminable were of equal power. Horch's findings, then, are compatible with the critical band hypothesis (Verrillo and Gescheider, 1975; Marks, 1979; Makous *et al.*, 1995).

The RA channel

After adaptation of the PC channel with a 300-Hz adapting stimulus, we observed a small improvement in discrimination of the low-frequency stimuli. The significance of this improvement is not clear. Perhaps the reduction in the activity of the PC system made for a clearer overall signal, mediated primarily by the RA channel, which in turn facilitated discrimination. In any case, it seems reasonable to conclude that the RA channel mediated discrimination of the low-frequency waveforms in experiments 1 and 3.

However, the pattern of discriminability across low-frequency waveform pairs does not correspond to our initial expectations. First, difference in peak amplitude is not a good predictor of discriminability, a surprising finding in view of the evidence (cf. Hollins *et al.*, 1996) that the RA channel faithfully encodes this property of complex waves; apparently, other properties play a more important role. Second, the two waveforms in the most discriminable pair (namely 90°–270°) are mirror images of each other (see Fig. 1) and thus could be construed as more similar to one another than any other members of the stimulus set.

Intensity code

It is therefore unclear what principles underlie the discrimination results obtained in the low-frequency conditions. We examined the ability of several candidate stimulus parameters to account for the discrimination data obtained in the low-frequency condition of experiment 3. The candidate parameters we explored were peak position, peak velocity, and the maximum of the product of position and velocity. These three parameters were chosen because Meissner corpuscles are thought to be sensitive to both position and velocity (Looft 1994, 1996a, 1996b). The value of each candidate parameter was computed for each waveform for each participant, and the absolute differences in these values were obtained for all pairs of waveforms and standardized within participant. The d's from the low-frequency condition of experiment 3 were also standardized within participant, and Spearman correlations between the standardized differences in candidate parameter values and d's were computed. Differences in both peak position and peak velocity were negatively correlated with discriminability (Spearman's $\rho = -0.48, -0.50$, $p < 0.05$ for position and velocity, respectively), which implies that neither peak position nor peak velocity plays a critical role in RA-mediated low-frequency

waveform discrimination. There was a small, nonsignificant correlation between differences in the peak product of position and velocity and discriminability (Spearman's $\rho = 0.12$, $p > 0.5$).

Temporal code

Since none of these putative intensity codes accounts for the discrimination data obtained in the low-frequency conditions of experiments 1 and 3, we explored the possibility that the waveforms were encoded temporally, as sequences of perceived stimulus peaks. The subjective impressions of the first author suggested that the basis for discrimination was differences in number and relative prominence of the peaks perceived within each stimulus cycle. For some stimuli, a sequence of peaks of a given perceived amplitude, repeated at constant intervals, was perceived; for other stimuli, each cycle seemed to consist of two peaks of different amplitude. Both the 270 and 180° waveforms consisted of one very robust component alternating with another much less salient component; the lesser component was almost negligible in the 270° waveform. The 90° stimulus was perceived as a higher-frequency waveform with no distinctive components. Finally, the perception of the 0° waveform was intermediate between the 90 and 180° waveforms.

The perceived composition of a waveform is clearly not a function of its objective form alone. If this were the case, 0° would be the most temporally homogeneous, since all its (above-baseline) peaks have the same amplitude; 180° would be the most heterogeneous since it is the waveform with the greatest discrepancy in the amplitudes of its individual peaks (see Fig. 1). Another mechanism seems to be operating which somehow alters the perceived train of stimulus peaks. Inspection of the objective and subjective waveforms suggests that one stimulus peak may partially suppress a subsequent, temporally contiguous peak. A possible physiological basis for such an interaction was found in RA afferents by Looft (1994), who observed that the occurrence of one impulse could result in the suppression of a second impulse. Freeman and Johnson (1982a, 1982b) attribute this suppression to a resetting of the receptor membrane upon production of an action potential. According to their model, after mechanically induced depolarization, the membrane potential decays exponentially at a rate characterized by the membrane time constant. If a stimulus peak partially depolarizes a Meissner corpuscle without eliciting an action potential, the corpuscle will be more highly excitable for a short time due to carryover from previous stimulus-induced depolarization. On the other hand, if depolarization is sufficient to induce an action potential, the membrane hyperpolarizes to its resting membrane potential and the corpuscle is less likely to produce an action potential in response to a subsequent stimulus peak.

Another factor that may contribute to the divergence between the objective and subjective configurations of the waveforms is a transformation of the signal as it ascends the sensory pathway. One impulse may suppress a subsequent impulse due to a transmission failure at one of the several synapses between the afferent and the somatosensory cortex. Mountcastle *et al.* (1969) observed a loss of precision in the entrainment of neurons in response to a sinusoidal stimulus at successively higher levels of the lemniscal pathway. They suggest that the synaptic transfers between the afferent and the postcentral pyramidal neurons act as low-pass filters. The probability of transmission failure is a function of the recent history of the synapse: The more recently it has transmitted an action potential, the more likely it is to fail to transmit another.

A central assumption underlying this hypothesis is that the perceived amplitude of a given peak is a function of the number of spikes it produces in a population of RA cortical neurons. A decrease in the number of spikes produced in this population of RA neurons will reduce the perceived amplitude of the peak. We propose, following the observations of Mountcastle *et al.*, that such a decrease is brought about as a result of a higher probability of transmission failure when one stimulus peak is immediately preceded by another. This perceptual suppression should be greater, the higher the amplitude of the first peak, and the shorter the interval between the two peaks.

A model of low-frequency waveform discrimination

In order to test whether intrastimulus forward suppression could underlie the pattern of discrimination and SJB's reported percepts of the low-frequency stimuli presented in experiments 1 and 3, we developed a simple simulation, applying the physiological principles sketched above. The simulation consisted of a simple receptor model coupled with a simple idealization of the synaptic transmission process.

The real-time receptor model, similar to Freeman and Johnson's (1982a, 1982b) transduction model, has the following characteristics:

(1) The membrane becomes depolarized as a linear function, α, of instantaneous change in stimulus position (instantaneous stimulus velocity), v_t, and only responds to positive displacements above baseline.
(2) The membrane time constant of the simulated mechanoreceptor, τ, is 60 ms (cf. Freeman and Johnson, 1982a, 1982b).
(3) When the membrane potential reaches its threshold, an action potential is produced.
(4) Following an action potential, the membrane is unresponsive for 1 ms (cf. a change proposed to the Freeman and Johnson model by Slavik and Bell, 1995). The membrane then returns to its resting potential.
(5) Time is divided into 0.1-ms increments and the membrane potential is updated according to the procedure outlined below. Let P_t be the membrane potential and v_t be the signal velocity at time t (in μm/s); let α be the sensitivity of the corpuscle, P_r, the resting membrane potential, τ, the membrane time constant in tenths of a millisecond, and T, the receptor threshold.

$P_t = P_{t-1} + \alpha v_t - (P_{t-1} - P_r)/\tau$;

if $P_t \geqslant T$, action potential ensues and $P_t = P_r$; membrane becomes unresponsive for 1 ms.

The receptor model differed from the Freeman and Johnson (1982a, 1982b) model in the following ways:

(1) The receptor membrane becomes depolarized as a function of the positive displacements above baseline rather than a square wave. This was implemented on the basis of Looft's (1994, 1996a, 1996b) findings that Meissner corpuscles seem to be primarily sensitive to stimulus velocity above baseline at high stimulus amplitudes; we did not include bidirectional corpuscles.
(2) There is a 1-ms absolute refractory period following an action potential (cf. Slavik and Bell, 1995).

The simplified transmission (or filtering) process is implemented as follows: When the afferent produces an action potential, the signal has probability p_t of being transmitted down the sensory pathway. This probability is initially 1. Upon successful transmission of an action potential, the probability drops to 0 and increases exponentially as a function of time with time constant τ',

$$p_t = (1 - p_{t-1})/\tau'.$$

We ran 100 cycles for 20 receptors—uniformly distributed in sensitivity over 6 dB in order to mimic the effects of position with respect to the stimulus probe and differences in corpuscle sensitivities—for each phase condition, and tallied the number of action of potentials elicited in each of 1000 bins within a stimulus cycle. The gain parameter for the most sensitive receptor for each subject was chosen such that its threshold was equal to that obtained psychophysically. Based on these cycle histograms, which simulated the population response of RA cortical neurons to each waveform, we tallied the number of action potentials elicited by each stimulus peak. We then computed the ratio of the number of action potentials elicited by the larger peak (i.e., that eliciting the most action potentials) to that of the smallest peak (i.e., that eliciting the fewest action potentials). This ratio is an index of stimulus shape in that stimuli whose peaks within each cycle elicit comparable responses will produce a ratio close to 1 while stimuli whose peaks elicit responses that are very different will produce ratios greater than 1. Our hypothesis was that, to the extent that this ratio differed between stimuli, those stimuli would be discriminable.

The ratio for each waveform was computed and the difference in ratios for each waveform pair was calculated and standardized within participant. The ratios for each phase condition differed across participants because the component amplitudes of the waveforms varied across subjects and the receptor sensitivity α was adjusted to reflect individual differences in absolute threshold. In order to fit the model to the data, we set the receptor time constant τ at 60 ms (cf. Freeman and Johnson, 1982a, 1982b), adjusted the receptor sensitivity to match observer thresholds, and manipulated the transmission time constant τ'. Our hypothesis was that activity elicited by one stimulus peak effectively decreases the activity elicited by a subsequent, temporally contiguous stimulus peak. Manipulating the time constant was then the most straightforward approach to investigate whether this hypothesis could account for the discrimination data and SJB's self reports within the context of this simple model.

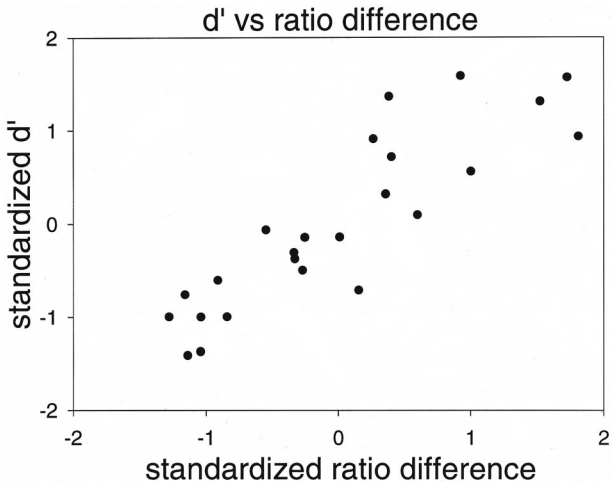


FIG. 7. Plot of d's from the low-frequency condition of experiment 3, standardized within participant, as a function of the difference in ratios, also standardized within participant, for each waveform pair. The model parameters used were: $P_r = -70$, $\tau = 600$, threshold $= -50 \pm 2$ [standard deviation (s.d.)], maximal α ranged from 1.0 to 1.35 across participants to reflect differences in threshold, and τ' ranged from 43.5 to 69.5 ms.

We increased the time constant, in increments of 0.5 ms, from 4 to 100 ms, until we obtained predictions from the model that fit the data for each subject separately. The best-fitting τ' varied from 43.5 to 69.5 ms. Figure 7 shows d's, standardized within participant, plotted as a function of these standardized differences in ratios. The Spearman correlation between d' and the standardized ratio differences was 0.89, $p < 0.01$.

The ratio index of waveform not only provided a good fit with the data but it also matched SJB's subjective reports: There was a consistent correspondence between reported waveform composition and ratio; the more homogeneous the waveform, the smaller the ratio.

This analysis provides a straightforward explanation for the discrimination data obtained in experiments 1 and 3 and is compatible with the subjective reports provided by SJB. One of the important findings in this study is that there appears to be a discrepancy between the objective configuration of waveform and its RA-mediated representation. Here we propose a simple simulation of the afferent and cortical responses to a mechanical waveform. The model we propose makes predictions as to the perceived temporal structure of a stimulus given its objective temporal structure. If indeed the RA channel encodes information temporally, this type of characterization could prove invaluable in future studies investigating the role of the RA channel in various contexts, e.g., texture perception.

CONCLUSION

We found that the high-frequency waveforms were difficult to discriminate, especially after the contribution of the RA channel had been minimized using a low-frequency adapting stimulus. Since all the waveforms presented in any given condition were of equal power, the results are compatible with the hypothesis that the PC channel integrates stimu-

lus energy over time and behaves as a critical band filter in touch (Verrillo and Gescheider, 1975; Marks, 1979; Makous *et al.*, 1995).

The results obtained using low-frequency test stimuli suggested that the RA system supports discrimination of complex waveforms as suggested by Hollins *et al.* (1996). A model of low-frequency complex waveform discrimination was proposed according to which the waveforms were discriminated on the basis of uniformity or nonuniformity of the sequences of perceived peaks within stimulus cycles. Within the context of the model, we put forth a novel way to characterize RA-mediated representations of waveform which we believe may prove useful in future research investigating the RA channel.

Hollins *et al.* (1998) have suggested that vibrotactile channels may be implicated in the perception of fine textural features. The results presented here suggest that the RA channel is well suited for encoding the complex vibrations elicited in the skin during the haptic exploration of textures. The PC channel may play a role in encoding texture by integrating the energy of the high-frequency components of these vibrations. Further elucidation of the mechanisms by which these channels code complex vibratory stimuli applied to the skin may help identify more precisely their respective roles in texture perception and in somesthesis as a whole.

ACKNOWLEDGMENTS

The authors are indebted to Jack Vevea and Tom Wallsten for valuable discussion, and to the participants for their many hours of painstaking observation. This research was supported by NSF Grant No. SBR-9514432.

Bolanowski, S. J., Gescheider, G. A., Verrillo, R. T., and Checkosky, C. M. (**1988**). ''Four channels mediate the mechanical aspects of touch,'' J. Acoust. Soc. Am. **84**, 1680–1694.

Connor, C. E., Hsiao, S. S., Phillips, J. R., and Johnson, K. O. (**1990**). ''Tactile roughness: Neural codes that account for psychophysical magnitude estimates,'' J. Neurosci. **10**, 3823–3836.

Connor, C. E., and Johnson, K. O. (**1992**). ''Neural coding of tactile texture: Comparison of spatial and temporal mechanisms for roughness perception,'' J. Neurosci. **12**, 3414–3426.

Dai, H., Verfeld, N. J., and Green, D. M. (**1996**). ''The optimum decision rules in the same-different paradigm,'' Percept. Psychophys. **58**, 1–9.

Freeman, A. W., and Johnson, K. O. (**1982a**). ''Cutaneous mechanoreceptors in macaque monkey: Temporal discharge patterns evoked by vibration, and a receptor model,'' J. Physiol. (London) **323**, 21–41.

Freeman, A. W., and Johnson, K. O. (**1982b**). ''A model accounting for effects of vibratory amplitude on responses of cutaneous mechanoreceptors in macaque monkey,'' J. Physiol. (London) **323**, 43–64.

Gescheider, G. A. (**1976**). ''Evidence in support of the duplex theory of mechanoreception,'' Sensory Process. **1**, 68–76.

Gescheider, G. A., Hoffman, K. E., Harrison, M. A., Travis, M. L., and Bolanowski, S. J. (**1994**). ''The effects of masking on vibrotactile temporal summation in the detection of sinusoidal and noise signals,'' J. Acoust. Soc. Am. **95**, 1006–1016.

Gescheider, G. A., and Verrillo, R. T. (**1979**). ''Vibrotactile frequency characteristics as determined by adaptation and masking procedures,'' in *Sensory Functions of the Skin of Humans*, edited by D. R. Kenshalo (Plenum, New York, 1979), pp. 183–206.

Gescheider, G. A., and Wright, J. H. (**1968**). ''Effects of sensory adaptation on the form of the psychophysical magnitude function for cutaneous vibration,'' J. Exp. Psychol. **77**, 308–313.

Gescheider, G. A., and Wright, J. H. (**1969**). ''Effects of vibrotactile adaptation on the perception of stimuli of varied intensity,'' J. Exp. Psychol. **81**, 449–453.

Goble, A. K., and Hollins, M. (**1993**). ''Vibrotactile adaptations enhances amplitude discrimination,'' J. Acoust. Soc. Am. **93**, 418–424.

Hollins, M., Bensmaïa, S., and Risner, R. (**1998**). ''The duplex theory of tactile texture perception,'' in *Fechner Day 98. Proceedings of the Fourteenth Annual Meeting of the International Society for Psychophysics*, edited by S. Grondin and Y. Lacouture (The International Society for Psychophysics, Québec, Canada, 1998), pp. 115–120.

Hollins, M., Delemos, K. A., and Goble, A. K. (**1991**). ''Vibrotactile adaptation on the face,'' Percept. Psychophys. **44**, 21–30.

Hollins, M., Delemos, K. A., and Goble, A. K. (**1996**). ''Vibrotactile adaptation of the RA system: A psychophysical analysis,'' in *Somesthesis and the Neurobiology of the Somatosensory Cortex*, edited by O. Franzén, R. Johansson, and L. Terenius (Birkhäuser Verlag, Basel, Switzerland), pp. 101–111.

Hollins, M., Goble, A. K., Whitsel, B. L., and Tommerdahl, M. (**1990**). ''Time course and action spectrum of vibrotactile adaptation,'' Somatosens. Motor Res. **7**, 205–221.

Hollins, M., and Risner, S. R. (**2000**). ''Evidence for the duplex theory of tactile texture perception,'' Percept. Psychophys. **62**, 695–705.

Horch, K. (**1991**). ''Coding of vibrotactile stimulus frequency by Pacinian corpuscle afferents,'' J. Acoust. Soc. Am. **89**, 2827–2836.

Looft, F. J. (**1994**). ''Response of monkey glabrous skin mechanoreceptors to random-noise sequences: I. Temporal response characteristics,'' Somatosens. Motor Res. **11**, 327–344.

Looft, F. J. (**1996a**). ''Response of monkey glabrous skin mechanoreceptors to random-noise sequences: II. Dynamic stimulus state analysis,'' Somatosens. Motor Res. **13**, 11–28.

Looft, F. J. (**1996b**). ''Response of monkey glabrous skin mechanoreceptors to random-noise sequences: III. Spectral analysis,'' Somatosens. Motor Res. **13**, 235–244.

Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press, Cambridge).

Makous, J. C., Friedman, R. M., and Vierck, Jr., C. J. (**1995**). ''A critical band filter in touch,'' J. Neurosci. **15**, 2808–2818.

Marks, L. E. (**1979**). ''Summation of vibrotactile intensity: An analog to auditory critical bands?'' Sensory Process. **3**, 188–203.

Mountcastle, V. B., Talbot, W. H., Sakata, H., and Hyvärinen, J. (**1969**). ''Cortical neuronal mechanisms in flutter vibration studied in unanesthetized monkeys. Neuronal periodicity and frequency discrimination,'' J. Neurophysiol. **32**, 452–484.

Noreen, D. L. (**1981**). ''Optimal decision rules for common psychophysical paradigms,'' in *Mathematical Psychology and Psychophysiology*, edited by S. Grossberg (Lawrence Erlbaum Associates, Inc., Hillsdale, NJ and London, England, 1981), pp. 237–280.

Slavik, P., and Bell, J. (**1995**). ''A mechanoreceptor model for rapidly and slowly adapting afferents subjected to period vibratory stimuli,'' Math. Biosci. **130**, 1–23.

Smith, W. D. (**1995**). ''Clarification of sensitivity measure A','' J. Math. Psychol. **39**, 82–89.

Sorkin, R. D. (**1962**). ''Extensions of the theory of signal detectability to matching procedures in psychoacoustics,'' J. Acoust. Soc. Am. **34**, 1745–1751.

Talbot, W. H., Darian-Smith, I., Kornhuber, H. H., and Mountcastle, V. B. (**1968**). ''The sense of flutter-vibration: Comparison of the human capacity with response patterns of mechanoreceptive afferents from the monkey hand,'' J. Neurophysiol. **31**, 301–334.

Verillo, R. T. (**1968**). ''A duplex mechanism of mechanoreception,'' in *The Skin Senses*, edited by D. R. Kenshalo (Thomas, Springfield, IL, 1968), pp. 139–156.

Verrillo, R. T., and Gescheider, G. A. (**1975**). ''Enhancement and summation in the perception of two successive vibrotactile stimuli,'' Percept. Psychophys. **18**, 128–136.

Verrillo, R. T., and Gescheider, G. A. (**1977**). ''Effect of prior stimulation on vibrotactile thresholds,'' Sensory Process. **1**, 292–300.

Perceptual calibration of *F*0 production: Evidence from feedback perturbation

Jeffery A. Jones
Department of Psychology, Queen's University, Kingston, Ontario K7L 3N6, Canada

K. G. Munhall
Department of Psychology and Department of Otolaryngology, Queen's University, Kingston, Ontario K7L 3N6, Canada

(Received 3 March 2000; accepted for publication 14 June 2000)

Hearing one's own speech is important for language learning and maintenance of accurate articulation. For example, people with postlinguistically acquired deafness often show a gradual deterioration of many aspects of speech production. In this manuscript, data are presented that address the role played by acoustic feedback in the control of voice fundamental frequency (*F*0). Eighteen subjects produced vowels under a control (normal *F*0 feedback) and two experimental conditions: *F*0 shifted up and *F*0 shifted down. In each experimental condition subjects produced vowels during a training period in which their *F*0 was slowly shifted without their awareness. Following this exposure to transformed *F*0, their acoustic feedback was returned to normal. Two effects were observed. Subjects compensated for the change in *F*0 and showed negative aftereffects. When *F*0 feedback was returned to normal, the subjects modified their produced *F*0 in the opposite direction to the shift. The results suggest that fundamental frequency is controlled using auditory feedback and with reference to an internal pitch representation. This is consistent with current work on internal models of speech motor control. © *2000 Acoustical Society of America.* [S0001-4966(00)03009-5]

PACS numbers: 43.70.Aj, 43.66.Hg [AL]

I. INTRODUCTION

There are many indications that fluent speech is controlled through the use of sophisticated internal representations as well as feedback processed on-line. For example, people with postlinguistically acquired deafness often show a deterioration of many aspects of speech production. Problems related to intensity and pitch control, as well as intonation, stress, and rate of speech are commonly seen quite soon after hearing loss. However, only after longer periods of deafness will variability in the production of vowels and consonants be observed (Cowie and Douglas-Cowie, 1992).

The finding that the precision of vowel and consonant production persists unaltered for a relatively long time after deafness onset supports the existence of a well-formed neural mapping between the motor system and the acoustic signals for segments. On the other hand, the finding that deafness more rapidly affects production parameters such as pitch and intensity implies that the mechanisms involved in suprasegmental control may be different than those for the control of segment production (Perkell *et al.*, 1997). The control of these parameters may be more directly sensitive to acoustic feedback. For example, speakers exposed to loud noise spontaneously and immediately compensate by increasing the volume of their speech (Lane and Tranel, 1971).

Uncovering how such a complex control system operates and determining the relative roles played by feedback and central representations is a daunting task requiring both empirical and modeling work. Recently, there has been considerable interest in the role of feedback and ''internal models'' in motor control in general. Internal models are hypothesized neural representations of the spatial (kinematic), force (dynamic), and/or proprioceptive characteristics of movements that could be used by the nervous system to predict movement outcome. These predictive models could provide internal feedback to planning and control systems without the delays associated with natural proprioceptive feedback [see Miall and Wolpert (1996); Kawato (1999) for discussions of the many roles of internal models in movement control].

Evidence for the existence of internal models comes primarily from the study of arm and hand movements. For example, Johanssen and Westling (1984) and Flanagan and Wing (1993) have shown that when grasping an object with the hand, grip force changes in synchrony with changes in load forces on the object. This synchrony could only result from control that predicts the loads on the object and thus the grip force needed to hold the object.

Although internal models can potentially reduce the need for closed-loop control, feedback still plays an important role in their acquisition and maintenance. Subjects exposed to novel conditions can acquire new internal models. Several investigators have shown that subjects exposed to an artificial force field while making point-to-point movements adapt and eventually make arm movements with natural trajectories. For example, Shadmehr and Mussa-Ivaldi (1994) had subjects move a robot manipulandum to targets while the robot imposed forces. Initially, the trajectories produced by the subjects were distorted; however, with practice, the subjects produced movement paths quite similar to movements produced prior to exposure to the force field. When the forces were suddenly removed, the subjects showed after-

effects in their movements for a few trials. For these trials, subjects moved as if they were encountering the experimental force field even though it was no longer present. This pattern of behavior suggests that subjects relearned the mapping between the kinematics of arm movements and the forces needed to control trajectories. In other words, they constructed a new internal model to accomplish the reaching task under the novel force conditions.

The aftereffects shown by Shadmehr and Mussa-Ivaldi are a form of sensorimotor adaptation similar to that observed with visual and vestibular perturbation paradigms [see Welch (1986) for a review]. For example, Held (1965) showed that subjects wearing prisms that displace the visual field quickly relearned the mapping between the visual space and the motor system. Initially, the subjects made reaching errors in the direction of the prism displacement. After a number of practice trials they returned to normal accuracy and normal movement speed. However, when the prisms were removed they made reaching errors in the opposite direction to the prism displacement. These data have been interpreted as evidence for a learned mapping between the movement and perceptual systems.

Similar experiments have been conducted in speech production research addressing the learned mappings between vocal tract movement and the resulting acoustics. The considerable variability that exists in vocal tract morphology means that talkers must learn the unique acoustic characteristics of their own vocal tract in order to produce the sounds of their language. In formal models of acoustic-articulatory mappings (Guenther, 1994; Hirayama *et al.*, 1994; Jordan, 1990, 1996; Kawato *et al.*, 1987) acoustic feedback plays a number of possible roles: (1) For speech sound development in children and adults and for learning new vocal tract arrangements, acoustic feedback provides the primary information about target achievement and thus is the vehicle for learning. (2) For fine motor control, the sound of the speaker's voice is used in closed-loop control of articulation. (3) For motor planning and control, the vocal acoustics provides an ongoing calibration of internal models of the speech motor system.

In this paper we explore the relative importance of the third role: the use of acoustic feedback in calibrating an internal model for the control of speaking fundamental frequency ($F0$). $F0$ is determined partly by individual anatomy and physiology and partly by a control system that relies on feedback to achieve a pitch ''target'' (Titze, 1994). The biomechanical and physiological contributions to the fundamental frequency of vocal fold vibration include the mass of the folds, the subglottic lung pressure, and tension on the folds from a network of muscles such as the cricothyroid and vocalis muscles. These biophysical factors are controlled by a complex network of cortical and brainstem centers (Larson, 1988) as well as proprioceptive (Kirchner and Wyke, 1965) and auditory (Sapir *et al.*, 1983) reflex mechanisms.

During normal conversation, the pitch of the voice varies as a function of speaking intensity, prosodic pattern, emotionality and speaking rate, but for any given individual this frequency range varies around an ''habitual'' vocal pitch (Zemlin, 1981). In this paper we test the extent to which this habitual pitch is controlled by an internal $F0$ target. As others have previously, we use a modified feedback approach. Several studies have demonstrated that when subjects hear their $F0$ feedback suddenly raised or lowered artificially, they compensate by shifting their pitch in the opposite direction (e.g., Burnett *et al.*, 1998; Kawahara, 1995). In the protocol used in this experiment, vocal pitch feedback was slowly shifted up or down in frequency without subjects' awareness. Our primary aim was to demonstrate adaptation to modified pitch feedback following return to normal feedback conditions.

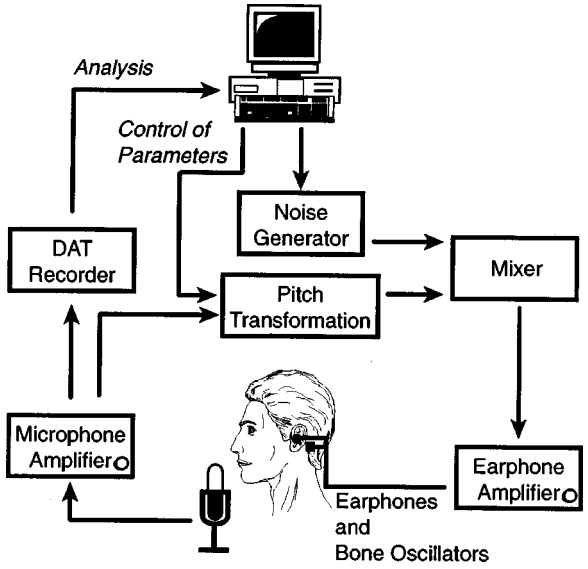


FIG. 1. Schematic of experimental acoustic feedback setup.

There are relatively few reports of auditory adaptation in speech production research. Houde and Jordan (1998) gave speakers real-time auditory feedback in which the formants they were producing were shifted enough to change the vowel's phonetic identity. Over many trials, Houde and Jordan found that speakers modified their vowel productions to compensate for the ongoing feedback transformations. In addition, the modified productions persisted in the absence of feedback, indicating an adaptive response involved in updating the acoustic-motor representation. In our experiment, if talkers show aftereffects of modified feedback conditions this will be clear evidence that habitual speaking $F0$ is controlled relative to an internally represented reference frequency.

II. METHOD

A. Subjects

Eighteen male speakers of Canadian English between 18 and 30 years of age (mean of 22.4 years) participated. The participants reported no hearing, speech, or language problems.

B. Apparatus

Figure 1 depicts the experimental setup. Utterances were recorded using a Telex PH-20 microphone. Prior to pitch shifting, the signals were amplified (Tucker-Davis MA2 mi-

crophone amplifier) and filtered (Tucker-Davis FT6-2) with a 9-kHz frequency cutoff. An Eventide Ultra Harmonizer (H3000-D/SX) transformed the pitch of the signals. The pitch-shift processing introduced only a small delay (3–4 ms). Trial initiation and the pitch processor were controlled by a computer. To reduce the amount of natural acoustic feedback, the pitch-shifted signals were mixed with pink noise (Grason-Stadler 901B) and multi-speaker babble (Auditec, St. Louis) and then attenuated by a Yorkville reference amplifier (model SR 300). The level of the masking noise was 75 dB SPL. Subjects received the auditory feedback through Etymotic (ER-2) earphones and through Radioear Model B-71 bone oscillators positioned on the left and right mastoid processes. Both the altered and unaltered signals were recorded at 48 kHz on DAT.

The experimental sessions recorded on DAT were later low-pass filtered (with a 5-kHz cutoff) and digitized with a sampling rate of 11 kHz. The fundamental frequency of utterances during each trial was calculated using an algorithm incorporated in the commercial software package CspeechSP (Milenkovic and Read, 1995). The median frequency value for each utterance made during the 3-s interval was used for subsequent analyses.

C. Procedure

Each subject was seated in a small room in front of a computer monitor. Bone oscillators were fixed to the mastoid processes using a flexible headband. The subject also wore a helmet that held a microphone at a fixed distance (7 cm) from the mouth. The transducers for the earphones were attached to a velcro strap around the subject's neck and foam inserts were positioned comfortably in the subject's ear canals.

Depicted on a computer monitor in front of the subject was the word ''awe.'' Below the word was a countdown from 3 to 0 s. Subjects were asked to produce the vowel /ɑ/ (represented orthographically by the word ''awe'') for the duration of the countdown and then to click on an icon at the bottom of the screen with a mouse to initiate the next trial.

The subjects were asked to try to produce the vowel the same way from trial to trial. However, the experimenter made no references to pitch or other voice characteristics and subjects were not made aware of the nature of the experimental manipulation.

D. Experimental design

There were three conditions in the experiment: a ''shift-up,'' ''shift-down,'' and ''control'' condition. Subjects participated in all three feedback conditions and the order of conditions was counterbalanced across subjects. The experimental sessions took place on different days to avoid vocal fatigue.

In the shift-up condition, subjects first produced ten utterances while receiving normal feedback. These ten utterances were later used to establish the subject's baseline $F0$ for the session. Following the ten baseline trials, subjects produced another 100 utterances. For each successive utterance, the pitch of their auditory feedback was increased by one cent. These trials were followed by 20 trials in which feedback was maintained at 100 cents above the subjects' true $F0$. Finally, subjects performed ten trials in which the feedback they heard was normal, that is, unaltered. From the subject's viewpoint, 140 trials were recorded without interruption. The stages of the experiment were implemented without any formal indication of changes in the feedback conditions.

The shift-down condition was conducted in exactly the same manner as in the up condition except subjects were exposed to decreasing pitch feedback to a maximum of -100 cents after the initial ten baseline trials. The pitch of their auditory feedback was decreased by one cent on each of 100 trials. Subjects were then exposed to 20 trials in which feedback was maintained at 100 cents below their true $F0$. These were followed by ten trials in which normal feedback was given.

In addition to the experimental manipulations, subjects also participated in a control condition in which they produced an equivalent number of trials without frequency manipulations. The condition was an attempt to control for $F0$ changes that may result from repeatedly producing the same sound in an experimental setting. Because insert earphones were used, the ''normal feedback'' condition was in reality a small transformation of the normal auditory feedback. Pinnae reflections change the quality of normal airborne feedback but were not present in our auditory feedback through the earphones.

The median frequency value for each utterance made during each trial was obtained and converted to cents based on the following formula:

$$\text{Cents} = 100(12 \log_2(F/B)),$$

where F is the median frequency for the utterance during the trial, B is the average of the median frequencies for the ten utterances during the baseline phase at the beginning of the experimental session.

III. RESULTS

Figure 2(a) presents the data in Hertz averaged across subjects for the control, shift-up, and shift-down conditions. As can be seen, when the pitch feedback is shifted down, the subjects raised their pitch compared to when the pitch feedback is shifted up. When the feedback was returned to normal, the $F0$ in both shift conditions changed. The mean differences between the final 20 training trials and the final 10 normal feedback trials of the experiment for the control, shift-up, and shift-down conditions were quite small in absolute terms (0.35, -2.8, and 1.5 Hz, respectively) but reliable. In the shift-up condition, $F0$ increased while the $F0$ in the shift-down condition dropped, generating a significant interaction $[F(2,34)=8.1,\ p=0.001]$. After normal feedback was returned for the test trials, the mean pitch for the shift-up condition increased significantly ($p=0.002$) and shift-down conditions decreased significantly ($p=0.047$).

Since the experimental sessions for each subject took place on different days, and because the subjects could have different baseline $F0$'s during the different sessions, we also converted the data to cents for comparison within and be-

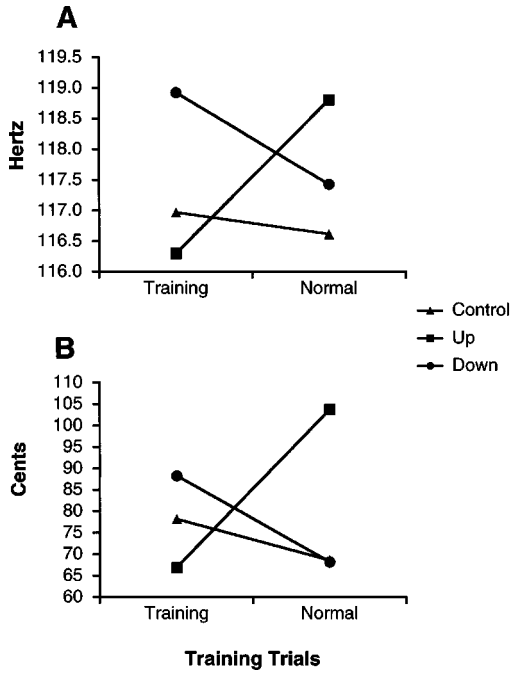


FIG. 2. Average fundamental frequency in Hertz (a) and cents (b) for the three feedback conditions (normal control, shifted up, shifted down) during the final 20 trials of the training period and when feedback was returned to normal.

tween subjects. The conversion of frequency values to cents served to normalize the data with respect to $F0$ baseline trials produced at the beginning of each experimental session.

Figure 2(b) shows the mean $F0$ in cents for the last 20 trials of the training period (i.e., $F0$ shifted 100 cents) and the $F0$ for the final 10 trials of the experiment (i.e., with normal $F0$ feedback). As can be seen, the data in cents show the same pattern. In the shift-up condition, $F0$ increased in response to normal feedback while the $F0$ in the shift-down condition dropped, generating a significant interaction $[F(2,34)=6.2,\ p=0.005]$.

Figure 3 shows the pitch patterns in cents for the three conditions during the 120 training trials. As can be seen, all conditions show an increase in $F0$ with respect to their baseline values with all conditions showing a significant linear trend $[F(1,17)=9.63,\ p=0.006]$. However, the shape of the function for each condition differed. The pitch values for the shift-down condition diverge from the control condition and become higher during the training than the other conditions. Conversely, the pitches during the shift-up condition diverged downward from the control condition. The pitches produced in the three conditions reached their maximum separation at the end of the 100 trials of training. This separation is reduced slightly and the pitch values for the shift-up and shift-down conditions tend to converge toward the control condition performance during the final 20 training trials in which pitch feedback was maintained at 100 cents above or below subjects' true $F0$. Figure 4 shows the produced and heard $F0$ in Hertz for the same 120 training trials. As noted above, the $F0$ shifts were intentionally small to avoid detection by subjects.

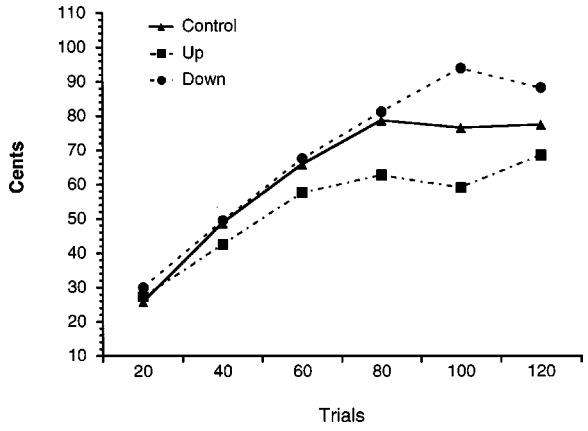


FIG. 3. Average fundamental frequency in cents as a function of blocks of 20 trials during the training period for the three feedback conditions (normal feedback control, shifted up, shifted down).

FIG. 4. Average produced and heard fundamental frequency in Hertz as a function of blocks of 20 trials during the training period for the two altered feedback conditions (shifted up, shifted down).

Subjects produced the vowels under loud auditory feedback conditions and it is possible that the observed $F0$ changes could have been due to shifts in speaking volume over the course of the training period. To test for this, the root-mean-square (rms) amplitude of the vowels was computed for the initial baseline ten trials and the final ten test trials for each of the three feedback conditions. An ANOVA showed no effects of training time [initial baseline versus final test trials: $F(1,17)=0.001,\ p>0.5$], feedback condition [upward, downward, control: $F(2,34)=0.592,\ p>0.5$], nor the interaction of these two variables $[F(2,34)=0.113,\ p>0.5]$. Thus pitch changes associated with speaking volume adjustment do not account for the observed $F0$ modifications.

In post-experimental interviews, none of the subjects reported being aware of the gradual shifts in $F0$ feedback. While all subjects were aware that something had been manipulated when $F0$ was returned to normal at the end of the training conditions, they were at a loss to explain what had transpired.

IV. DISCUSSION

The data in this study show two related effects of feedback transformation on $F0$. During the training period sub-

jects *compensated* for the pitch shifts in an apparent attempt to maintain habitual pitch targets under feedback control. When pitch was shifted up subjects lowered their pitch relative to a control condition; when pitch was shifted down they raised their pitch relative to the control. The subjects also showed evidence of sensorimotor *adaptation*. Aftereffects resulted from the relatively short period of exposure to the altered $F0$ feedback. When subjects heard $F0$ feedback that was higher than their true $F0$, the pitch of their voice increased when they were unexpectedly given normal, unaltered auditory feedback. The opposite effect was observed when subjects heard $F0$ feedback lower than it actually was.

Multiple components must be involved in the vocal pitch control system responsible for this behavior. Some have suggested that there is an ''optimum'' pitch range (Zemlin, 1981) determined by the anatomy and physiology of the vocal mechanism. Since fundamental frequency varies as a function of prosody, speaking volume, social situation, emotional state, etc., a mechanism must exist for the controlled modulation of this ''natural level.'' A number of studies (e.g., Kawahara, 1995; Larson *et al.*, 2000) and the present compensation data have shown that auditory feedback can be used in a closed-loop fashion to control fundamental frequency. Presumably, talkers also use kinesthetic feedback or its perceptual concomitant, vocal effort, to aid in this control. In addition, the present data suggest that some type of internal model or representation plays a role in the long-term calibration of vocal pitch. This conclusion is supported by the aftereffect or adaptation data.

The current adaptation results are analogous to those found in classic prism experiments (e.g., Held, 1965). Following a training period wearing displacing prisms, subjects make errors for a short time in the opposite direction of the prism displacement. This aftereffect is widely considered to result from adaptation. The mechanism of the adaptation remains controversial, but it is clear, that at some level a remapping between retinal space and body space must occur. In our data, the subjects acted as if a remapping between perceived and produced pitch had taken place.

It is difficult to determine at what level this $F0$ remapping is taking place. It may be that a representation of a base or neutral pitch level is modulated during the training phase. However, habitual pitch (i.e., average $F0$ from day to day) appears to be quite variable and thus a narrow $F0$ target range seems unlikely. In a study by Coleman and Markham (1991), habitual pitch was found to vary as much as plus or minus three semitones or approximately 18% (cf. Titze, 1994). On the other hand, subjects may attempt to match pitch with perceived vocal effort or kinesthetic feedback (Guenther, 1994). When pitch is shifted up or down, it may not be the absolute pitch value that drives compensation, but the discrepancy between the kinesthetic and auditory feedback.

In this study, there was a tendency for subjects to gradually increase their pitch during the experimental session independent of the feedback condition (see Fig. 3). Since we did not ask subjects to maintain a particular loudness level one possible explanation for this pattern is that subjects increased their speaking volume during the session, causing an increase in pitch (Gramming *et al.*, 1988). However, there was no significant difference between the rms amplitudes of the utterances during the sessions. It is also possible that the increased pitch is related to vocal fatigue developed over the session. Unfortunately, there is no established method for assessing vocal fatigue from the acoustic record (Titze, 1994). It should be noted that the tendency for vocal pitch to increase even when there is no feedback manipulation underscores the value of the control condition. If only one shift condition had been tested with no control group, the effect size of the feedback manipulation could not be determined.

While this trend for pitch to increase is controlled for, it obscures the sensitivity of subjects to the pitch shifts. None of the subjects was consciously aware that the pitch feedback was being modified but their control systems ultimately responded to the changes. The $F0$ patterns in each training condition could be used to test at what size of pitch shift the $F0$ control system began to compensate. Figures 3 and 4 show that for small shifts at the beginning of the training period all three conditions show a similar function and only with larger shifts do the experimental conditions diverge from the control. For upward shifts, the subjects diverged from the control condition earlier than for the downward shifts. Unfortunately, all conditions were tested on different days, so the conditions producing the gradual increase may not have been constant across days. This question must await further study.

Klatt (1973) has shown that subjects can make quite fine perceptual discriminations (between 0.3 and 0.5 Hz) in the $F0$ of synthesized vowels with flat $F0$ contours. When the vowels were synthesized with linearly decreasing $F0$ or as a diphthong with a natural $F0$ contour, the discrimination threshold rose above 2 Hz. Since the final pitch shifts in the experiment were small and in this range, our subjects may have been operating at their perceptual limen. However, evidence from visual-motor control (e.g., Milner and Goodale, 1995) indicates that there can exist perceptual systems for the control of action that are separate from the perceptual system used in categorical judgements. A recent magnetoencephalography study has provided evidence that suggests that this also could be true for the auditory system in speech. Houde *et al.* (2000) have shown that the auditory cortex responds differently to hearing one's own speech while producing it versus listening to recordings of one's own speech. Thus the Klatt threshold data may not be relevant to the issue of sensitivity to feedback modification.

The kind of short-term learning that was observed in this study has been reported in many speech studies previously. Subjects adapt to various static [e.g., bite block (McFarland and Baum, 1995; palatal prostheses, Baum and McFarland, 1997; Hamlet and Stone, 1976, 1978; Hamlet *et al.*, 1978)] or dynamic (Gracco and Abbs, 1986) physical perturbations and auditory feedback transformations (e.g., Houde and Jordan, 1998). We have chosen to consider this learning in the context of internal models. The general concept of an internal model is not a new one. Similar roles have been played by motor programs (e.g., Keele, 1968), efference copy (e.g., von Holst and Mittelstaedt, 1950) and feedforward control (e.g., Arbib, 1981). In proposals about speech production,

internal models for vocal tract geometry, kinesthetic, and acoustic mappings have been postulated. Our work suggests that the acoustic mapping must be differentiated to include an *F*0 model as well [see Kawato (1999) for a discussion of multiple internal models].

This suggestion is consistent with evidence from clinical populations. Post-linguistically deafened individuals often have difficulty producing normal intonations soon after their hearing is lost (Cowie and Douglas-Cowie, 1992). Perkell and his colleagues have shown that better *F*0 control is achieved after activation of cochlear implants (Perkell *et al.*, 1992). Perkell *et al.* (1997) have proposed that the auditory system uses information regarding conditions for intelligibility (e.g., ambient noise, social context) in a closed-loop fashion to rapidly make adjustments in *F*0 and vocal intensity. The data in the present study suggest that in addition to this closed-loop control, auditory feedback may also play a role in establishing a baseline for the controlled parameters.

ACKNOWLEDGMENTS

This work was funded by NIH Grant No. DC-00594 from the National Institute of Deafness and other Communications Disorders and NSERC. Chuck Larson and David McFarland made helpful comments on an earlier draft of this manuscript.

Arbib, M. A. (**1981**). ''Perceptual structures and distributed motor control,'' *Handbook of Physiology—The Nervous System II*, edited by V. B. Brooks (American Physiology Society, Bethesda, MD).

Baum, S. R., and McFarland, D. H. (**1997**). ''The development of speech adaptation to an artificial palate,'' J. Acoust. Soc. Am. **102**, 2353–2359.

Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (**1998**). ''Voice *F*0 responses to manipulations in pitch feedback,'' J. Acoust. Soc. Am. **103**, 3153–3161.

Coleman, R. F., and Markham, I. W. (**1991**). ''Normal variations in habitual pitch,'' J. Voice **5**, 173–177.

Cowie, R., and Douglas-Cowie, E. (**1992**). ''Postlingually acquired deafness,'' in *Trends in Linguistics, Studies and Monographs* (Mouton de Gruyter, New York), Vol. 62.

Flanagan, J. R., and Wing, A. M. (**1993**). ''Modulation of grip force with load force during point-to-point arm movements,'' Exp. Brain Res. **95**, 131–143.

Gracco, V. L., and Abbs, J. H. (**1986**). ''Variant and invariant characteristics of speech movements,'' Exp. Brain Res. **65**, 156–166.

Gramming, P., Sundberg, J., Ternström S., Leanderson, R., and Perkins, W. H. (**1988**). ''Relationship between changes in voice pitch and loudness,'' J. Voice **2**, 118–126.

Guenther, F. (**1994**). ''A neural network model of speech acquisition and motor equivalent production,'' Biol. Cybern. **72**, 43–53.

Hamlet, S., and Stone, M. (**1976**). ''Compensatory vowel characteristics resulting from the presence of different types of experimental dental prosthesis,'' J. Phonetics **4**, 199–218.

Hamlet, S., and Stone, M. (**1978**). ''Compensatory alveolar consonant production induced by wearing a dental prosthesis,'' J. Phonetics **6**, 227–248.

Hamlet, S., Stone, M., and McCarty, T. (**1978**). ''Conditioning dentures viewed from the standpoint of speech adaptation,'' J. Prosthet. Dent. **40**, 60–66.

Held, R. (**1965**). ''Plasticity in sensory-motor systems,'' Sci. Am. **213**, 84–94.

Hirayama, M., Vatikiotis-Bateson, E., and Kawato, M. (**1994**). ''Inverse dynamics of speech motor control,'' in *Advances in Neural Information Processing Systems*, edited by S. J. Hanson, J. D. Cowan, and C. L. Giles (Morgan Kaufmann, San Mateo, CA), Vol. 6, pp. 1043–1050.

Houde, J. F., and Jordan, M. I. (**1998**). ''Sensorimotor adaptation in speech production,'' Science **279**, 1213–1216.

Houde, J., Nagarajan, S., and Mersenich, M. (**2000**). ''Modulation of auditory cortex during speech production: An MEG study,'' *Proceedings of the 5th Seminar on Speech Production: Models and Data*, pp. 249–252.

Johansson, R. S., and Westling, G. (**1984**). ''Roles of glabrous skin receptors and sensorimotor memory in automatic-control of precision grip when lifting rougher or more slippery objects,'' Exp. Brain Res. **56**, 550–564.

Jordan, M. I. (**1990**). ''Motor learning and the degrees of freedom problem,'' *Attention and Performance*, edited by M. Jeannerod (Lawrence Erlbaum, Hillsdale, NJ), Vol. XIII, pp. 796–836.

Jordan, M. I. (**1996**). ''Computational aspects of motor control and motor learning,'' *Handbook of Perception and Action*, edited by H. Heuer and S. Keele (Academic, New York), Vol. 2, pp. 71–120.

Kawahara, H. (**1995**). ''Transformed auditory feedback: The collection of data from 1993.1 to 1994.12 by a new set of analysis procedures,'' ATR Technical Report, TR-H-120.

Kawato, M. (**1999**). ''Internal models for motor control and trajectory planning,'' Curr. Opin. Neurobiol. **9**, 718–727.

Kawato, M., Furukawa, K., and Suzuki, R. (**1987**). ''A hierarchical neural-network model for control and learning of voluntary movement,'' Biol. Cybern. **57**, 169–185.

Keele, S. W. (**1968**). ''Movement control in skilled motor performance,'' Psychol. Bull. **70**, 387–403.

Kirchner, J. A., and Wyke, B. D. (**1965**). ''Articular reflex mechanisms in the larynx,'' Ann. Otol. Rhinol. Laryngol. **74**, 748–768.

Klatt, D. (**1973**). ''Discrimination of fundamental frequency contours in synthetic speech: Implications for models of pitch perception,'' J. Acoust. Soc. Am. **53**, 8–16.

Lane, H., and Tranel, B. (**1971**). ''The Lombard sign and the role of hearing in speech,'' J. Speech Hear. Res. **14**, 677–709.

Larson, C. R. (**1988**). ''Brain mechanisms involved in the control of vocalization,'' J. Voice **2**, 301–311.

Larson, C. R., Burnett, T. A., Kiran, S., and Hain, T. C. (**2000**). ''Effects of pitch-shift velocity on voice *F*0 responses,'' J. Acoust. Soc. Am. **107**, 559–564.

McFarland, D. H., and Baum, S. R. (**1995**). ''Incomplete compensation to articulatory perturbation,'' J. Acoust. Soc. Am. **97**, 1865–1873.

Miall, R. C., and Wolpert, D. M. (**1996**). ''Forward models for physiological motor control,'' Neural Networks **9**, 1265–1279.

Milenkovic, P., and Read, C. (**1995**). *CSpeechSP Quick Reference Manual*, University of Wisconsin.

Milner, A. D., and Goodale, M. A. (**1995**). *The Visual Brain in Action* (Oxford University Press, New York).

Perkell, J., Lane, H., Svirsky, M. A., and Webster, J. (**1992**). ''Speech of cochlear implant patients: A longitudinal study of vowel production,'' J. Acoust. Soc. Am. **91**, 2961–2979.

Perkell, J., Matthies, M., Lane, H., Guenther, F., Wilhelms-Tricarico, R., Wozniak, J., and Guoid, P. (**1997**). ''Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models,'' Speech Commun. **22**, 227–250.

Sapir, S., McClean, M. D., and Larson, C. R. (**1983**). ''Human laryngeal responses to auditory stimulation,'' J. Acoust. Soc. Am. **73**, 315–321.

Shadmehr, R., and Mussa-Ivaldi, F. A. (**1994**). ''Adaptive representation of dynamics during learning of a motor task,'' J. Neurosci. **14**, 3208–3224.

Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs, NJ).

von Holst, E., and Mittelstaedt, H. (**1950**). *The principle of reafference: Interactions between the central nervous system and the peripheral organs*, edited by P. C. Dodwell (Appleton-Century-Crofts, New York, 1971).

Welch, R. B. (**1986**). ''Adaptation of space perception,'' in *Handbook of Perception and Human Performance*, edited by K. R. Boff, L. Kaufman, and J. P. Thomas (Wiley, New York).

Zemlin, W. R. (**1981**). *Speech and Hearing Science: Anatomy and Physiology*, 2nd ed. (Prentice-Hall, Englewood Cliffs, NJ).

Acoustic characteristics of English fricatives

Allard Jongman[a)]
Linguistics Department, Blake Hall, The University of Kansas, Lawrence, Kansas 66045

Ratree Wayland
Program in Linguistics, Yon Hall, University of Florida at Gainesville, P. O. Box 115454, Gainesville, Florida 32611-5454

Serena Wong
Columbia University College of Physicians & Surgeons, 630 West 168th Street, New York, New York 10032

(Received 3 March 2000; accepted for publication 11 June 2000)

This study constitutes a large-scale comparative analysis of acoustic cues for classification of place of articulation in fricatives. To date, no single metric has been found to classify fricative place of articulation with a high degree of accuracy. This study presents spectral, amplitudinal, and temporal measurements that involve both static properties (spectral peak location, spectral moments, noise duration, normalized amplitude, and $F2$ onset frequency) and dynamic properties (relative amplitude and locus equations). While all cues (except locus equations) consistently serve to distinguish sibilant from nonsibilant fricatives, the present results indicate that spectral peak location, spectral moments, and both normalized and relative amplitude serve to distinguish all four places of fricative articulation. These findings suggest that these static and dynamic acoustic properties can provide robust and unique information about all four places of articulation, despite variation in speaker, vowel context, and voicing. © *2000 Acoustical Society of America.* [S0001-4966(00)02909-X]

PACS numbers: 43.70.Fq [AL]

I. INTRODUCTION

One of the primary goals of speech research is to characterize the defining properties of speech sounds that occur in natural language, and to determine how the listener extracts these properties in the process of speech perception. Phonetic research of the past 50 years has demonstrated that the identification of acoustic cues which uniquely characterize particular (classes of) speech sounds is a serious challenge. A major obstacle in this endeavor is the variability typically found in the speech signal, often resulting in a defective one-to-one correspondence between acoustic cue and phonetic percept (Liberman *et al.*, 1967). This lack of invariance arises from a variety of sources, including speaker size, phonetic context, and speaking rate (see Pisoni and Luce, 1986, for an overview). The basic problem, then, is how perceptual constancy or invariance is achieved in the presence of such varying information.

Much research has been devoted to the question of whether distinct spectral patterns that correspond to phonetic dimensions, such as place and manner of articulation, can be derived from the acoustic waveform. Early studies failed to find any consistent mapping between acoustic properties and phonetic features (e.g., Cooper *et al.*, 1952; Schatz, 1954; Delattre *et al.*, 1955). Some recent research, however, suggests that stable, consistent acoustic properties may indeed be found in the speech signal, with appropriate analyses (e.g., Stevens and Blumstein, 1981; Kewley-Port, 1983; Lahiri *et al.*, 1984; Forrest *et al.*, 1988; Sussman *et al.*, 1991). Such research has predominantly focused on the search for properties distinguishing place of articulation in (English) stop consonants. In contrast, fricatives have been studied in much less detail. Moreover, it is uncertain whether the classification metrics proposed for stop consonants can be successfully applied to fricatives. The current study contributes to the body of research on the mapping between acoustic properties and phonetic categories by providing a detailed look at this mapping for English fricatives.

Fricatives are produced with a very narrow constriction in the oral cavity. A rapid flow of air through the constriction (the position of which depends on the particular fricative) creates turbulence in the flow, and the random velocity fluctuations in the flow act as a source of sound (e.g., Stevens, 1971, 1998; Shadle, 1990). English fricatives are usually grouped into four classes according to their place of articulation: labiodental /f,v/, (inter)dental /θ,ð/, alveolar /s,z/, and palato-alveolar /ʃ,ʒ/. Most studies of fricatives exclude /h/, since it is considered the voiceless counterpart of the abutting vowel (e.g., Pike, 1943; Ladefoged, 1982), and for that reason /h/ will not be considered in the present study either.

Previous studies of fricatives have concentrated on four attributes: spectral properties of the frication noise, amplitude of the noise, duration of the noise, and spectral properties of the transition from the fricative into the following vowel. In general, these studies have documented acoustic differences between the sibilant (/s,z,ʃ,ʒ/) and nonsibilant (/f,v,θ,ð/) fricatives, which involve spectrum, amplitude, and duration of the frication noise. Additionally, /s,z/ may be distinguished from /ʃ,ʒ/ on the basis of noise spectrum, while there are some data suggesting that /f,v/ may be distinguished from /θ,ð/ on the basis of transition information.

[a)]Electronic mail: jongman@ukans.edu

However, no cue has been identified so far that can uniquely distinguish all four places of articulation.

The present study is a comprehensive comparative analysis of acoustic cues to place of articulation in English fricatives. Spectral parameters include spectral peak location, spectral moments, locus equations, and $F2$ onset. Amplitudinal parameters include overall noise amplitude as well as relative amplitude. Temporal measurements consist of fricative noise durations. The data reported here thus concern both static and dynamic properties. Static properties pertain to acoustic information that is measured at one location of the speech signal, while dynamic properties pertain to changes in acoustic information during the fricative and/or adjacent segments. Spectral peak location, spectral moments, $F2$ onset frequency, noise amplitude, and noise duration are considered static properties. Dynamic properties include locus equations and relative amplitude. Inclusion of both static and dynamic parameters may result in a more comprehensive characterization of fricative acoustics. In particular, the goal of this study is to identify stable acoustic cues to place of articulation, to evaluate the nature of these cues: are they primarily in terms of spectrum, amplitude, or duration, and, finally, to determine their location: are these cues uniformly distributed throughout the fricative, or are some regions more informative than others?

A. Spectral properties

1. Frication noise: Spectral peak location and spectral moments

The overall spectral shape of each fricative is determined by the size and shape of the oral cavity in front of the constriction. The longer this anterior cavity, the more defined the resulting spectrum (e.g., Stevens, 1998). As a result, the alveolar and palato-alveolar fricatives are characterized by well-defined, distinct spectral shapes while labiodental and (inter)dental fricatives display a relatively flat spectrum (e.g., Strevens, 1960; Jassem, 1965; Behrens and Blumstein, 1988a). In particular, /ʃ,ʒ/ typically exhibit a midfrequency spectral peak at around 2.5–3 kHz which often corresponds to $F3$ of the following vowel. Alveolar /s,z/ are produced with a shorter anterior cavity relative to /ʃ,ʒ/ and therefore display a primary spectral peak at higher frequencies, around 4 to 5 kHz. In addition, since for these fricatives the airstream hits the teeth, the high-frequency turbulence is very intense. Both /f,v/ and /θ,ð/ are characterized by a relatively flat spectrum with no clearly dominating peak in any particular frequency region.

Previous studies reveal that the local spectral properties of frication noise serve to distinguish the sibilant fricatives /s,z,ʃ,ʒ/ as a group from the nonsibilants /f,v,θ,ð/. Within the sibilants, /s,z/ can also be distinguished from /ʃ,ʒ/ on the basis of the spectral properties of the noise (e.g., Hughes and Halle, 1956; Strevens, 1960; Heinz and Stevens, 1961; Shadle, 1990; Behrens and Blumstein, 1988a; Evers, Reetz, and Lahiri, 1998). However, the location of the spectral peaks in the frication noise is to some extent speaker dependent (Hughes and Halle, 1956) and vowel dependent (Soli, 1981). Recently, Tabain (1998) obtained high classification rates for sibilants and moderate rates for nonsibilants. Averaged spectra were calculated based on a series of fast Fourier transforms (FFTs) across each fricative. These spectra were then subjected to a classification algorithm based on a Bayesian distance measure. Classification across five male and five female speakers averaged 97% for the sibilants but only 70% for the nonsibilants. Unfortunately, for the nonsibilants, there were no consistent differences in the spectra which correlated with classification accuracy. In other words, it was not clear which acoustic properties contributed to correct classification.

Spectral moments analysis involves a statistical procedure for classifying obstruents, capturing both local (mean frequency) and global (spectral tilt and peakedness) aspects of speech sounds. These analyses may be based on one or multiple regions of the speech signal. In Forrest *et al.* (1988), a series of FFTs was calculated every 10 ms from the onset of the word-initial obstruent. Each FFT was treated as a random probability distribution from which the first four moments (mean, variance, skewness, and kurtosis) were computed. Mean and variance reflect the average energy concentration and range, respectively. Skewness is an indicator of a distribution's asymmetry. A skewness of zero indicates a symmetrical distribution around the mean. Skewness is positive when the right tail of the distribution extends further than the left tail. Likewise, skewness is negative when the left tail of the distribution extends further than the right tail (e.g., Newell and Hancock, 1984). In phonetic terms, skewness refers to spectral tilt, the overall slant of the energy distribution. Positive skewness suggests a negative tilt with a concentration of energy in the lower frequencies. Negative skewness is associated with a positive tilt and a predominance of energy in the higher frequencies. Finally, kurtosis is an indicator of the peakedness of the distribution. Positive kurtosis values indicate a relatively high peakedness (the higher the value, the more peaked the distribution), while negative values indicate a relatively flat distribution. Positive kurtosis thus suggests a clearly defined spectrum with well-resolved peaks, while negative kurtosis indicates a flat spectrum without clearly defined peaks. The spectral moments metric thus incorporates both local (spectral peak) and more global (spectral shape) information.

Forrest *et al.* (1988) derived spectral moments for a small corpus of syllable-initial fricatives (''see, she, fought, thought, fat'') produced by five females and five males. These moments were then entered into a discriminant analysis for classification in terms of place of articulation. Classification based on the first 20 ms of the fricative was good for sibilants (85% for /s/, 95% for /ʃ/); however, classification of nonsibilants was poor (58% for /θ/, 75% for /f/). Classification rates for individual moments were not reported.

Although promising as a technique to quantify spectral properties of obstruents, surprisingly little research has attempted to replicate or extend the Forrest *et al.* (1988) findings. In a preliminary report, Shadle and Mair (1996) analyzed all eight English fricatives produced by only one female and one male speaker. Moments were computed at the beginning, middle, and end of each fricative. Moments did capture some important fricative characteristics: the second moment (variance) was large for the nonsibilant frica-

tives, and /ʃ/ was uniquely characterized by a low first moment (mean). Nevertheless, the authors concluded that spectral moments did not reliably differentiate fricative place of articulation.

The most comprehensive study to date is that by Tomiak (1990), who reported all moments for the four voiceless fricatives and /h/ as produced by six American speakers. Although Tomiak (1990) did not subject her measurements to analyses of variance, she reported the following observations: /θ/ displayed a greater standard deviation, skewness, and kurtosis than /f/; /s/ was distinct from /ʃ/, having a higher mean, lower standard deviation, and greater kurtosis. Discriminant analysis yielded poor classification rates for the nonsibilant fricatives (67% for /f/, 44% for /θ/) and high rates for the sibilants (96% for both /s/ and /ʃ/).

Most studies using spectral moments have concentrated on the spectral mean and report that /ʃ/ has a lower mean than /s/ (e.g., Nittrouer *et al.*, 1989; Tjaden and Turner, 1997 for normal controls). Nittrouer (1995) and McFarland *et al.* (1996) have reported that spectral moments 1, 3, and 4 (mean, skewness, and kurtosis, respectively) distinguish /s/ from /ʃ/ across male and female adult speakers and different vowel contexts. Specifically, /ʃ/ was characterized by a lower spectral mean, positive skewness, and smaller kurtosis, indicating a slightly flatter spectrum.

In sum, while some spectral moments distinguished /s/ from /ʃ/, spectral moments have not been shown to reliably differentiate the nonsibilants.

2. Transition information: Locus equations and F2 onset

Locus equations are based on the second formant frequency ($F2$) at vowel onset and at vowel midpoint (e.g., Sussman *et al.*, 1991; Sussman, 1994) and constitute a dynamic representation of speech sounds since they express a relation between $F2$ at different points in the speech signal. Results indicate that the apparent $F2$ starting frequency of a vowel preceded by an obstruent provides information about the articulatory configuration used to generate the consonant. Although locus equations have recently been successful in the classification of place of articulation in voiced stop consonants, researchers have only just begun to apply this method to fricatives (e.g., Wilde, 1993; Fowler, 1994; Sussman, 1994; Sussman and Shore, 1996; Yeou, 1997). At present, there are very few data on fricative locus equations, and the results are contradictory: Fowler (1994) and Yeou (1997) obtained good classification of fricatives, with each place of articulation characterized by a distinct slope and *y* intercept. Yeou (1997) investigated locus equations for Arabic stops and fricatives. Slope and *y*-intercept values uniquely distinguished those fricatives that are common to Arabic and English (/f,ð,s,ʃ/). However, overlap occurred between postalveolar /ʃ/ and pharyngeal /h/ and between labial /f/ and uvular /χ/ in terms of both slope and *y* intercept.

Unfortunately, there is little correspondence between the values observed across these two studies for each place of articulation. The only qualitative agreement is that the labiodental place has the highest slope and lowest *y*-intercept value (Fowler: /v/ 0.73 and 337 Hz, respectively; Yeou: /f/ 0.92 and 61 Hz, respectively). In two smaller-scale studies, Wilde (1993) and Sussman (1994) did not obtain unique classification. In his analysis of the voiced fricatives /v,ð,z,ʒ/ of four speakers, Sussman (1994) found that only labiodental /v/ was significantly different in terms of slope (0.74) from the other three places of articulation. Similarly, although Wilde did not provide any statistics, only /f,v/ seem different from the other three places of articulation.

Wilde (1993) provides preliminary data suggesting that the onset of $F2$ alone at the fricative–vowel boundary or its range varies systematically as a function of place of articulation. Based on data from two speakers, Wilde (1993) observed that, for a given vowel context, $F2$ onset is progressively higher as the place of constriction moves back in the oral cavity. Studies investigating effects of formant transition information on perception of the /s/–/ʃ/ distinction also typically employ synthetic stimuli in which $F2$ onset frequency for /ʃ/ is substantially higher (approximately 100–300 Hz) than for /s/ (e.g., Mann and Repp, 1980; Whalen, 1981; Nittrouer, 1992). In addition, Wilde (1993) presented data that indicate that the range of $F2$ onset is progressively smaller as place of constriction moves further back, as had been previously reported for stop consonants by Kewley-Port (1982). These findings are also consistent with Recasens' (1985) observation that consonants with a greater degree of tongue-body raising (and thus typically a more posterior place of articulation) are more resistant to coarticulation.

B. Amplitude

1. Overall noise amplitude

Most research concerned with frication amplitude has investigated the overall amplitude of fricatives. These studies (e.g., Strevens, 1960; Behrens and Blumstein, 1988a) have focused on voiceless fricatives and converge on similar findings: sibilant /s,ʃ/ have a substantially greater (10–15 dB) amplitude than nonsibilant /f,θ/. Within each group, however, the two fricatives are not different from each other.

2. Relative amplitude

It has been suggested that overall amplitude may not be the relevant parameter; instead, a change in amplitude of the frication relative to the vowel in a specific frequency region may vary with place of articulation (Stevens, 1985). However, to date, no systematic acoustic study has been conducted to determine the magnitude of differences in relative amplitude as a function of place of articulation. Instead, research on relative amplitude has focused on its role in perception (e.g., Stevens, 1985; Hedrick and Ohde, 1993; Hedrick, 1997; Hedrick and Carney, 1997). For example, in order to create appropriate synthetic stimuli, Hedrick and Ohde (1993) measured relative amplitude for /s,S/ in the context of /ɑ/ produced by a female speaker. Relative amplitude, defined as the difference between fricative and vowel amplitude in the $F3$ region for sibilants, was -17 dB for /s/ and $+16$ dB for /ʃ/, suggesting that relative amplitude may distinguish sibilant fricatives in terms of place. Indeed, relative amplitude was shown to be a cue to perception of the place contrast between /s/ and /ʃ/ (Stevens, 1985; Hedrick

and Ohde, 1993). In addition, in an /s–θ/ labeling task, relative amplitude values of −20 to 0 dB were shown to yield /θ/ responses, while values of 10 to 20 dB elicited /s/ responses (Hedrick and Ohde, 1993). Unfortunately, no relative amplitude measures were provided for /θ/, nor has anyone investigated relative amplitude in the labiodental fricatives.

C. Noise duration

Noise duration serves to distinguish sibilant from nonsibilant fricatives, with /s,ʃ/ being longer than /f,θ/ (e.g., Behrens and Blumstein, 1988a). However, Behrens and Blumstein (1988a) found no difference in duration between /s/ and /ʃ/ and only a trend for /θ/ to be shorter than /f/. Noise duration does provide a robust cue to the voicing distinction in syllable-initial position, with voiceless fricatives having longer noise durations than voiced fricatives. This observation holds both for fricatives in isolated syllables (e.g., Behrens and Blumstein, 1988a; Baum and Blumstein, 1987) and in connected speech (Crystal and House, 1988).

In sum, acoustic studies focusing on the frication noise have shown that properties of the spectrum, amplitude, and duration of the noise can all serve to distinguish the sibilant /s,z,ʃ,ʒ/ from the nonsibilant /f,v,θ,ð/ fricatives. In addition, spectral properties serve to distinguish /s/ from /ʃ/, with /s/ having a concentration of energy in higher frequencies than /ʃ/. None of the noise properties alone, however, seems adequate to distinguish /f,v/ from /θ,ð/. More recent metrics such as spectral moments, locus equations, and relative amplitude show some promise for the distinction between labiodental and dental fricatives, although studies examining all eight fricatives with these metrics are few. The present study therefore consists of a comprehensive investigation of English fricatives, incorporating both recent and more traditional approaches with the aim of establishing stable acoustic cues to all four places of fricative articulation.

II. EXPERIMENT

A. Method

1. Participants

Twenty speakers (ten females and ten males) were recruited from the Cornell University student population. All were native speakers of American English, representing a variety of regional backgrounds. No participants reported any known history of either speech of hearing impairment. Participants were paid for their participation.

2. Materials

The eight English fricatives /f,v,θ,ð,s,z,ʃ,ʒ/ were recorded in consonant–vowel–consonant (CVC) syllables in the carrier phrase ''Say — again.'' The fricatives were in initial position, followed by each of six vowels /i,e,æ,ɑ,o,u/. The final consonant was always /p/. Each CVC token was repeated three times, yielding a total of 144 tokens per subject (8 fricatives×6 vowels×3 repetitions).

3. Procedure and analysis

Speakers were recorded in the Cornell Phonetics Laboratory, in a soundproof booth (IAC) with a high-quality microphone (Electro-Voice RE20), microphone pre-amp (Gaines Audio MP-1), and cassette deck (Carver TD1700). The microphone was placed at approximately a 45-deg angle and 15 cm away from the corner of the speaker's mouth, to prevent turbulence due to direct airflow from impinging on the microphone.

All recordings were sampled at 22 kHz (16-bit quantization, 11-kHz low-pass filter) on a Sun SPARCstation 5. All measurements were made using Entropics Systems' WAVES +/ESPS software. Fricative segmentation involved the simultaneous consultation of waveform and wideband spectrogram. Fricative onset was defined as the point at which high-frequency energy first appeared on the spectrogram and/or the point at which the number of zero crossings rapidly increased. Frication offset for voiceless fricatives was defined as the intensity minimum immediately preceding the onset of vowel periodicity. For voiced fricatives, the earliest pitch period exhibiting a change in the waveform from that seen throughout the initial frication was identified. The zero crossing of the preceding pitch period was then designated as the end of the voiced fricative (see Yeni-Komshian and Soli, 1981). Word duration was defined as the interval between fricative onset and the syllable-final /p/ release burst.

Spectral peak location of the fricatives was examined using a 40-ms full Hamming window placed in the middle of the frication noise. This larger window size yields better resolution in the frequency domain, at the expense of resolution in the temporal domain. Since fricatives are characterized by a relatively stationary articulatory configuration, the advantage of increased frequency resolution outweighs the disadvantage of decreased temporal resolution. A previous comparison of spectral properties of fricatives as measured at onset, midpoint, and offset of the frication noise showed that these properties are relatively stable throughout the noise portion, with high-frequency peaks more likely to emerge in the middle and end of the noise (Behrens and Blumstein, 1988a). Spectral peak estimation was based on spectra generated by means of FFT (fast Fourier transform) and LPC (linear predictive coding). For both FFT and LPC, a 40-ms full Hamming window was used, with a pre-emphasis factor of 98%. For LPC, 24 poles were used. LPC spectra were computed to examine if their peaks matched those of the FFT spectra. Spectral peak is defined here as the highest-amplitude peak of the FFT spectrum.

Spectral moments were computed following the procedures described by Forrest *et al.* (1988) with a few modifications. FFTs were calculated using a 40-ms full Hamming window (as compared to Forrest *et al.*'s 20-ms window) at four different locations in the fricative: onset, middle, and end, as well as centered over fricative offset. For example, the first window included the first 40 ms of the fricative, while the last window spanned the final 20 ms of the fricative and the first 20 ms of the following vowel. Each FFT was treated as a random probability distribution from which the first four moments were calculated. Moments were calculated from both linear and bark-transformed spectra. Only

moments based on linear spectra are reported here, since there was no substantial difference between them and bark-transformed spectra.

Locus equations were derived using the procedure described by Sussman and Shore (1996) for fricatives. For both voiced and voiceless fricatives, $F2$ was measured at vowel onset and midway in the vowel. Specifically, $F2$ at vowel onset was estimated by means of FFT spectra, with a 23.3-ms full Hamming window (similar to Sussman and Shore, 1996, and Fowler, 1994) starting at the first glottal pulse following cessation of the fricative. (These data were also used in the analysis of $F2$ onset.) Similarly, $F2$ at vowel nucleus was estimated by placing a 23.3-ms window at the vowel's midpoint. In the case of the diphthongized vowels /e/ and /o/, data points from the vowel offglide were excluded. In addition to FFT spectra, wideband spectrograms and LPC spectra were also consulted.

Root-mean-square (rms) amplitude in dB was measured for the entire noise portion of each fricative token. In order to normalize for intensity differences among speakers, a difference of fricative amplitude minus vowel amplitude (''normalized amplitude'') was calculated, where vowel amplitude was defined as rms amplitude (in dB) averaged over three consecutive pitch periods at the point of maximum vowel amplitude (see Behrens and Blumstein, 1988b).

Relative amplitude in dB was measured as described in Hedrick and Ohde (1993). Briefly, for the vowel, a discrete Fourier transform (DFT) was derived at vowel onset, using a 23.3-ms Hamming window. The amplitude (in dB) of the component at $F3$ for /s,z,ʃ,ʒ/ and at $F5$ for /f,v,θ,ð/ was measured. For the fricative, a DFT was then derived at the center of the fricative, using a 23.3-ms Hamming window. The amplitude (in dB) of the component in the same frequency region as that selected for the vowel was measured. Relative amplitude was then expressed as the difference between fricative amplitude and vowel amplitude.

B. Results

1. Spectral properties

a. Spectral peak location. A four-way analysis of variance (ANOVA) (place×voicing×vowel×gender) revealed a main effect for place of articulation [$F(3,2876)=1083.72$, $p<0.0001$; $\eta^2=0.512$]. Averaged across speakers, voicing, and vowel context, spectral peak location for the labiodentals was 7733 Hz, for dentals 7470 Hz, for alveolars 6839 Hz, and for palato-alveolars 3820 Hz. Spectral peak location thus decreases in frequency as place of articulation moves further back in the oral cavity. Bonferroni *post hoc* tests indicated that all four places of articulation were significantly different from each other in terms of spectral peak location ($p<0.003$ for the contrast between labiodentals and dentals, $p<0.0001$ for all other contrasts).

A main effect of voicing [$F(1,2876)=30.65$, $p<0.0001$; $\eta^2=0.01$] indicated that voiceless fricatives had spectral peaks at a significantly higher frequency (6612 Hz) than voiced fricatives (6310 Hz). A place×voicing interaction [$F(3,2876)=12.14$, $p<0.0001$; $\eta^2=0.012$] and subsequent *post hoc* tests revealed that the difference in spectral peak location between voiceless and voiced fricatives was carried by the nonsibilant fricatives. As shown in Fig. 1, while there was no difference between /s/ and /z/ and between /ʃ/ and /ʒ/, the differences in spectral peak between /θ/ and /ð/ (832 Hz) and between /f/ and /v/ (340 Hz) were significant.

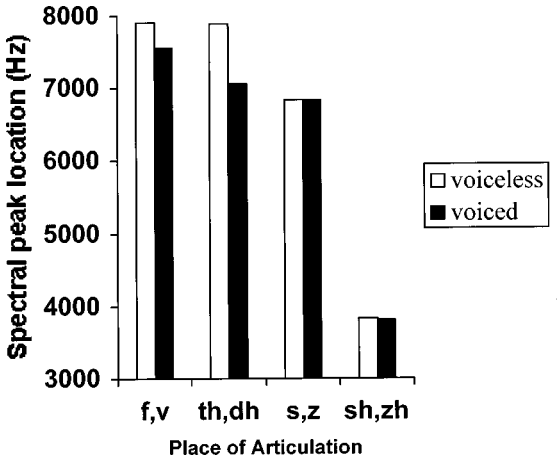


FIG. 1. Mean spectral peak location as a function of place of articulation and voicing (in Hz, averaged across vowels, and male and female speakers). Spectral peak location was computed over a 40-ms window placed in the middle of the fricative noise.

A main effect of gender [$F(1,2876)=154.15$, $p<0.0001$; $\eta^2=0.047$] indicated that, as expected, mean spectral peak location was significantly higher for female (6800 Hz) than for male (6122 Hz) speakers. A place×gender interaction [$F(3,2876)=34.61$, $p<0.0001$; $\eta^2=0.032$] and subsequent *post hoc* tests revealed that the pattern of males and females was not entirely parallel. As shown in Fig. 2, male speakers show a pattern in which spectral peak frequency decreases as place moves back; however, female speakers are different in that their dentals have a higher spectral peak frequency than their labiodentals.

Finally, no main effect was observed for vowel ($p>0.878$). A significant place×vowel interaction [$F(15,2876)=3.67$, $p<0.001$; $\eta^2=0.017$] and *post hoc* tests indicated that spectral peak location of only /s,z/ varied

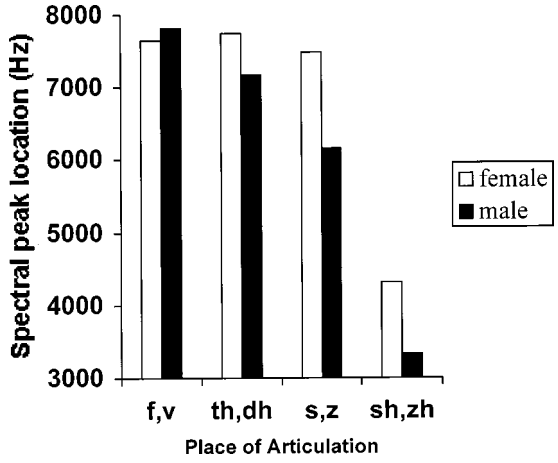


FIG. 2. Mean spectral peak location for male and female speakers as a function of place of articulation (in Hz, averaged across vowels, and voiced and voiceless tokens).

TABLE I. Mean spectral moment values for each place of articulation, averaged across speakers, window location, voiced and voiceless tokens, and vowel context.

Place of articulation	Spectral mean (Hz)	Variance (MHz)	Skewness	Kurtosis
/f,v/	5108	6.37	0.077	2.11
/T,D/	5137	6.19	−0.083	1.27
/s,z/	6133	2.92	−0.229	2.36
/S,Z/	4229	3.38	0.693	0.42

as a function of vowel context: spectral peak for /s,z/ was significantly lower in the context of the back-rounded vowels /o,u/.

b. Spectral moments. One-way ANOVAs were conducted for place, voicing, and gender across window locations with the four moments as dependent variables. For spectral mean, a main effect obtained for place of articulation [$F(3,11520)=488.16$, $p<0.0001$; $\eta^2=0.113$]. As shown in Table I, spectral mean was highest for /s,z/ (6133 Hz) and lowest for /ʃ,ʒ/ (4229 Hz), and this difference was significant ($p<0.0001$). Spectral mean values for /f,v/ (5108 Hz) and /θ,ð/ (5137 Hz) fell in between and were not significantly different from each other ($p>0.9$). For spectral variance, a main effect obtained for place of articulation [$F(3,11520)=1216.02$, $p<0.0001$; $\eta^2=0.241$]. Variance was low for the sibilant fricatives and high for the nonsibilants. Differences among all places were highly significant ($p<0.0001$) except that between /f,v/ and /θ,ð/ which was only marginally so ($p>0.066$). A main effect for skewness [$F(3,11520)=332.24$, $p<0.0001$; $\eta^2=0.080$] and subsequent *post hoc* tests revealed that skewness distinguished all four places of articulation ($p<0.0001$). Skewness was highest for /ʃ,ʒ/, indicating that the palato-alveolars had the strongest concentration of energy in the lower frequencies. Finally, there was a main effect for kurtosis [$F(3,11520)=90.69$, $p<0.0001$; $\eta^2=0.023$]. Kurtosis failed to distinguish /f,v/ from /s,z/ ($p>0.293$), both of which had high values indicating peaked spectra. All other comparisons were significant ($p<0.0001$).

A main effect was obtained for voice for all four moments. Effect size was rather small, with η^2 ranging from 0.001 for kurtosis to 0.069 for variance. Voiceless fricatives were characterized by higher values for spectral mean (5267 Hz), skewness (0.238), and kurtosis (1.70) than voiced fricatives (5036 Hz; −0.009; and 1.38, respectively). Thus, compared to voiced fricatives, the spectra of voiceless fricatives had a concentration of energy towards slightly lower frequencies and slightly better defined peaks. In addition, voiced fricatives had a significantly greater variance (5.56 MHz) than voiceless ones (3.87 MHz).

Finally, a main effect for gender indicated that females exhibited significantly higher values than males for spectral mean (5286 vs 5018 Hz), variance (4.9 vs 4.5 MHz), and kurtosis (1.64 vs 1.44), while female skewness values were significantly lower than those of males (0.084 vs 0.145). Thus, compared to males, the spectra of female speakers had clearer peaks and a concentration of energy towards higher frequencies. It must be noted that effect size was very small,

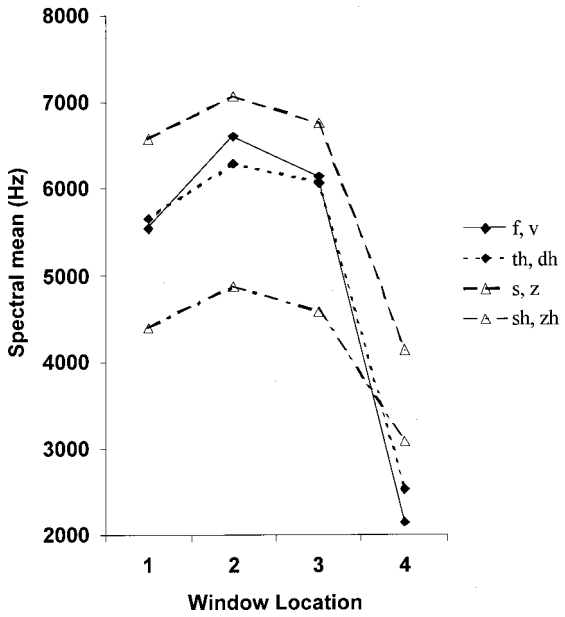


FIG. 3. Spectral mean (moment 1) in Hz (averaged across vowels, voiced and voiceless tokens, and male and female speakers), for each window location, as a function of place of articulation. Window locations 1, 2, and 3 refer to the first, middle, and last 40 ms of the fricative noise, respectively; window location 4 includes the final 20 ms of the fricative and the first 20 ms of the following vowel.

with η^2 ranging from 0.001 for skewness to 0.004 for spectral mean. A table with values for each moment at each window location for voiced and voiceless tokens and female and male speakers can be found in the Appendix.

In order to assess the importance of acoustic information at different positions in the speech signal, four-way ANOVAs (place×voicing×vowel×gender) and subsequent Bonferroni *post hoc* tests were conducted for each moment at each window location. Figures 3 through 6 show moment values for each place of articulation as a function of window location, for moments 1 through 4, respectively. Results of the statistical tests are summarized in Table II. This table shows the number of places of articulation differentiated by a given moment at a given window location. It is clear that spectral moments distinguish at least three places of articulation at all window locations, and four places in the majority of cases. All but two confusions involved a lack of differentiation between /f,v/ and /θ,ð/.

M1 (spectral mean) (Fig. 3) distinguishes all four places of articulation at the second and fourth window locations. In general, /s,z/ have the highest spectral mean, and /ʃ,ʒ/ the lowest. The nonsibilants' spectral means fall in between. M2 (variance) (Fig. 4) distinguishes all places at all but the second window location. Variance is low for the sibilant fricatives and high for the nonsibilants. M3 (skewness) (Fig. 5) distinguishes all places at all but the third window location. Skewness is always positive for /ʃ,ʒ/, indicating a concentration of energy in the lower frequencies. Skewness increases substantially at the fricative–vowel transition (window 4) for the nonsibilants, reflecting the predominance of low-frequency over high-frequency energy as the vowel begins. M4 (kurtosis) (Fig. 6) distinguishes all places at only the first window location. Kurtosis is highest for /s,z/, indi-

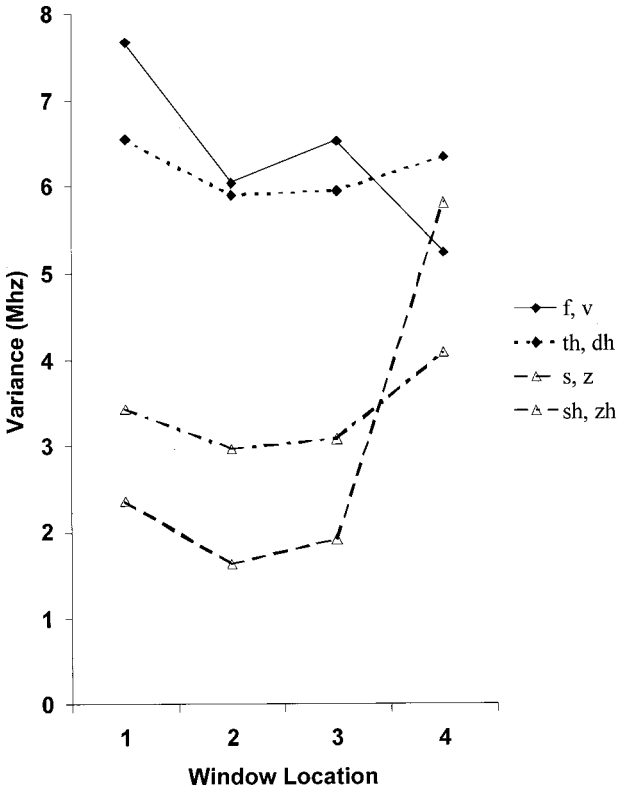


FIG. 4. Spectral variance (moment 2) in MHz (averaged across vowels, voiced and voiceless tokens, and male and female speakers), for each window location, as a function of place of articulation.

cating a spectrum with clearly defined peaks. Kurtosis yields the only confusions that do not involve /f,v/ and /θ,ð/; instead, /f,v/ and /ʃ,ʒ/ are nondistinct at fricative offset while /s,z/ and /ʃ,ʒ/ are not differentiated in the transition region between fricative and vowel. The effect for each moment is quite sizable at nearly every window location. For spectral mean, η^2 ranges from 0.296 to 0.387, for variance from

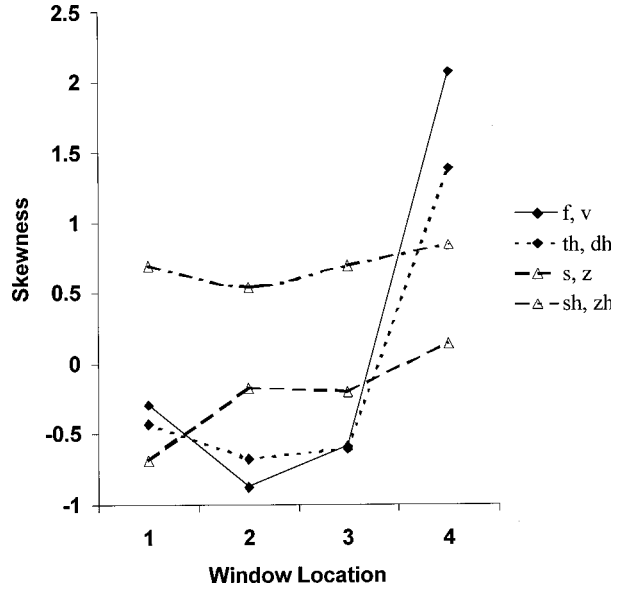


FIG. 5. Spectral skewness (moment 3; averaged across vowels, voiced and voiceless tokens, and male and female speakers), for each window location, as a function of place of articulation.

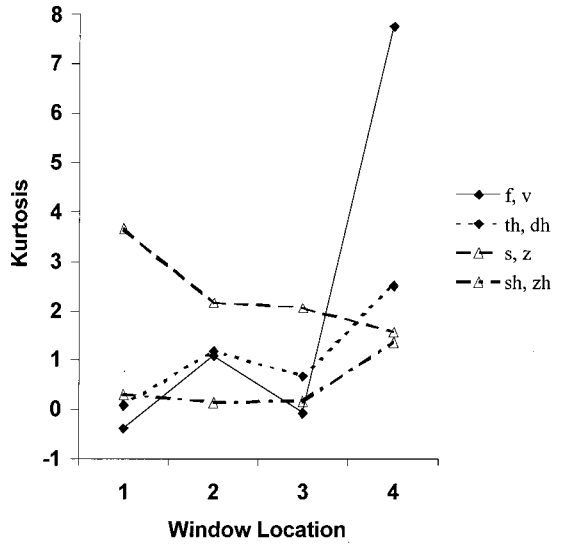


FIG. 6. Spectral kurtosis (moment 4; averaged across vowels, voiced and voiceless tokens, and male and female speakers), for each window location, as a function of place of articulation.

0.103 to 0.545, and for skewness from 0.321 to 0.380. Finally, effects were somewhat weaker for kurtosis, with η^2 ranging from 0.066 to 0.281.

2. Transition information

a. Locus equations. Following Sussman *et al.* (1991), slope and *y*-intercept values were derived for each place of articulation for each speaker, averaged across vowel context. Table III presents slope and *y*-intercept values for each place of articulation for females and males, averaged across all vowel contexts.[1] A two-way ANOVA (place×gender) for slope revealed a main effect for place of articulation [$F(3,72)=33.25$, $p<0.0001$; $\eta^2=0.581$]. *Post hoc* tests indicated that only the slope value for /f,v/ was significantly different from that of the other three places of articulation. For the *y* intercept, a main effect was observed for place [$F(3,72)=51.32$, $p<0.0001$; $\eta^2=0.681$], with subsequent *post hoc* tests revealing that, while *y*-intercept values were distinct for /f,v/ and /ʃ,ʒ/, they were not for /θ,ð/ and /s,z/. A main effect was also observed for gender [$F(1,72)=19.79$, $p<0.0001$; $\eta^2=0.216$], indicating that the *y* intercept was significantly higher for females (900 Hz) than for males (708 Hz).

b. F2 onset values. Table IV presents *F*2 onset values for each place of articulation, averaged across all speakers

TABLE II. Number of places of articulation (out of 4) distinguished by each moment for each window location. A score of 3 was in all but two cases due to confusion of /f,v/ with /θ,ð/. The exceptions were the confusion of /f,v/ with /ʃ,ʒ/ for kurtosis at fricative offset, and of /s,z/ with /ʃ,ʒ/ for kurtosis at the fricative–vowel transition.

	Window			
Moment	Onset	Middle	Offset	Transition
Spectral mean	3	4	3	4
Variance	4	3	4	4
Skewness	4	4	3	4
Kurtosis	4	3	3	3

TABLE III. Mean slope (k) and y intercept (c in Hz) (averaged across voiced and voiceless tokens and vowels) as a function of place of articulation and speaker gender.

	/f,v/		/θ,ð/		/s,z/		/ʃ,ʒ/	
	k	c	k	c	k	c	k	c
Females	0.766	413	0.530	940	0.501	1004	0.452	1242
Males	0.770	299	0.529	819	0.533	825	0.557	887
Mean	0.768	356	0.530	879	0.517	914	0.505	1065

and vowel contexts. A four-way ANOVA (place×voicing ×vowel×gender) revealed a main effect for place [$F(3,2876)=147.25$, $p<0.0001$; $\eta^2=0.133$]. $F2$ onset values generally increased as place of articulation moved further back in the vocal tract. However, Bonferroni *post hoc* tests indicated that the difference between /θ,ð/ and /s,z/ was not significant.

A main effect of vowel [$F(5,2876)=481.74$, $p<0.0001$; $\eta^2=0.456$] obtained: $F2$ onset was 2334 Hz in the context of /i/, 2010 Hz before /e/, 1820 Hz before /æ/, 1710 Hz before /u/, 1526 Hz before /o/, and 1512 Hz before /A/. *Post hoc* tests indicated that $F2$ onset values were higher for front vowels compared to back vowels and that $F2$ onset values significantly increased as a function of increasing vowel height. All differences among vowels were significant except that between /o/ and /A/. There was no main effect of voicing. A place×vowel interaction [$F(15,2876)=22.52$, $p<0.0001$; $\eta^2=0.105$] and *post hoc* tests revealed that while $F2$ onset differed significantly with each vowel for /f,v/ and /s,z/, the vowel context effects for /θ,ð/ and /ʃ,ʒ/ were restricted to /i,e/. A place×voicing interaction [$F(3,2876)=6.85$, $p<0.0001$; $\eta^2=0.007$] and *post hoc* tests revealed that while there was no difference in $F2$ onset between voiced and voiceless tokens of the labiodental, dental, and alveolar fricatives, $F2$ onset was significantly higher for /ʒ/ (2040 Hz) than for /ʃ/ (1925 Hz). Finally, as expected, there was a main effect for gender [$F(1,2876)=563.9$, $p<0.0001$; $\eta^2=0.164$]; $F2$ onset was significantly higher for females (1967 Hz) than for males (1689 Hz).

3. Amplitude

a. Overall amplitude. Table V shows mean noise amplitude, vowel amplitude, and the difference between the two (''normalized amplitude'') as a function of place of articulation. Using normalized amplitude as the dependent variable, a four-way ANOVA (place×voicing×vowel×gender) revealed a main effect for place [$F(3,2876)=1489.51$, $p<0.0001$; $\eta^2=0.591$]. Bonferroni *post hoc* tests indicated that all four places of articulation were significantly different from each other in terms of normalized amplitude.

TABLE IV. Mean $F2$-onset values (Hz) (averaged across voiced and voiceless tokens, and vowels) as a function of place of articulation and speaker gender.

	/f,v/	/θ,ð/	/s,z/	/ʃ,ʒ/
Females	1815	1969	1967	2115
Males	1509	1701	1697	1849
Mean	1661	1833	1832	1982

TABLE V. Mean noise amplitude, vowel amplitude (in dB, averaged across speakers and vowels), and normalized amplitude for each fricative. Normalized amplitude refers to noise amplitude minus vowel amplitude in dB. Mean normalized amplitude refers to normalized amplitude for each place of articulation.

Fricative	Noise amplitude	Vowel amplitude	Normalized amplitude	Mean norm. ampl.
/f/	55.7	76.5	−20.8	
				−17
/v/	63.2	76.3	−13.1	
/θ/	54.7	76.6	−21.9	
				−18
/ð/	62.7	76.7	−14.0	
/s/	64.9	75.9	−11.0	
				−10
/z/	67.7	76.7	−9.0	
/ʃ/	66.4	76.3	−9.9	
				−9
/ʒ/	68.2	76.5	−8.3	

A main effect of voicing [$F(1,2876)=1644.06$, $p<0.0001$; $\eta^2=0.347$] indicated that voiced fricatives (−15.9 dB) had a significantly smaller amplitude relative to the vowel than their voiceless counterparts (−11.1 dB). A main effect of vowel [$F(5,2876)=11.94$, $p<0.0001$; $\eta^2=0.019$] was obtained. The normalized amplitude preceding /o/ was −14 dB, /u/: −13.8 dB, /e/: −13.8 dB, /æ/: −13.6 dB, /ɑ/: −13 dB, /i/: −12.7 dB. Bonferroni *post hoc* tests indicated that only the amplitude difference for /i/ and /ɑ/ differed from that for all other vowels. There was no main effect of gender. Finally, a place×voicing interaction [$F(3,2876)=214.15$, $p<0.0001$; $\eta^2=0.172$] and *post hoc* tests indicated that the difference between voiced and voiceless fricatives was much greater for the nonsibilants than for the sibilants.

b. Relative amplitude. Figure 7 presents relative amplitude values for each place of articulation for voiced and voiceless tokens. A four-way ANOVA (place×voicing ×vowel×gender) revealed a main effect for place [$F(3,2876)=458.27$, $p<0.0001$; $\eta^2=0.308$]. Bonferroni *post hoc* tests indicated that all four places of articulation were significantly different ($p<0.0001$ for all comparisons).

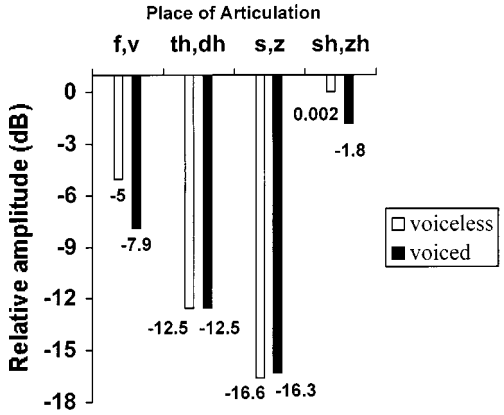


FIG. 7. Relative amplitude (dB) as a function of place of articulation and voicing (see Sec. II A 3 for calculation).

TABLE VI. Mean frication duration, total word duration (in ms, averaged across speakers and vowels), and normalized duration for each fricative. Normalized duration refers to the ratio of frication duration over word duration. Mean normalized duration refers to normalized duration for each place of articulation.

Fricative	Frication duration	Word duration	Normalized duration	Mean norm. dur.
/f/	166	395	0.420	
				0.333
/v/	80	326	0.245	
/θ/	163	393	0.415	
				0.340
/ð/	88	333	0.264	
/s/	178	406	0.438	
				0.382
/z/	118	362	0.326	
/ʃ/	178	397	0.448	
				0.393
/ʒ/	123	364	0.338	

A main effect of voicing [$F(1,2876)=14.03$, $p<0.0001$; $\eta^2=0.005$] indicated that voiceless fricatives (−8.5 dB) had a significantly greater relative amplitude than their voiced counterparts (−9.6 dB). A main effect of vowel [$F(5,2876)=6.36$, $p<0.0001$; $\eta^2=0.01$] was also obtained. Relative amplitude preceding /e/ was −10.4 dB, /o/: −9.1 dB, /i/, /æ/: −8.9 dB, /A/: −8.6 dB, /u/: −7.6 dB. Bonferroni *post hoc* tests indicated that only the relative amplitude for /e/ and /u/ differed from that for all other vowels. Finally, a main effect of gender [$F(1,2876)=28.73$, $p<0.0001$; $\eta^2=0.009$] indicated that relative amplitude values were smaller for females (−9.8 dB) than for males (−8.1 dB).

A place by vowel interaction [$F(15,2876)=4.95$, $p<0.0001$; $\eta^2=0.023$] revealed that while vowel-intrinsic differences in relative amplitude were similar across most places of articulation, labiodental /f,v/ deviated from this pattern, showing much lower values for the back vowels /ɑ,o,u/ as compared to the front vowels /i,e,æ/. A place by voicing interaction [$F(3,2876)=4.82$, $p<0.002$; $\eta^2=0.005$] was due to the fact that while there was no difference in relative amplitude between /θ/ and /ð/ or between /s/ and /z/, the difference in relative amplitude between /f/ and /v/ (2.9 dB) and that between /ʃ/ and /ʒ/ (1.8 dB) was significant. A place by gender interaction [$F(3,2876)=6.01$, $p<0.0001$; $\eta^2=0.006$] indicated that the gender difference in relative amplitude was most pronounced for /θ,ð/. Finally, a voicing by gender interaction [$F(1,2876)=13.74$, $p<0.0001$; $\eta^2=0.004$] was obtained because the voicing difference in relative amplitude was mostly due to the male speakers.

4. Noise duration

Table VI shows mean frication duration, word duration, and their ratio as a function of place of articulation. Analyses involving duration have typically focused on absolute frication duration (e.g., Behrens and Blumstein, 1988a). An initial four-way ANOVA (place×voicing×vowel×gender) with fricative duration as the dependent variable revealed a main effect for place [$F(3,2876)=327.69$, $p<0.0001$; $\eta^2=0.092$]. However, Bonferroni *post hoc* tests indicated that all four places were not significantly different but that noise duration of the nonsibilant fricatives was significantly shorter than that of the sibilant fricatives. Because absolute duration may vary as a function of speaking rate, a four-way ANOVA (place×voicing×vowel×gender) was conducted using ''normalized duration,'' defined as the ratio of fricative duration over word duration, as the dependent variable. A main effect for place [$F(3,2876)=236.56$, $p<0.0001$; $\eta^2=0.187$] and subsequent *post hoc* tests indicated that only the difference between /f,v/ and /θ,ð/ was not significant. All other comparisons were significant at the $p<0.0001$ level, except that between /s,z/ and /ʃ,ʒ/ ($p<0.001$). A main effect of voicing [$F(1,2876)=4547.30$, $p<0.0001$; $\eta^2=0.595$] indicated normalized duration was significantly greater for voiceless fricatives (0.429) than for voiced ones (0.293). A place×voicing interaction [$F(3,2876)=58.28$, $p<0.0001$; $\eta^2=0.053$] and *post hoc* tests indicated that the effect of voicing was more pronounced for the nonsibilants than for the sibilants. A main effect of gender [$F(1,2876)=66.32$, $p<0.0001$; $\eta^2=0.021$] indicated that fricatives produced by female speakers (0.351) had slightly smaller normalized durations than those produced by male speakers (0.368). Finally, a main effect of vowel [$F(5,2876)=138.04$, $p<0.0001$; $\eta^2=0.182$] was obtained. Bonferroni *post hoc* tests indicated that normalized duration decreased with decreasing vowel height: normalized duration preceding /i/ was 0.390, /u/: 0.400, /e/: 0.356, /o/: 0.357, /æ/: 0.324, and /ɑ/: 0.324. Differences between vowels of different heights were all significant ($p<0.0001$); differences between vowels of the same height were not significant (/i/ vs /u/: $p>0.098$; /e/ vs /o/ and /æ/ vs /ɑ/: $p>0.90$).

5. Discriminant analysis

Discriminant analysis was performed to evaluate the extent to which the acoustic parameters reported here could categorize the fricatives in terms of place of articulation. All acoustic parameters discussed above, except for locus equations because they are not a property of individual productions, were entered as predictors. For the moments, each moment at each window location was entered. A stepwise linear discriminant analysis was conducted with 21 predictors (spectral peak location, 4 moments×4 window locations, $F2$ onset, normalized amplitude, relative amplitude, and normalized duration). Classification results are based on the jackknife method, whereby each speaker in turn was used as the testing speaker with training being done on the 19 remaining speakers. Final classification scores were then averaged across the 20 testing speakers.

Classification scores for each place of articulation based on the jackknife method are shown in Table VII. Overall classification accuracy was 77%. While classification of all four places of articulation was significantly above chance, it was clearly better for the sibilants (88%) than for the nonsibilants (66%). Classification errors rarely crossed the sibilant/nonsibilant distinction. That is, labiodentals and dentals were mostly confused with each other, and the same was true of alveolars and palato-alveolars.

In order to assess the contribution of each predictor variable to the discriminant functions, the standardized canonical discriminant function coefficients were analyzed (Klecka,

TABLE VII. Predicted group membership (%) in terms of fricative place of articulation. Classification is based on a stepwise linear discriminant analysis with all acoustic measures as predictors (see the text). Bold percentages indicate correct classification rates. Overall correct classification was 77%.

	Predicted group membership			
	/f,v/	/θ,ð/	/s,z/	/ʃ,ʒ/
/f,v/	**68**	27	3	2
/T,D/	26	**64**	6	4
/s,z/	1	4	**85**	9
/S,Z/	4	0	5	**91**

1980). These coefficients suggested that spectral peak location, normalized amplitude, relative amplitude, and spectral mean at fricative onset and midpoint were the main parameters used for fricative classification. A subsequent discriminant analysis with only those five predictors yielded an overall classification rate of 69%. Exclusion of the spectral mean at onset and midpoint only slightly decreased classification accuracy to 67%. Combinations of only two predictors yielded substantially lower rates, below 60% accuracy. Overall, then, spectral peak location, normalized amplitude, and relative amplitude served to distinguish the fricatives in terms of place of articulation with reasonable accuracy. Classification rates for this analysis were as follows: /f,v/: 53%, /θ,ð/: 48%, /s,z/: 81%, and /ʃ,ʒ/: 88%.

C. Discussion

The present results from 20 speakers indicate that spectral and amplitudinal information provide the most critical information to place of articulation in fricatives. In agreement with previous research on spectral properties of the frication noise (e.g., Hughes and Halle, 1956; Strevens, 1960; Heinz and Stevens, 1961; Jassem, 1965; Shadle, 1990; Behrens and Blumstein, 1988a), spectral peak location distinguishes sibilants from nonsibilants, and alveolar /s,z/ from palato-alveolar /ʃ,ʒ/. Importantly, however, the present results indicate that spectral peak location also distinguished /f,v/ from /θ,ð/. Thus, contrary to previous reports, spectral peak location does distinguish all four places of articulation.

Spectral moments also served to distinguish all four places of articulation. If the success of a moment is measured in terms of the number of places it distinguished at each location in the fricative, m2 (variance) and m3 (skewness) performed best (see Table II). Across moments, a comparison of window locations suggests that window locations 1 and 4 (noise onset and fricative–vowel transition region, respectively) contain the most distinctive information (see Table II).

Few studies report all four moment values or values for all fricatives. Most studies have focused on the spectral mean of /s/ and /ʃ/ (e.g., Nittrouer *et al.*, 1989; Baum and McNutt, 1990; Waldstein and Baum, 1991). Interestingly, those studies that did include more moments usually excluded spectral variance, perhaps because Forrest *et al.* (1988) excluded it from their original analysis since it did not appear to distinguish among any of the obstruents in their study. The fact that variance turns out to be a robust cue to place in the present study may be the result of sampling a larger and more representative number of speakers and tokens, as compared to the rather small database of Forrest *et al.* (1988), consisting of only five target words and no voiced fricatives.

Generally, in those cases in which direct comparisons could be made, there is good agreement in terms of spectral mean between the present study and previous research (Tomiak, 1990; Nittrouer, 1995; Avery and Liss, 1996; McFarland *et al.*, 1996; Tjaden and Turner, 1996). In terms of spectral variance, there is good agreement with Tomiak (1990), the only other study reporting values for the second moment.

With respect to the third moment, the present finding of a negative skewness for /s/ and a positive skewness for /ʃ/ is supported by previous findings by Nittrouer (1995) and McFarland *et al.* (1996) but differs from Tomiak (1990) and Avery and Liss (1996), who reported a greater positive skewness for /s/ than for /ʃ/. Shadle and Mair (1996) did report that variance was a more reliable indicator of fricative place than skewness, although the authors report no overall analysis of place of articulation for these moments and only one female and one male speaker were included. Finally, our finding of a large positive kurtosis for /s/ and a small positive or negative kurtosis for /ʃ/ is in agreement with Tomiak (1990), Nittrouer (1995), Avery and Liss (1996), and McFarland *et al.* (1996). In general, the present data clearly show that four places of articulation were distinguished by most moments at most window locations.

Both normalized and relative amplitude properties were also found to be consistent cues to fricative place of articulation. In terms of normalized amplitude, sibilant fricatives had a greater noise amplitude than nonsibilants; moreover, within the group of sibilants, palato-alveolar /ʃ,ʒ/ had a greater noise amplitude than alveolar /s,z/, while for the nonsibilants labiodental /f,v/ had a greater amplitude than interdental /θ,ð/. Previous research supports the role of noise amplitude in the sibilant/nonsibilant distinction (e.g., Strevens, 1960; Behrens and Blumstein, 1988a, b). In particular, in their study of /f,θ,s,ʃ/, Behrens and Blumstein (1988a, b) reported overall amplitude differences of similar magnitude as the present study. However, contrary to these studies, the present study also indicates that normalized amplitude can distinguish place of articulation *within* these two groups. One of the three speakers analyzed by Behrens and Blumstein (1988a) showed a significantly greater amplitude for /f/ compared to /θ/, suggesting that the difference in the present study may be due to our larger sampling of speakers and tokens.

Relative amplitude also distinguished all four places of articulation. Relative amplitude was small for the palato-alveolars, indicating that /ʃ,ʒ/ has a major concentration of energy in the region corresponding to $F3$ of the following vowel. For the other places, relative amplitude was seen to decrease as place moved further back in the oral cavity. Fricative amplitude in the $F5$ region is smaller for /θ,ð/ than /f,v/. In addition, the large difference between fricative and vowel amplitude in the $F3$ region for /s,z/ supports the notion that these fricatives have their major energy in a frequency region well above $F3$. The present findings are qualitatively in line with those of Hedrick and Ohde (1993), who also reported a much greater relative amplitude for /ʃ/ than

for /s/ for their speaker. The present value for /s,z/ (−16.5 dB) is very similar to that used by Hedrick and Ohde (−17 dB), while that for /ʃ,ʒ/ (−0.9 dB) is much lower than theirs (16 dB). As mentioned previously, research on relative amplitude has exclusively focused on perception, which makes it impossible to compare the present findings to earlier work in any detail. However, the present acoustic data are corroborated by perceptual data on relative amplitude. Perceptually, the crossover boundary between /s/ and /ʃ/ has been shown to correspond to a relative amplitude of approximately −7.5 dB (Hedrick and Ohde, 1993), which is also halfway in between the relative amplitude measurements for /s/ and /ʃ/ reported here.

A comparison of the results from ANOVA and discriminant analysis reveals a high degree of agreement. Using η^2 to select those acoustic parameters from the ANOVAs that contribute most to distinguishing all four places of articulation, normalized amplitude ($\eta^2=0.591$), spectral peak location ($\eta^2=0.512$), relative amplitude ($\eta^2=0.308$), and spectral variance (M2) ($\eta^2=0.287$) were identified as the primary contributors. Results from the discriminant analyses also identified three of these parameters (spectral peak location as well as normalized and relative amplitude) as the strongest predictors of group membership.

In the present study, a number of measures were shown not to distinguish fricative place of articulation. These include *F*2 transition properties and noise duration. Properties specific to the *F*2 transition failed to distinguish all fricative places of articulation. Both the combination of slope and intercept values of the locus equations and the *F*2 onset values could only single out the labiodental and palato-alveolar fricatives. As for normalized noise duration, sibilant fricatives were longer than nonsibilants, supporting similar findings by Behrens and Blumstein (1988a) based on absolute noise duration. In addition, voiceless fricatives were substantially longer than their voiced counterparts (see also Baum and Blumstein, 1987; Behrens and Blumstein, 1988a; Crystal and House, 1988; Jongman, 1989). A new finding in the current study is that normalized duration also distinguished /s,z/ from /ʃ,ʒ/, which may be due to the use of normalized duration rather than absolute duration. However, even normalized duration failed to distinguish /f,v/ from /θ,ð/. Given the present findings with other parameters and the extent of the present database, it must be concluded that *F*2 transition properties and noise duration do not reliably distinguish place of articulation in fricatives.

In sum, the present study indicates that several acoustic properties serve to distinguish all four places of fricative articulation. These properties include both spectral (spectral peak location, spectral moments) and amplitudinal (normalized and relative amplitude) parameters, as well as both static (spectral peak location, spectral moments, normalized amplitude) and dynamic (relative amplitude) information. This finding suggests that, contrary to earlier reports, acoustic properties can provide robust information about all four places of articulation, despite variation in speaker, vowel context, and voicing. Future research will need to address the extent to which the properties identified here contribute to perception of place of articulation in fricatives.

ACKNOWLEDGMENTS

Portions of this research were conducted while the first author was at Cornell University. This research was supported (in part) by Research Grant Number 1 R29 DC 02537-01A1 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health. The authors thank Scott Gargash, Eric Evans, Michelle Spence, and Yue Wang for assistance, and Anders Lofqvist, Marija Tabain, Joan Sereno, Roel Smits, and an anonymous reviewer for comments on a previous version of this paper. Portions of this research were presented at the 135th meeting of the Acoustical Society of America and at the 1999 meeting of the American Speech Language and Hearing Association.

APPENDIX

Table AI gives values of the four spectral moments for each window location, as a function of voicing and speaker gender. Moments 1, 2, 3, and 4 refer to spectral mean (Hz), variance (MHz), skewness, and kurtosis, respectively. Window locations 1, 2, and 3 refer to the first, middle, and final 40 ms of the frication noise, respectively. Window location 4 refers to a window spanning the last 20 ms of the fricative and the first 20 ms of the following vowel.

TABLE AI. Values of the four spectral moments as a function of voicing and speaker gender.

		Females				Males			
		Window location				Window location			
Moment		1	2	3	4	1	2	3	4
1	voiceless	6149	6858	6320	2457	5822	6426	5862	2244
	voiced	5230	5883	5763	3629	4957	5652	5606	3573
2	voiceless	3.77	3.10	3.67	5.37	3.92	2.94	3.70	4.53
	voiced	6.26	5.68	5.44	5.93	6.06	4.80	4.67	5.64
3	voiceless	−0.1882	−0.3798	−0.2111	1.5576	−0.1064	−0.3139	−0.1081	1.6543
	voiced	−0.2624	−0.2580	−0.1906	0.6060	−0.1600	−0.2337	−0.2026	0.6286
4	voiceless	0.6238	0.9031	0.6125	4.619	0.2943	1.0144	0.6045	4.9541
	voiced	2.2613	1.0994	0.7272	2.3101	1.1438	1.2629	0.9329	1.3235

[1]These data were previously reported in a brief commentary (Jongman, 1998). They are repeated here in more detail and for the sake of completeness.

Avery, J. D., and Liss, J. M. (**1996**). ''Acoustic characteristics of less-masculine-sounding male speech,'' J. Acoust. Soc. Am. **99**, 3738–3748.

Baum, S. R., and Blumstein, S. E. (**1987**). ''Preliminary observations on the use of duration as a cue to syllable-initial fricative consonant voicing in English,'' J. Acoust. Soc. Am. **82**, 1073–1077.

Baum, S. R., and McNutt, J. C. (**1990**). ''An acoustic analysis of frontal misarticulation of /s/ in children,'' J. Phonetics **18**, 51–63.

Behrens, S. J., and Blumstein, S. E. (**1988a**). ''Acoustic characteristics of English voiceless fricatives: A descriptive analysis,'' J. Phonetics **16**, 295–298.

Behrens, S. J., and Blumstein, S. E. (**1988b**). ''On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants,'' J. Acoust. Soc. Am. **84**, 861–867.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., and Gerstman, L. J. (**1952**). ''Some experiments on the perception of synthetic speech sounds,'' J. Acoust. Soc. Am. **24**, 597–606.

Crystal, T., and House, A. (**1988**). ''Segmental durations in connected-speech signals: Current results,'' J. Acoust. Soc. Am. **83**, 1553–1573.

Delattre, P. C., Liberman, A. M., and Cooper, F. S. (**1955**). ''Acoustic loci and transitional cues for consonants,'' J. Acoust. Soc. Am. **27**, 769–773.

Evers, V., Reetz, H., and Lahiri, A. (**1998**). ''Crosslinguistic acoustic categorization of sibilants independent of phonological status,'' J. Phonetics **26**, 345–370.

Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (**1988**). ''Statistical analysis of word-initial voiceless obstruents: Preliminary data,'' J. Acoust. Soc. Am. **84**, 115–124.

Fowler, C. A. (**1994**). ''Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation,'' Percept. Psychophys. **55**, 597–611.

Hedrick, M. (**1997**). ''Effect of acoustic cues on labeling fricatives and affricates,'' J. Speech Lang. Hear. Res. **40**, 925–938.

Hedrick, M. S., and Carney, A. E. (**1997**). ''Effect of relative amplitude and formant transitions on perception of place of articulation by adult listeners with cochlear implants,'' J. Speech Lang. Hear. Res. **40**, 1445–1457.

Hedrick, M. S., and Ohde, R. N. (**1993**). ''Effect of relative amplitude of frication on perception of place of articulation,'' J. Acoust. Soc. Am. **94**, 2005–2027.

Heinz, J. M., and Stevens, K. N. (**1961**). ''On the properties of voiceless fricative consonants,'' J. Acoust. Soc. Am. **33**, 589–596.

Hughes, G. W., and Halle, M. (**1956**). ''Spectral properties of fricative consonants,'' J. Acoust. Soc. Am. **28**, 303–310.

Jassem, W. (**1965**). ''Formants of fricative consonants,'' Lang. Speech **8**, 1–16.

Jongman, A. (**1989**). ''Duration of fricative noise required for identification of English fricatives,'' J. Acoust. Soc. Am. **85**, 1718–1725.

Jongman, A. (**1998**). ''Are locus equations sufficient or necessary for obstruent perception?,'' Behav. Brain Sci. **21**, 271–272.

Kewley-Port, D. (**1982**). ''Measurement of formant transitions in naturally produced stop consonant–vowel syllables,'' J. Acoust. Soc. Am. **72**, 379–389.

Kewley-Port, D. (**1983**). ''Time-varying features as correlates of place of articulation in stop consonants,'' J. Acoust. Soc. Am. **73**, 322–335.

Klecka, W. (**1980**). *Discriminant Analysis*, Sage University Paper Series, Quantitative Applications in the Social Sciences, Series no. 07-019 (Sage, Beverly Hills, CA and London).

Ladefoged, P. (**1982**). *A Course in Phonetics* (Harcourt Brace Jovanovich College Publishers, New York).

Lahiri, A., Gewirth, L., and Blumstein, S. E. (**1984**). ''A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: Evidence from a cross-language study,'' J. Acoust. Soc. Am. **76**, 391–404.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (**1967**). ''Perception of the speech code,'' Psychol. Rev. **74**, 431–461.

Mann, V. A., and Repp, B. H. (**1980**). ''Influence of vocalic context on perception of the [S]–[s] distinction,'' Percept. Psychophys. **28**, 213–228.

McFarland, D. H., Baum, S. R., and Chabot, C. (**1996**). ''Speech compensation to structural modifications of the oral cavity,'' J. Acoust. Soc. Am. **100**, 1093–1104.

Newell, K. M., and Hancock, P. A. (**1984**). ''Forgotten moments: A note on skewness and kurtosis as influential factors in inferences extrapolated from response distributions,'' J. Motor Behav. **16**, 320–335.

Nittrouer, S. (**1992**). ''Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries,'' J. Phonetics **20**, 351–382.

Nittrouer, S. (**1995**). ''Children learn separate aspects of speech production at different rates: Evidence from spectral moments,'' J. Acoust. Soc. Am. **97**, 520–530.

Nittrouer, S., Studdert-Kennedy, M., and McGowan, R. S. (**1989**). ''The emergence of phonetic segments: evidence from the spectral structure of fricative-vowel syllables spoken by children and adults,'' J. Speech Lang. Hear. Res. **32**, 120–132.

Pike, K. (**1943**). *Phonetics* (University of Michigan Press, Ann Arbor).

Pisoni, D., and Luce, P. (**1986**). ''Speech perception: Research, theory, and the principal issues,'' in *Pattern Recognition by Humans and Machines*, edited by E. Schwab and H. Nusbaum (Academic, New York), pp. 1–50.

Recasens, D. (**1985**). ''Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences,'' Lang. Speech **28**, 97–114.

Schatz, C. D. (**1954**). ''The role of context in the perception of stops,'' Language **30**, 47–56.

Shadle, C. H. (**1990**). ''Articulatory-acoustic relationships in fricative consonants,'' in *Speech Production and Speech Modeling*, edited by W. Hardrastle and A. Marchal (Kluwer, Dordrecht, 1990), pp. 187–209.

Shadle, C. H., and Mair, S. J. (**1996**). ''Quantifying spectral characteristics of fricatives,'' Proc. Int. Conf. Spoken Lang. Proc. (ICSLP), pp. 1521–1524 (unpublished).

Soli, S. D. (**1981**). ''Second formants in fricatives: Acoustic consequences of fricative–vowel coarticulation,'' J. Acoust. Soc. Am. **70**, 976–984.

Stevens, K. N. (**1971**). ''Airflow and turbulence for noise for fricative and stop consonants: Static considerations,'' J. Acoust. Soc. Am. **50**, 1182–1192.

Stevens, K. N. (**1985**). ''Evidence for the role of acoustic boundaries in the perception of speech sounds,'' in *Phonetic Linguistics*, edited by V. A. Fromkin (Academic, New York), pp. 243–255.

Stevens, K. N. (**1998**). *Acoustic Phonetics* (The MIT Press, Cambridge, MA).

Stevens, K. N., and Blumstein, S. E. (**1981**). ''The search for invariant acoustic correlates of phonetic features,'' in *Perspectives of the Study of Speech*, edited by P. D. Eimas and J. L. Miller (Erlbaum, Hillsdale, NJ).

Strevens, P. (**1960**). ''Spectra of fricative noise in human speech,'' Lang. Speech **3**, 32–49.

Sussman, H. M. (**1994**). ''The phonological reality of locus equations across manner class distinctions: Preliminary observations,'' Phonetica **51**, 119–131.

Sussman, H. M., McCaffrey, H. A., and Matthews, S. A. (**1991**). ''An investigation of locus equations as a source of relational invariance for stop place categorization,'' J. Acoust. Soc. Am. **90**, 1309–1325.

Sussman, H. M., and Shore, J. (**1996**). ''Locus equations as phonetic descriptors of consonantal place of articulation,'' Percept. Psychophys. **58**, 936–946.

Tabain, M. (**1998**). ''Nonsibilant fricatives in English: Spectral information above 10 kHz,'' Phonetica **55**, 107–130.

Tjaden, K., and Turner, G. S. (**1997**). ''Spectral properties of fricatives in Amyotrophic Lateral Sclerosis,'' J. Speech Lang. Hear. Res. **40**, 1358–1372.

Tomiak, G. R. (**1990**). ''An acoustic and perceptual analysis of the spectral moments invariant with voiceless fricative obstruents,'' Doctoral dissertation, SUNY Buffalo.

Waldstein, R. S., and Baum, S. R. (**1991**). ''Anticipatory coarticulation in the speech of profoundly hearing-impaired and normally hearing children,'' J. Speech Lang. Hear. Res. **34**, 1276–1285.

Whalen, D. (**1981**). ''Effects of vocalic formant transitions and vowel quality on the English /s–S/ boundary,'' J. Acoust. Soc. Am. **69**, 275–282.

Wilde, L. (**1993**). ''Inferring articulatory movements from acoustic properties at fricative vowel boundaries,'' J. Acoust. Soc. Am. **94**, 1881.

Yeni-Komshian, B., and Soli, S. (**1981**). ''Recognition of vowels from information in fricatives: Perceptual evidence of fricative–vowel coarticulation,'' J. Acoust. Soc. Am. **70**, 966–975.

Yeou, M. (**1997**). ''Locus equations and the degree of coarticulation of Arabic consonants,'' Phonetica **54**, 187–202.

Intelligibility of bandpass speech: Effects of truncation or removal of transition bands

Richard M. Warren,[a] James A. Bashford, Jr.,[b] and Peter W. Lenz[c]
Department of Psychology, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, Wisconsin 53201-0413

(Received 10 May 2000; revised 24 May 2000; accepted 25 May 2000)

An intelligibility of over 90% was reported for keywords in ''everyday'' 1/3-octave sentences centered on 1500 Hz and having steep transition band slopes of 100 dB/octave [Warren *et al.*, Percept. Psychophys. **57**, 175–182 (1995)]. A subsequent study by Warren and Bashford [J. Acoust. Soc. Am. **106**, L47–L52 (1999)] found that it was not the 1/3-octave passband, but the transition bands that were chiefly responsible for this high intelligibility: When the passband and transition bands were segregated using filter slopes of 1000 dB/octave, the isolated passband had an intelligibility score of only 24%, while the pair of transition bands had a score of over 80%. In the present study, experiment 1 examined the distribution of information along the transition bands' slopes by truncation at graded downpoints: Truncation at downpoints of 40 dB or more produced no significant change in intelligibility. Experiment 2 closed the gap separating the transition bands so that their slopes intersected at 1500 Hz. This triangular band had a negligible passband (as defined conventionally by 3-dB downpoints) and an intelligibility score of 60%; truncation at downpoints of 50 dB or more produced no significant change in intelligibility. Experiment 3 determined the intelligibilities of rectangular bands (1000-dB/octave slopes) centered on 1500 Hz. Their bandwidths ranged from 3 to 12 semitones in 1-semitone steps, resulting in intelligibility scores increasing monotonically from 14% to 94%. Calculations based upon experiments 2 and 3 showed that the triangular band truncated at 30-dB downpoints had half the intelligibility of a rectangular band having the same frequency range.
[S0001-4966(00)01009-2]

PACS numbers: 43.71.Es, 43.71.An [KRK]

INTRODUCTION

As vocal productions become more complex and approximate normal discourse, identification becomes more accurate and resistant to distortions, masking, and deletions. At the level of sentences, a narrow spectral band can be sufficient to achieve high intelligibility. Warren *et al.* (1995) investigated the extent to which keywords in narrow-band sentences could be identified when they lacked spectral components required for the accurate identification of isolated phonemes and words. The stimuli chosen for this study had been designed to represent ''everyday American speech'' by Silverman and Hirsh (1955). When listeners heard 1/3-octave bands centered on 1500 Hz with steep slopes of 100 dB/octave, it was found that intelligibility was well over 90%. Stickney and Assmann (1997) verified the high intelligibility of narrow-band sentences, and also demonstrated that intelligibility dropped when the keywords in the sentences used for scoring had a lower contextual probability. When the narrow-band speech consisted of isolated words, a further large drop in performance occurred.

In order to examine the basis for the high intelligibility of words in 1/3-octave sentences, Warren and Bashford (1999) used the same digitally recorded stimuli employed in the earlier Warren *et al.* study. It was found that the average score of over 90% correct identification for keywords was due not to the nominal 1/3-octave passband, but rather to the flanking 100-dB/octave transition bands. This was determined by separating the passband (as defined conventionally by the 3-dB downpoints) and the pair of transition bands using 2000-order finite impulse response (FIR) filtering that produced nearly vertical slopes of approximately 1000 dB/octave. The isolated rectangular 1/3-octave passband had an intelligibility of only 24%, while the isolated transition bands (two right triangles separated by a 1/3-octave gap) had a much higher intelligibility score of 83% despite their severe spectral tilt.

The contrast between the low intelligibility of the passband and the remarkably high intelligibility of the steep transition bands raised several new questions that were addressed by the present study, using the same digitally recorded 1/3-octave sentences as the earlier experiments.

I. OVERALL PLAN AND RATIONALE

Experiment 1 examined the distribution of functional information along the transition bands by truncating the slopes at graded downpoints using 2000-order FIR filtering. In order to avoid ceiling effects that would obscure the loss of information, the passband was excised, leaving the isolated transition bands with a below-ceiling intelligibility score of approximately 80% prior to truncation.

Since the pair of transition bands employed in experiment 1 were separated by a 1/3-octave gap, the speech spectrum being sampled was discontinuous. It has been reported that two widely separated narrow bands of the everyday sen-

[a]Electronic mail: rmwarren@csd.uwm.edu
[b]Electronic mail: bashford@csd.uwm.edu
[c]Electronic mail: plenz@csd.uwm.edu

tences, when presented together, have intelligibility scores well above the scores for each of the bands heard separately (Warren *et al.*, 1995). To avoid such interactive enhancement associated with dual bands, experiment 2 closed the gap separating the transition bands, creating a single band consisting of slopes of 100 dB/octave intersecting at 1500 Hz. This triangular band had a negligible bandwidth, as defined conventionally by the 3-dB downpoints along the slopes. Preliminary observations had indicated that closing the 1/3-octave gap reduced intelligibility, but that the pair of intersecting slopes nevertheless had a considerably higher intelligibility than the 1/3-octave rectangular passband having the same center frequency. In experiment 2, the slopes of the triangular band were truncated at graded downpoints using the 2000-order FIR filtering as in experiment 1, and the intelligibilities were determined for various bandwidths (defined in this case as the range between the truncation downpoints) for this band consisting solely of 100-dB/octave slopes.

In experiment 3, the intelligibility was determined for ten bands having bandwidths ranging from 3 to 12 semitones in 1-semitone steps. These bands were each centered logarithmically on 1500 Hz, and were truncated at the passband borders with slopes of 1000 dB/octave. While these rectangular bands lack appreciable transition bands, the triangular bands lack appreciable passbands. By comparing the intelligibilities of truncated triangular bands (as determined in experiment 2) with the intelligibilities of rectangular bands extending over the same frequency range (as determined in experiment 3), we can determine the extent to which the listeners were able to compensate for the 100-dB/octave distortion of the normal spectral profile. In addition, the rectangular bands in experiment 3 provide a measure of the passband bandwidth needed to achieve an intelligibility comparable to the near ceiling value obtained with the intact 1/3-octave band (passband plus 100-dB/octave transition bands) reported previously for the same sentences.

II. SUBJECTS

The 60 listeners (20 in each of the three experiments) were native English speakers with no known hearing problems, who were recruited from the introductory Psychology courses at the University of Wisconsin-Milwaukee. They received course credit for their participation.

III. STIMULUS PROCESSING CID (CENTRAL INSTITUTE FOR THE DEAF)

The filtered sentences employed in this study consisted of the 100 (10 lists of 10) ''everyday'' sentences (Silverman and Hirsh, 1955). They contain 500 keywords that are used for scoring of intelligibility. Another 25 sentences served as practice stimuli and were taken from the high-predictability sublist of the revised SPIN (speech perception in noise) test (Kalikow *et al.*, 1977; Bilger *et al.*, 1984). All sentences had been recorded in digital form (44.1-kHz sampling, 16-bit quantization), and were used previously to study the intelligibility of 1/3-octave sentences (Warren *et al.*, 1995; Warren and Bashford, 1999). They were produced by a male speaker with no evident regional accent, who had an average voicing frequency of about 100 Hz.

Prior to preparation of the filtered stimuli, all sentences were transduced by a Sennheiser HD 250 Linear II headphone. The slow-rms peak level of each sentence was matched to within ± 0.2 dBA using an artificial ear in conjunction with a Brüel & Kjaer Model 2230 digital sound-level meter.

The two types of filtering described by Warren and Bashford (1999) were used in this study. Slopes of approximately 1000 dB/octave were produced using MATLAB function *fir 1* with order $n = 2000$, and when slopes of 100-dB/octave were required, they were produced by low-pass followed by high-pass filtering using MATLAB function *ellip* with order $n = 7$.

IV. PROCEDURE

A. General procedure

A repeated measures design was used, with the 10 CID lists presented pseudorandomly in a balanced order. Before receiving each stimulus condition, listeners were familiarized with the effects of filtering by listening to a set of SPIN sentences (five for each of the five conditions in experiments 1 and 2, and two for each of the ten conditions in experiment 3). These practice sentences were first presented broadband and then filtered in the same manner as the test sentences that followed. Listeners were tested individually while seated with the experimenter in an audiometric chamber. The stimuli were delivered diotically through Sennheiser HD 250 Linear II headphones. Subjects were told to repeat each sentence as accurately as possible, and to guess if unsure. The number of keywords reported correctly for each condition was recorded by the experimenter. Testing occurred in a single session lasting approximately 30 min.

B. Experiment 1: Truncation of isolated transition bands

The levels of the transition bands were the same as those employed by Warren and Bashford (1999), and were based upon a slow-rms peak level of 75 dB for the intact 1/3-octave sentences. This resulted in an average peak level of 69 dB for the pair of transition bands as measured after excision of the 1/3-octave passband. Listeners received 20 of the 100 sentences in each of the five conditions. These consisted of the intact transition bands with slopes of 100 dB/octave plus four truncated derivatives having 1000-dB/octave cutoffs at downpoints along both slopes of -30, -40, -50, and -60 dB, respectively (see Fig. 1). All truncations had a negligible effect on overall levels (less than 0.1 dB). The spectra of the stimuli are shown in Fig. 1.

C. Experiment 2: Truncation of triangular bands

The triangular band sentences were produced by setting the cutoff frequencies of both the low-pass and high-pass 100-dB/octave filters at 1500 Hz. As in experiment 1, in addition to the intact transition bands, four additional stimuli were employed having nearly vertical truncations of the

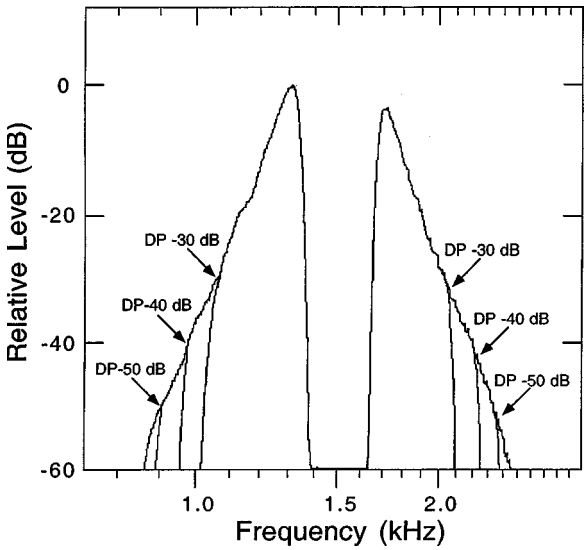


FIG. 1. Long-term averaged spectra based upon the 100 1/3-octave CID sentences having the nominal passband removed through steep FIR filtering. The isolated transition bands, having slopes of about 100-dB/octave, were presented either intact or truncated as shown with cutoffs at downpoints (DPs) along slopes of −30, −40, −50, or −60 dB, relative to the peak of the respective transition band. The plots were generated offline via discrete Fourier transform (DFT) with linear averaging of spectra obtained using a 8192-point (186 ms, 5-Hz resolution) Blackman window across the entire set of digitally recorded sentences. Metric Halo Laboratories, Inc. produced the DFT software.

slopes occurring at downpoints of −30, −40, −50, and −60 dB, respectively. The level of all five stimulus sets was 75 dB (as in experiment 1, truncation had no appreciable effect upon the overall level). The spectra of the stimuli are shown in Fig. 2.

D. Experiment 3: Graded widths of rectangular bands

Listeners heard ten sets of ten rectangular-band sentences having slopes of 1000 dB/octave. All 100 sentences

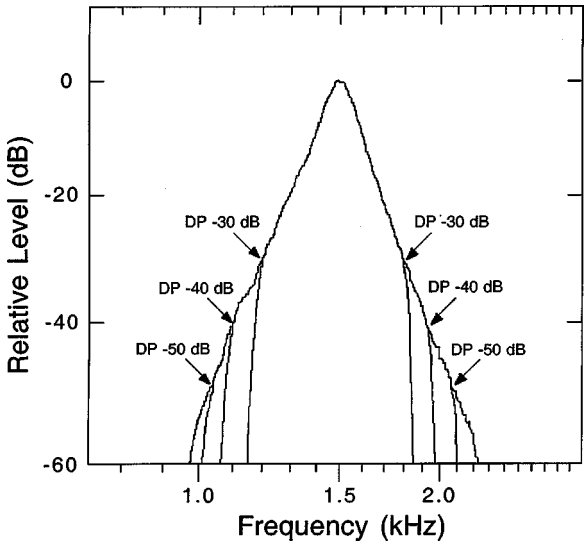


FIG. 2. Long-term averaged spectra of the 100 ''triangular band'' CID sentences produced with a common low-pass and high-pass cutoff of 1500 Hz for the 100-dB/octave filters. These contiguous transition bands were presented either intact or truncated with cutoffs at downpoints along slopes of −30, −40, −50, or −60 dB, relative to the peak of the triangular band. The plots were generated offline as described for Fig. 1.

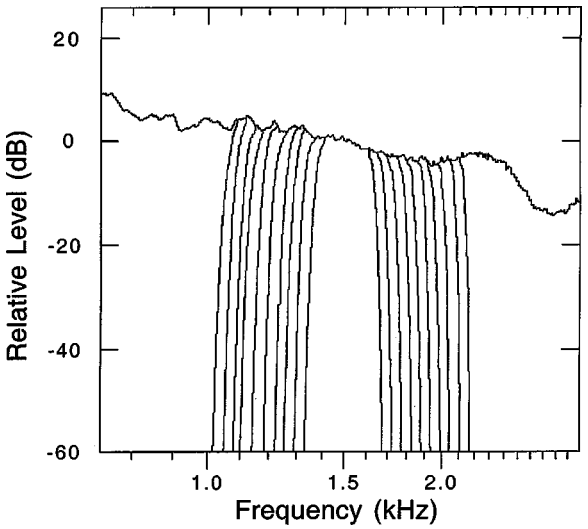


FIG. 3. Long-term averaged spectra of the 100 CID sentences subjected to steep FIR filtering to produce rectangular bands with nearly vertical slopes that were each geometrically centered at 1500 Hz. Their bandwidths ranged from 3 to 12 semitones in steps of 1 semitone. The plots were generated offline as described for Fig. 1.

were centered logarithmically on 1500 Hz, and their bandwidths varied in 1-semitone steps from 3 semitones (1/4 octave) to 12 semitones (one octave). The one-octave band had a level of 75 dBA. The narrower bands, which maintained the same spectrum level for corresponding frequencies, were consequently reduced in overall peak level (the lowest average peak level was 68 dB for the three-semitone band). The spectra of the stimuli are shown in Fig. 3.

V. RESULTS AND DISCUSSION

The intelligibility scores obtained in experiments 1 and 2 are shown in Fig. 4, and scores for experiment 3 are shown in Fig. 5. These scores were subjected to separate analyses of variance (ANOVAs), which yielded significant main effects for truncation of the transition bands in experiment 1 [$F(4,76)=14.94$, $p<0.0001$], for truncation of the triangular band in experiment 2 [$F(4,76)=14.90$, $p<0.0001$], and for

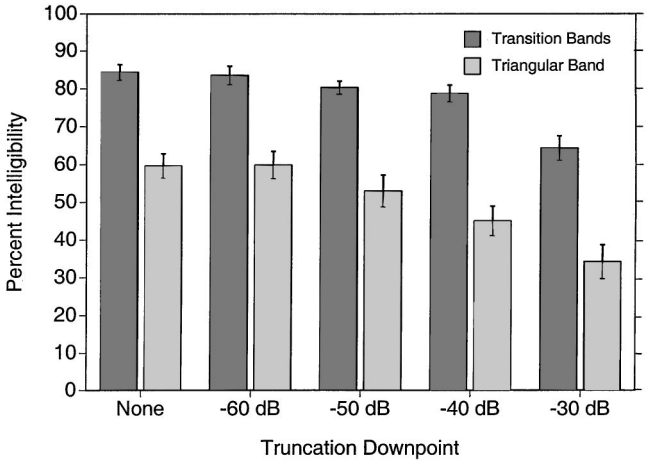


FIG. 4. Mean intelligibility scores and standard error confidence limits ($N=20$) for the transition band stimuli in experiment 1 and the triangular band stimuli in experiment 2, which were presented both intact and truncated at various downpoints.

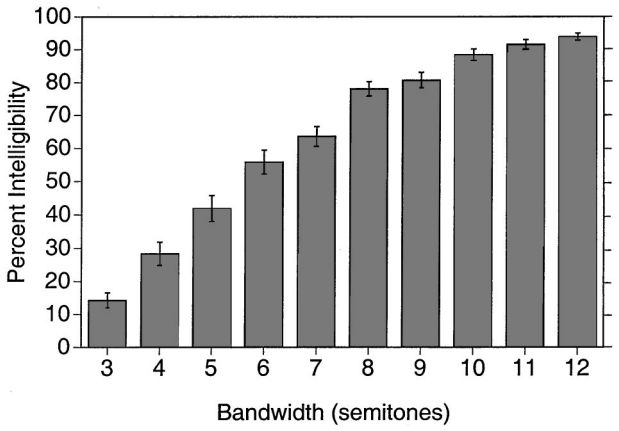


FIG. 5. Mean intelligibility scores and standard error confidence limits ($N = 20$) for rectangular bands ranging in width from 3 to 12 semitones (1/4 octave to 1 octave) as presented in experiment 3.

variation of the width of the rectangular bands in experiment 3 [$F(9,261) = 135.13$, $p < 0.0001$]. Subsequent Tukey HSD tests indicated that truncation of the transition bands in experiment 1 lowered intelligibility only when the frequencies beyond the 30-dB downpoints were removed ($p < 0.05$). In contrast, intelligibility of the triangular band in experiment 2 was reduced by truncation not only at minus 30 dB, but also when frequencies beyond the 40-dB downpoints were removed ($p < 0.05$ by Tukey HSD test). Finally, Helmhert single-df contrast analysis of the results from experiment 3 indicated that intelligibility improved consistently with increasing width of the rectangular band ($F \geqslant 14.4$, $p < 0.001$ or better) until reaching asymptote in the ten-semitone bandwidth condition.

As shown in Fig. 4, there was a small monotonic decrease in average scores in experiment 1 as the truncation downpoint moved from 60 to 40 in 10-dB steps, but no significant change from the score for the intact transition bands occurred until the 30-dB cutoff was reached. If we accept the simplifying assumption that only frequencies between the two 30-dB downpoints on the transition bands (1043 and 2067 Hz, respectively) contributed to intelligibility, then frequencies spanning approximately one octave were responsible for the near ceiling intelligibility of the nominal 1/3-octave band (passband plus intact transition bands) reported by Warren and Bashford (1999).

As shown in Fig. 4, intelligibility scores decreased when the gap separating the two transition bands in experiment 1 was closed, resulting in the high-pass and low-pass slopes intersecting at 1500 Hz. This decrease in scores (from 84% to 60%) presumably occurred because of the greater redundancy of information in a restricted spectral range (Pollack, 1968; Steeneken and Houtgast, 1999). The effects of truncation obtained in experiment 2 were also somewhat different from those obtained in experiment 1: As shown in Fig. 4, the mean intelligibility score for the intact triangular band was unchanged by truncation at the 60-dB downpoints, but the scores decreased in a linear fashion as truncations moved up the slopes. The reduction in intelligibility relative to the intact triangular band became significant at the 40-dB downpoints. Using the conservative assumption that only information limited to the frequencies lying between the two 40-dB downpoints contributed to intelligibility, then the effective bandwidth of the triangular band (in terms of intelligibility) extended from 1083 to 1947 Hz (10.2 semitones).

Figure 5 shows the monotonic increase in intelligibility in experiment 3 occurring when the bandwidths of rectangular bands centered upon 1500 Hz were increased from 3 semitones (intelligibility of 14%) to 12 semitones (intelligibility of 94%). Warren and Bashford (1999) reported that the 1/3-octave (four semitone) rectangular passband of the everyday sentences had an intelligibility of 24%, a value close to the 28% obtained for this bandwidth in the present study.[1] It also can be seen that intelligibility scores increase monotonically up to the 12-semitone limit, but scores from 10 semitones (88%) to 12 semitones (94%) increase only slightly.

In experiment 2, when the two transition bands were brought together and intersected at 1500 Hz, the intelligibility was 60%. Despite the absence of an appreciable passband, this triangular band had an intelligibility score corresponding to a rectangular bandwidth between six and seven semitones as measured in experiment 3. When the triangular band was truncated at 30 dB, the cutoff frequencies of 1178 and 1820 Hz spanned a range of 7.5 semitones, and the truncated band had an intelligibility score of 34%. A rectangular band at the same center frequency and spanning the same 7.5-semitone range had an intelligibility of about 70% as calculated from the data in experiment 3. Hence, the spectral tilt of 100-dB/octave reduced the scores of the triangular band by about 50% over this frequency range.

The ability of listeners to recognize words in spectrally limited and distorted sentences demonstrates the redundancy of acoustic information in speech. It seems probable, as suggested by Greenberg (1996) and by Shannon *et al.* (1995), that the patterns of temporal modulation occurring at local frequencies can allow listeners to comprehend degraded speech. Listeners may have the ability to utilize this temporal information despite the distortion of relative spectral amplitudes by (1): Using familiarity with the normal spectral profile of speech to compensate for filter slopes by applying an inverse transfer function of the type suggested by Watkins and Makin (1996); or (2) by applying a direct cross-frequency integration of information concerning local patterns of amplitude fluctuation without the need to restore the normal spectral profile (this latter possibility is currently being tested experimentally).

It appears that listeners having mastery of a language can achieve comprehension in a variety of ways. Although narrow-band sentences can be understood despite the absence and/or amplitude distortion of formants normally present in speech, Remez *et al.* (1994) have shown that sentences can be understood when the information consists solely of sinusoids that track the frequency and amplitude fluctuation of formants.

It is suggested that appropriate processing strategies can be selected and their results compared and combined as required by the specific task demands, integrating not only top–down and bottom–up, but also processing across similar hierarchical levels reached by different routes. This flexible

nature of speech processing is emphasized by the LAME (lateral access from multilevel engrams) model elaborated by Warren (1999) that encompasses this mélange of strategies allowing listeners to understand speech both when intact and when acoustically degraded.

ACKNOWLEDGMENT

This work was supported in part by NIH, through Grant No. DC 00208.

[1]In another study in our laboratory, Eric Healy used the same recorded sentences in his 1998 doctoral dissertation on the intelligibility of tones and noise modulated by narrow-band speech. Using a between-groups design, he also obtained data for three different rectangular bandwidths of unmodulated speech. The intelligibility scores were similar to those obtained for equivalent bands using a within-group design as shown in Fig. 5.

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., and Rzeczkowski, C. (**1984**). ''Standardization of a test of speech perception in noise,'' J. Speech Hear. Res. **27**, 32–48.

Greenberg, S. (**1996**). ''Understanding speech understanding: Towards a unified theory of speech perception,'' in *Proceedings of the Workshop on the Auditory Basis of Speech Perception*, edited by W. A. Ainsworth and S. Greenberg (Keele University, UK), pp. 1–8.

Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (**1977**). ''Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability,'' J. Acoust. Soc. Am. **61**, 1337–1351.

Pollack, I. (**1968**). ''Effect of high pass and low pass filtering on the intelligibility of speech in noise,'' J. Acoust. Soc. Am. **20**, 259–266.

Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., and Lang, J. M. (**1994**). ''On the perceptual organization of speech,'' Psychol. Rev. **101**, 129–156.

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

Silverman, S. R., and Hirsh, I. J. (**1955**). ''Problems related to the use of speech in clinical audiometry,'' Ann. Otol. Rhinol. Laryngol. **64**, 1234–1245.

Steeneken, H. J. M., and Houtgast, T. (**1999**). ''Mutual dependence of the octave-band weights in predicting speech intelligibility,'' Speech Commun. **20**, 109–123.

Stickney, G. S., and Assmann, P. F. (**1997**). ''Intelligibility of bandpass-filtered speech,'' J. Acoust. Soc. Am. **102**, 3134 (Abstract).

Warren, R. M. (**1999**). *Auditory Perception: A New Analysis and Synthesis* (Cambridge University Press, Cambridge).

Warren, R. M., and Bashford, Jr., J. A. (**1999**). ''Intelligibility of 1/3-octave speech: Greater contribution of frequencies outside than inside the nominal passband,'' J. Acoust. Soc. Am. **106**, L47–L52.

Warren, R. M., Riener, K. R., Bashford, Jr., J. A., and Brubaker, B. S. (**1995**). ''Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits,'' Percept. Psychophys. **57**, 175–182.

Watkins, A. J., and Makin, S. J. (**1996**). ''Effects of spectral contrast on perceptual compensation for spectral-envelope distortion,'' J. Acoust. Soc. Am. **99**, 3749–3757.

The relationship between speech perception and electrode discrimination in cochlear implantees

Belinda A. Henry,[a] Colette M. McKay, Hugh J. McDermott, and Graeme M. Clark
The University of Melbourne, Department of Otolaryngology, Parkville 3052, Australia

(Received 27 July 1999; revised 21 January 2000; accepted 24 May 2000)

Speech Intelligibility Index (SII) procedures were used to measure the amount of speech information perceived in five frequency bands (170–570, 570–1170, 1170–1768, 1768–2680, and 2680–5744 Hz) by 15 users of the Cochlear Ltd. CI-22M implant and Spectra-22/SPEAK processor. The speech information perceived was compared to that perceived by normal-hearing listeners. The ability of these subjects to discriminate between stimulation on adjacent electrodes corresponding to each frequency band was also investigated, using a 4IFC procedure with random current level variations of between 0% and 60% of the dynamic range. Relative to normal-hearing listeners, speech information was, on average, significantly more reduced in the four frequency regions between 170 and 2680 Hz than in the region 2680–5744 Hz. There was a significant correlation between electrode discrimination ability (when the random level variation encompassed 20% or more of the dynamic range) and the amount of speech information perceived in the four frequency regions between 170 and 2680 Hz. There was no such correlation in the region 2680–5744 Hz, regardless of the extent of random level variation. These results indicate that speech information in the low to medium frequencies is more difficult for implantees to perceive, that this difficulty is correlated with the difficulty in discriminating electrode place in the presence of random loudness variations, and that fine spectral discrimination may be relatively more important in the vowel-formant regions than in higher frequency regions. [S0001-4966(00)01109-7]

PACS numbers: 43.71.Ky, 43.66.Fe, 43.66.Ts [DOS]

INTRODUCTION

Although the majority of multichannel cochlear implant users with the most recently developed speech processors are able to enjoy some speech understanding when listening with their implant alone, a wide range of speech perception abilities remains (e.g., Skinner *et al.*, 1994). The broad factors affecting speech perception with a cochlear implant include the ability of the particular system to provide accurate and salient speech cues, and individual factors, such as the number and function of surviving spiral ganglion cells, the placement of the electrodes, and the survival of more-central neural elements. While it is not possible to measure most individual factors directly in human implantees, it has been shown that related factors such as age at implantation, duration of implant use, age at onset of deafness, as well as the duration of deafness, and its etiology, are related to speech perception ability (e.g., Battmer *et al.*, 1995; Blamey *et al.*, 1992, 1996; Dorman *et al.*, 1990; Gantz *et al.*, 1993; Parkin *et al.*, 1989). However, these correlations do not account for the majority of the variability in speech perception performance (Blamey *et al.*, 1996).

To characterize the factors that vary among implantees more precisely, it has been suggested that it may be helpful to study the ability of individual implantees to discriminate psychophysical stimuli, and to analyze the way in which this ability is related to speech understanding (e.g., Shannon, 1983, 1993; Tong *et al.*, 1982). One of the factors that may directly affect speech perception ability with a multichannel implant is the ability to discriminate stimulation on different electrodes. The few previous studies which have systematically examined the relationship between the ability to perceive place of stimulation and the ability to perceive speech in postlinguistically deafened adult listeners have yielded mixed results. Nelson *et al.* (1995) investigated the relationship between performance on a pitch ranking task and closed-set consonant recognition for nine users of the Cochlear Limited CI-22M Mini Speech Processor/MultiPeak (MSP/MPEAK) strategy. They found a significant negative correlation between consonant perception and the electrode separation required in the ranking task to achieve a specific perceptual sensitivity. Collins *et al.* (1997) also found a significant correlation between pitch ranking and pitch scaling ability, and speech perception ability in 11 MSP/MPEAK users. Throckmorton and Collins (1999) found a significant correlation between electrode discrimination and speech perception ability in seven Cochlear Ltd. Spectra-22/SPEAK users. However, Zwolan *et al.* (1997) found no correlation between electrode discrimination ability and performance for any of several speech perception measures for 11 MSP/MPEAK users, although significant improvements in speech perception were observed in individuals when the electrode discrimination data were used to optimize the speech processing programs.

The ability to discriminate electrodes as a function of overall level has also received some recent attention. Both McKay *et al.* (1999) and Pfingst *et al.* (1999) studied the effect of stimulus intensity on the ability of Cochlear Ltd.

[a]Electronic mail: belinda-henry@uiowa.edu

CI-22M users to discriminate electrodes, and showed a tendency toward better discrimination at higher intensities, although the effects were variable both across the electrode array and among subjects.

The previous studies, discussed above, which have investigated electrode discrimination ability, used psychophysical stimuli which were loudness-balanced. However, the ability of implantees to discriminate simple, loudness-balanced sounds may not be the best predictor of their performance with more complex sound patterns, such as speech, which is a complex, dynamic stimulus, varying in both frequency content and intensity. Thus an important topic for investigation is whether speech perception ability is related to the ability to discriminate place of stimulation with random loudness variations. In the study described below, the amount of speech information perceived by implantees in different spectral regions was investigated. The amount of information perceived by implantees was then compared to that of normal-hearing listeners. In addition, the relationship between the amount of speech information perceived and the ability to discriminate electrodes with stimuli varying randomly in level was investigated.

The procedures used to measure the amount of speech information perceived were based on the principles of Speech Intelligibility Index (SII) theory. The SII [based on Articulation Index (AI) theory (French and Steinberg, 1947)] is a measure of the amount of information contained in the speech signal that is available to a listener, and is defined by

$$\mathrm{SII}=\sum_{i=1}^{n} I_i \times W_i , \tag{1}$$

where n is the number of frequency bands, and I_i and W_i are the values, associated with the frequency band i, of the importance function (I), representing the relative contribution of different frequency bands to speech perception, and the audibility function (W), characterizing the effective proportion of the speech dynamic range which is audible within each band. The calculated SII, a value between 0.0 and 1.0, may be interpreted as the proportion of the total information present in the speech signal that is available to the listener. Speech recognition (measured, for example, by a score on a speech test) is related to the SII by the transfer function. The shape of both the frequency importance and the transfer functions depend on both the speech material used and the talker. One useful aspect of measuring and using the SII instead of speech perception scores is that it is assumed that the SII contributed by each of the separate bands which comprise a frequency region is independent of the contributions of the other bands. Therefore, the SII can be added across different frequency bands to estimate theoretically the total SII for the combined bands.

Although the SII has been applied by numerous authors to the prediction of the speech perception of listeners with a sensorineural hearing loss (e.g., Ching *et al.*, 1997, 1998; Dirks *et al.*, 1986; Kamm *et al.*, 1985; Ludvigsen, 1987; Pavlovic, 1984; Pavlovic *et al.*, 1986; Schum *et al.*, 1991; Skinner *et al.*, 1982), this method has not previously been applied to investigations of the speech perception abilities of cochlear implantees. In studies with hearing-impaired listeners, the SII has been shown to provide accurate predictions of speech perception for listeners with a mild to moderate sensorineural hearing loss, but tends to overestimate speech perception scores for listeners with a severe loss. This implies that factors additional to signal audibility affect the speech perception of severely hearing-impaired listeners. This effect has been modeled by Pavlovic *et al.* (1986) as an additional factor, the ''desensitization factor,'' in the SII equation. This desensitization factor, which is frequency dependent, is specified as a function of the hearing loss, having a value of 1 for hearing threshold levels less than 15 dBHL, then decreasing linearly to a value of 0 when hearing levels exceed 95 dBHL. This modification provided more accurate predictions of speech perception compared with the unmodified procedure in a study with four hearing-impaired listeners.

In a similar manner to severely hearing-impaired listeners, the poorer ability of cochlear implantees to perceive speech compared to normal-hearing listeners is due to the speech information being reduced by factors additional to signal audibility. Audibility is not the primary factor affecting variability in performance with an implant because the audibility of a conversational-level speech signal is determined mostly by the sensitivity setting of the speech processor and does not vary significantly among implantees, providing threshold and comfortable levels are set optimally. In this paper, a similar model to that of Pavlovic *et al.* (1986) has been applied to cochlear implantees, where the poorer ability of implant users to perceive speech compared to normal-hearing listeners is modeled as a reduction of the information available in the signal. This reduction is in addition to any restriction on the information due to the lower audibility of the speech signal in some frequency regions for implantees.

For implant users, Eq. (1) was altered by introducing a reduction factor, R_i,

$$\mathrm{SII}=\sum_{i=1}^{n} I_i \times W_{C_i} \times R_i , \tag{2}$$

where I_i is the frequency importance function measured with normal-hearing listeners for the particular speech material, and W_{C_i} is the audibility function for implantees, associated with the frequency band i. I_i is therefore a population factor (for normal-hearing listeners), while W_{C_i} is an individual factor, which depends on the speech processor and its sensitivity setting (but was held constant across subjects during this experiment). The value of R will be equal to 1 if there is no information reduction in a particular band compared to normal-hearing listeners (other than that predicted by audibility), and will decrease as the amount of information reduction increases. The factors which may affect R may include a loss or distortion of information in the speech signal when it is processed and transformed into an electrical signal, an effect which would be expected to affect individuals with the same implant and processor in a similar way. Information reduction may also be associated with poor neural survival at the acoustic nerve level or more-central changes in neural processing ability. Variations in these peripheral and central factors between implantees may lead to differences in pat-

terns of neural responses to the same electrical signal, and[2] therefore may contribute to differences in the amount of information reduction among implant users and between frequency regions. The value of R therefore incorporates both population (e.g., implant/speech processor) and individual (e.g., neural survival) factors.

The aim of experiment 1 (speech perception experiment) was to measure the amount of speech information reduction for individual cochlear implantees. The derived reduction factors for each subject and frequency region were then analyzed to determine whether the amount of information reduction varied consistently across the spectral bands for implant users as a group, and whether the pattern of reduction across frequency bands varied significantly among implantees with different speech perception abilities. In experiment 2 (psychophysical discrimination experiment), the relationship was investigated between the amount of information perceived in different frequency regions and the ability to discriminate adjacent electrodes associated with the same regions, using stimuli varying randomly in level.

I. EXPERIMENT 1: SPEECH PERCEPTION MEASUREMENTS

A. Methods

1. Subjects

Fifteen postlinguistically deafened adults (9 female, 6 male), using the Cochlear Ltd. CI-22M implant and the Spectra-22/SPEAK processor, participated in both experiments 1 and 2. The subjects had at least 16 electrodes activated by their speech processor, and had a wide range of speech perception abilities. Individual subject details are shown in Table I in which the subjects (S1 to S15) are ranked from better to worse speech perception ability (based on percent-correct phoneme scores). For reference, the audition-alone phoneme scores for unfiltered CNC (consonant-nucleus-consonant) words (see below), are also shown.

2. Stimuli

The test material consisted of 30 lists of 50 CNC words, which have a consonant–vowel–consonant structure, spoken by a female talker having an average Australian accent (Henry *et al.*, 1998). The word lists were recorded digitally onto compact disc, then copied onto high-quality audio cassette tapes for use in the experiment. The speech stimuli were presented to the subjects with the Spectra-22 processor and SPEAK processing strategy (Seligman and McDermott, 1995; Skinner *et al.*, 1994). The CI-22M implant consists of an intracochlear array of 22 active electrodes, spaced 0.75-mm apart, and numbered from 1 to 22 in a basal to apical direction. The electrodes are activated by an implanted receiver-stimulator, which receives digitally encoded signals from the speech processor via a transcutaneous inductive link. The current pulses may be delivered to the electrodes in one of two modes: bipolar (BP) stimulation, in which the current is passed between two electrodes, or common ground (CG) stimulation, in which current passes between the selected electrode and all other electrodes on the array, which are connected together and serve as a return path for the current. In BP stimulation, the number of electrodes separating the two electrodes can be varied (e.g., in BP+1 stimulation, the members of the active bipolar pair are separated by one inactive electrode).

3. Procedures

The amount of speech information (SII) perceived in five frequency regions was determined for each of the subjects. The method, described in detail below, involved measuring speech perception ability with each of the five frequency regions removed from the electrical stimulation, and

TABLE I. Individual subject details. Etiology: Cong./prog.=congenital/progressive. Phoneme score: unfiltered CNC words presented by audition alone (scores obtained from experiment 1).

Subject	Etiology	Age (yrs)	Duration of profound deafness (yrs)	Implant experience (yrs)	Phoneme score (% correct)	Number of electrodes in normal map	Electrodes used in experiment 1 map	Mode of stimulation	Loudness coding	Pulse width (μs/phase) (experiment 2)
1	Cong./prog.	48	1	9	92.5	16	22–12, 8, 6–3	CG	SL	100
2	Trauma	51	17	8	89.5	16	22–7	CG	SL	100
3	Otosclerosis	41	8	4	83.2	18	20–14, 11–3	BP+1	SL	100
4	Meningitis	56	45	6	83.0	16	20–5	BP+1	SL	100
5	Otosclerosis	60	16	9	82.2	18	22, 20, 18–15, 13–4	CG	CL	300
6	Unknown	57	2	11	74.0	16	20–19, 17–4	variable[a]	SL	400
7	Unknown	69	41	4	73.2	18	19–4	BP+2	SL	200
8	Unknown	43	1	8	70.2	19	20–7, 5–4	BP+1	SL	250
9	Unknown	23	10	5	68.4	19	19–4	BP+2	SL	400
10	Unknown	73	3	12	66.0	16	20–5	BP+1	SL	400
11	Cong./prog.	48	5	1	61.5	16	20–5	BP+1	SL	200
12	Unknown	50	1	8	59.9	16	18–3	BP+3	CL	185/400
13	Otosclerosis	60	14	4	47.7	16	20, 16, 14–1	variable[b]	SL	100
14	Cong./prog.	66	30	10	43.7	16	20–14, 12–4	BP+1	SL	100
15	Unknown	72	49	12	33.2	18	18–3	BP+2	SL	250

Mode of stimulation: CG=common ground; BP+1=bipolar plus one, etc.
[a]Variable mode, BP+1, BP+2.
[b]BP+1, BP+2, BP+4. Loudness coding: SL=stimulus level, CL=current level.

then transforming the speech perception scores into SII values, using the previously determined transfer function for this material (Henry *et al.*, 1998). The SII was derived in each band by subtracting the SII in the condition with the band removed from that in the condition with all bands present. The reduction factors (R_i) were derived for each subject in each band using Eq. (2), and the measured SII values, the CNC word-test frequency importance function for normal-hearing listeners (Henry *et al.*, 1998), and the audibility function for implantees.

Each subject was programmed with an experimental ''map'' which was used only for the experimental sessions. (A map contains the data required in the sound processor to control stimulation on each active electrode within a range of levels appropriate to the individual cochlear implant user.) The experimental map used exactly 16 electrodes, and had a frequency range encompassing 170 to 5744 Hz. The normal map for each individual was used to determine the electrodes to be activated, the mode of stimulation to be used, and the method of controlling loudness. In the Spectra-22/SPEAK system, loudness can be controlled by means of current levels, which use pulse height alone with a constant pulse width selected for the individual subject and electrode, or stimulus levels, which vary both the pulse height and pulse width using a single parameter. These details are shown in Table I, along with the number of electrodes in the normal maps of each subject, which varied between 16 and 19. If an individual's normal map contained more than 16 electrodes, the most-basal electrodes were removed for the experimental map. The use of the same number of active electrodes in all of the subjects' experimental maps minimized differences between the maps, so that the only subject-dependent parameters were the method of loudness coding, and the particular electrodes and modes used in their map. Note that, throughout this paper, the term ''electrode position'' refers to the electrodes assigned to the 16 filters used in the experiment, while the term ''electrode'' or ''electrode number'' refers to the active electrode ring in common ground mode, or the more-basal member of a pair in bipolar mode. Electrode positions can be converted to electrode numbers in individual subjects by referral to Table I. Threshold levels (*T*-levels) and comfortably loud levels (*C*-levels) were determined for each electrode, and loudness balancing was conducted at the *T*-level, *C*-level, and at 50% of the *T*- to *C*-level range, using standard clinical procedures. These data were used to create the experimental map for each subject.

The five frequency bands which were investigated were 170–570, 570–1170, 1170–1768, 1768–2680, and 2680–5744 Hz. The frequency ranges of these five bands were selected, based on the filters available in the speech processor, to approximate five bands of similar importance for normal-hearing listeners. The electrode positions assigned to the five frequency regions are shown in Fig. 1. The amount of speech information perceived from each of the five regions was determined for each of the subjects by measuring their speech perception with the stimulation associated with each of these regions removed. Instead of filtering the acoustic signal, as is conventional in the derivation of the frequency importance function for normal-hearing listeners, the signal was ''filtered'' after processing by removing stimulation on selected sets of electrodes in the individual's map, so that the particular band of frequencies usually mapped to those electrodes was not represented in the stimulation. The reason for filtering the electrical, rather than the acoustic, signal was that filtering the acoustic signal before processing would have resulted in the speech signal in the passband being treated differently by the speech processor. To remove stimulation in a frequency region, selected pulses were diverted to an unused, or ''dummy'' electrode. The current on the dummy electrode was set to the minimum possible level (1 current step) to ensure the stimulation was inaudible, and the pulse width was adjusted so that the average rate of stimulation was approximately the same as that for the full map.

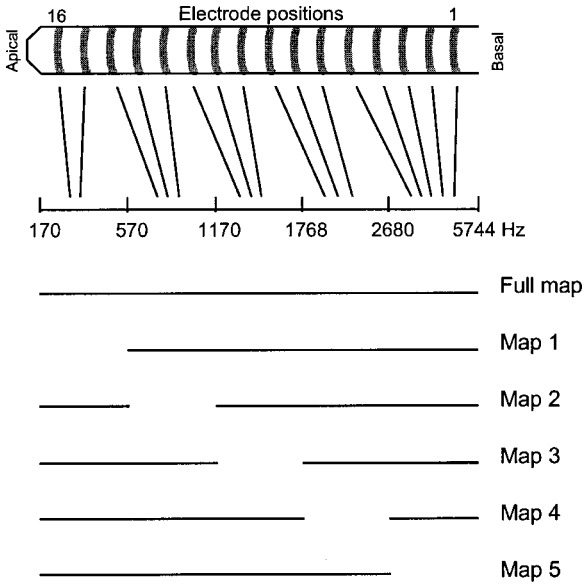


FIG. 1. Schematic representation of the assignment of electrode positions to the five frequency bands used in experiment 1. The electrode array is represented at the top of the figure, showing the 16 active electrode positions and the frequency bands to which each group of electrodes was allocated. The full-band map contained all 16 electrodes, while each of the five reduced band maps had one group of electrodes, and therefore one frequency band, removed.

Each subject was tested in a sound-treated room using the same speech processor, with four lists of CNC words under each of the six test conditions, at an average level of 70 dB SPL in the free field. The sensitivity setting of the speech processor was set to 3.5. In order to minimize individual variation in the audibility of the speech signal across subjects, the sensitivity setting was constant across subjects. There was no take-home practice for any of the maps, but the subject and tester engaged in general conversation for at least 10 min prior to testing to enable the subject to become familiar with the map. Testing for both experiments 1 and 2 was completed for each subject in approximately 15 two-hour sessions. Subjects were presented with at least four practice word lists in the full-map condition in the first test session, prior to commencing the experiment. The order of testing of the experimental conditions was pseudorandom, where the conditions were chosen so that the repeat measurements for each of the test conditions were distributed through

the sessions to minimize any learning effects. One CNC list was administered per condition per session, with multiple conditions tested in each session. The word lists were randomly selected without replacement for each test session.

Mean percentage-correct phoneme scores for each of the six test conditions were determined for each subject, and the mean scores were converted into SII values using the previously published CNC word-test transfer function for normal-hearing listeners (Henry *et al.*, 1998). In order to determine the SII in each of the five frequency regions, the SII for each of the reduced-band maps was subtracted from the SII for the full-band map.

In order to derive the reduction factors (R) for each of the five bands and 15 subjects [using Eq. (2)], the audibility function (WC) was measured in one-third-octave bands. The audibility function for implantees was estimated using

$$WC_i = (SP/\theta)_i/30. \tag{3}$$

This audibility function is based on that used in acoustic experiments. In Eq. (3), the range of speech levels contributing to intelligibility is estimated by finding the speech peaks to threshold (SP/θ) ratio, which is equal to the difference, in dB, between the 1% speech peaks [SP, the root-mean-square (RMS) level of the speech signal exceeded 1% of the time] and the signal detection threshold (θ), where θ was defined as the warbled-tone thresholds of activation of the implant. The 1% speech peak levels were derived from the previously published speech peak levels for the CNC material (Henry *et al.*, 1998). θ was derived experimentally by measuring the output of the speech processor, with the threshold of activation of the implant defined as the acoustic level of a warble tone at which the implant produced 30 pulses in a 1.5-s presentation period (i.e., an average rate of 20 Hz) on the electrode most activated by that frequency. Theoretically, if the T-levels are set appropriately for each individual, the thresholds of activation of the implant should be approximately equal to the psychophysical thresholds for acoustic tones through the processor. This assumption was verified by measuring the acoustic psychophysical thresholds in four subjects (S5, S9, S11, and S15) with the experimental (full-band) maps. These behavioral thresholds were measured in one-third octave bands, for warbled tones presented in the free-field. A t-test (paired for test frequency) showed no significant difference between the behavioral thresholds (averaged across subjects) and the objective thresholds ($p = 0.09$).

The derived one-third octave band values of W_i were multiplied by the one-third octave band values of I_i, and a cumulative $I_i \times WC_i$ curve was derived by adding these values across frequency. This cumulative function was then used to derive the $I_i \times WC_i$ values for the five frequency bands used in the experiment. Finally, the values of R in each band were derived using Eq. (2).

B. Results

1. The measured SII

The SII values for subjects 1 to 15 in each frequency band are shown in Fig. 2. The data for each subject are denoted by a different symbol (joined by nonsolid lines), with S1 to S5 in panel A, S6 to S10 in panel B, and S11 to S15 in panel C. The upper two lines in each of the three panels show the $I_i \times WN_i$ and the $I_i \times WC_i$ values in each of the five frequency bands (solid squares and solid circles, respectively). $I \times WN$ (SII available to normal-hearing listeners in each of the five frequency bands, with the normal-hearing audibility function denoted WN to distinguish it from the implantee audibility function, WC) was determined using the same procedure as for $I \times WC$, but with WN_i set to 1. Comparing the $I \times WN$ and $I \times WC$ values shows that lower audibility of the speech signal for implantees reduces the predicted amount of speech information available, mainly in the lowest and the highest frequency bands.

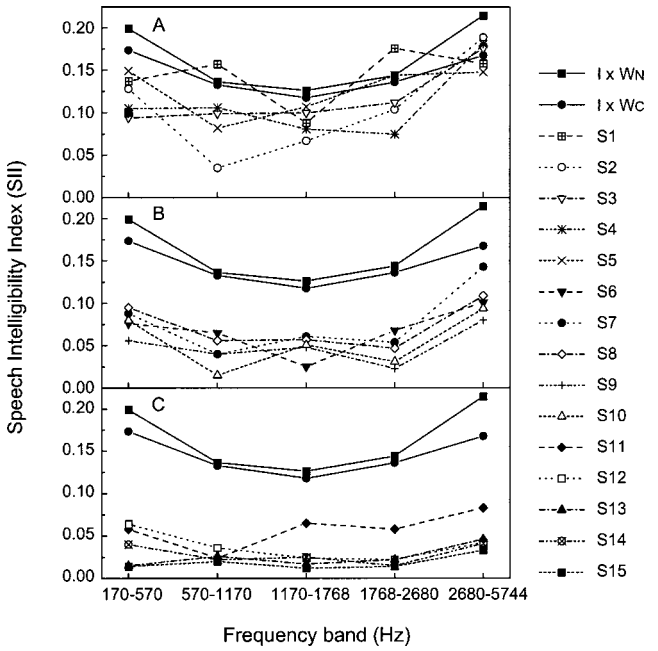


FIG. 2. The measured SII values for each subject in each of the five frequency bands. The subjects are ranked from S1, who obtained the highest phoneme score in the full-map condition, to S15, who obtained the lowest phoneme score in the full-map condition. The SII in each band is shown for S1 to S5 in A, S6 to S10 in B, and S11 to S15 in C. Also shown in each of the three panels are $I \times WN$ and $I \times WC$ (see text for details).

2. Information reduction across frequency

The variation in information reduction across frequency was investigated for the subjects as a group. Two-factor analysis of variance (ANOVA), with frequency band as one (fixed) factor and subject as the other (random) factor, showed a significant effect of both the subject ($p < 0.0001$), and the frequency band ($p = 0.001$). The mean R values are shown in Fig. 3. *Post-hoc* analysis (using Tukey's pairwise comparisons) of the effect of the frequency band showed a significantly higher R factor for the band 2680–5744 Hz than for each of the bands 170–570, 570–1170, 1170–1768, and 1768–2680 Hz. There were no significant differences in the R factors between the four lower frequency bands. Therefore, on average, the information in the 2680–5744 Hz band was reduced the least.

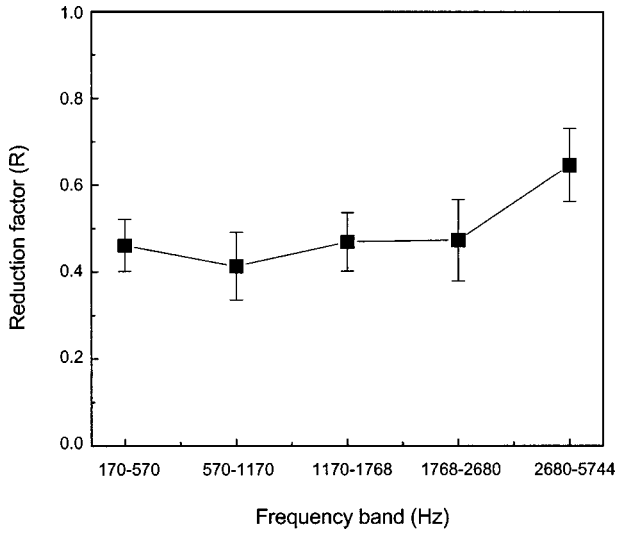


FIG. 3. The mean reduction factor in each of the five frequency bands. Speech information was significantly more reduced in the four frequency regions between 170 and 2680 Hz compared to the region 2680–5744 Hz. The error bars represent ± one standard error across subjects.

3. The dependence of the pattern of information reduction across frequency on performance level

It was hypothesized that the relative pattern of reduction across frequency bands (i.e., the reduction in one band relative to the reductions in the other bands for the same subject) may be dependent on the overall performance level. To test this hypothesis, the relative pattern of reduction across frequency bands was normalized for each subject, by dividing each R value by the average R across bands for that subject. If the hypothesis is correct, then these normalized values should be correlated with the absolute performance level. The results of regression analysis between the normalized reduction factors and the overall SII are shown in Fig. 4. Each subject's data are shown as a different symbol. While there was no significant relationship between the normalized R value and the SII in four of the five frequency bands, there was a significant correlation in the frequency band 1768–2680 Hz ($r=0.60$, $p=0.02$). The r^2 value of 0.36 indicates the proportion of the variance in the reduction factors that is accounted for by the overall performance (SII). This relationship suggests that the poorest performers had relatively greater difficulty perceiving information in the region 1768–2680 Hz than their overall performance deficit would predict, indicating that the subject-dependent reduction factors have a larger impact in this band than they do on average across frequency bands.

4. The independence of the information in the different frequency bands

One of the basic assumptions of SII theory is that the separate frequency bands which comprise the full frequency region give an independent contribution of speech information to the total information transmitted, and thus that the sum of the information in the separate bands is equal to the total information measured in the full frequency region. If the bands were not completely independent, the total SII

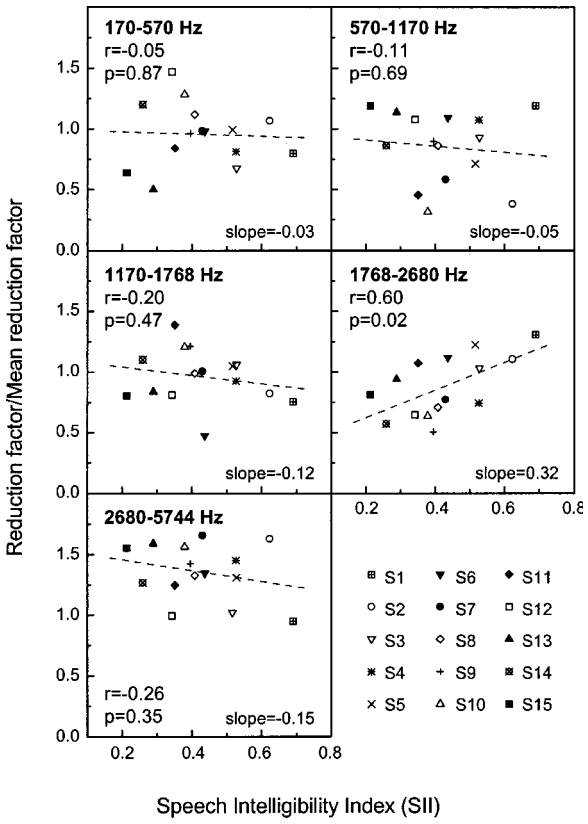


FIG. 4. The relationship in each frequency band between the normalized reduction factor and the measured SII for the 15 subjects. r is the correlation coefficient between the normalized reduction factor and the SII, and a p-value <0.05 indicates a significant correlation.

measured in the full frequency region would be larger than that derived by summing the SII values from the separate frequency regions.

Ching *et al.* (1998) found, in an investigation of the predictive accuracy of the SII model for hearing-impaired listeners, that the assumption of band independence was not met. The measured scores of 40 adult listeners with sensorineural hearing loss for broadband stimuli were consistently better than predictions (which were made using the measured results of the same subjects in a filtering experiment) based on the summed contributions of octave bands within the same bandwidth. Similar calculations with 14 normal-hearing subjects showed the expected independence of frequency bands. In order to investigate whether the assumption of independent information was true for the cochlear implant results presented here, the total SII determined from the full band in the experiment (full band SII) was plotted against the total SII determined by summing the SII values derived from each of the five individual frequency regions (summed SII) for each subject (Fig. 5). The dashed line in Fig. 5 represents the situation where the total SII measured in the full band is equal to that summed over the five separate frequency regions. In a similar manner to the results of Ching *et al.*, it can be seen that, for the poorer performing implantees, the summed SII is less than the full-band SII. The results of both the present study and the Ching *et al.* study suggest that, for hearing-impaired listeners and for poorer performing cochlear implantees, the separate frequency bands do not pro-

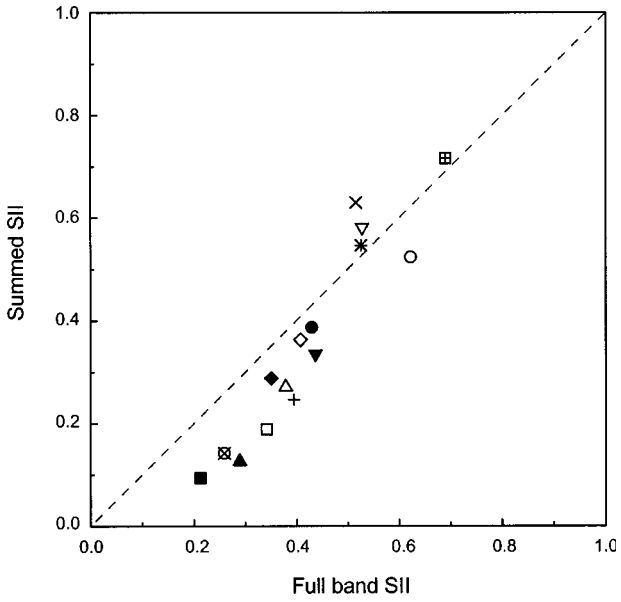


FIG. 5. The values of the summed SII relative to the full-band SII. Each data point is for an individual subject. The dashed line represents the situation where the total SII measured in the full band is equal to that summed over the five separate frequency regions.

vide a totally independent contribution to the total SII.

In this experiment, the procedure of subtracting the reduced-band SII from the full-band SII to calculate the SII in a particular band will therefore tend to underestimate the actual amount of information in that band for poorer performing subjects. To determine whether any particular band gave a disproportionate contribution to this information dependence, a similar analysis to that described above was performed for each of the reduced-band maps (maps 2 to 6). For each subject, the SII for the particular map was plotted against the total SII determined by summing the SII values derived from each of the four individual frequency regions comprising the particular map. Regression analysis showed that the slope of the regression lines were greater than 1 in all cases, and that there was no significant difference between the slopes of the regression lines for maps 1 to 6.

II. EXPERIMENT 2: RELATIONSHIP WITH ELECTRODE DISCRIMINATION

A. Methods

1. Stimuli

The stimuli used for electrode discrimination tasks were 500-ms-duration trains of biphasic current pulses delivered at a rate of 250 Hz. Current amplitude was used to control loudness, while the pulse width remained fixed for the experiment. The pulse widths of the stimuli were selected for each subject (see Table I), depending on the amount of current required to reach *C*-level, such that moderate current amplitudes could be used. The pulse widths remained constant for all electrode comparisons for all subjects except S12 (as described below). For any two electrodes under comparison, the same pulse width was used. The mode of stimulation was the same for each subject as that used both in their normal maps and in their experimental maps for experiment 1 (Table I). Electrical stimuli were presented to subjects using a speech processor interfaced with an IBM-compatible personal computer, which precisely controlled the stimulation and recorded the subject's responses using software specifically designed for the purpose.

2. Procedures

The ability to discriminate between pulse trains delivered to adjacent electrode positions was assessed for each of the 15 available adjacent positions in each subject's experimental 16-electrode map from experiment 1. The two electrode positions under comparison will be denoted in brackets (e.g., [3,4] for a comparison of electrode positions 3 and 4). Since some of the normal maps contained deactivated electrodes (see Table I), the electrode position comparisons were not always with spatially adjacent electrodes.

Prior to beginning each comparison, *T*-levels and initial *C*-levels were determined using the standard clinical procedure. For S12, electrode comparisons [13,14], [14,15], and [15,16] were made with a pulse width of 400 μs phase (in order to reach *C*-level), while the remaining comparisons were made with a pulse width of 185 μs/phase. The initial *C*-level currents on the two electrodes were then adjusted so that the percepts evoked were of equal loudness. In the loudness-balancing procedure, the more apical of the two electrodes (*A*) was chosen to be the reference, and the other (*B*) was matched to it in loudness. The two stimuli were presented at *C*-level, alternating repeatedly, separated by a silent interval of 500 ms. The subject adjusted the current of the variable stimulus (*B*) using an unmarked knob until the loudness of the two stimuli was judged to be equal. The subject was encouraged to approach the equal loudness level from both above and below before converging on the final setting. The procedure was then repeated with the current on electrode *B* fixed at the level judged as balanced by the subject in the first part of the procedure. The current on electrode *A* was then adjusted by the subject using the knob until the two stimuli were again judged to be equal in loudness. The procedure was then repeated, and the average difference in current on the two electrodes for the four trials was used to set the level of electrode *B* relative to that of electrode *A*. The dynamic range was then calculated for each electrode as the difference in current steps between the *T*-level and the adjusted *C*-level.

The ability to discriminate between the two electrodes was assessed using a four-interval forced choice (4IFC) task, in which three intervals contained stimulation on the more apical (reference) electrode, and the remaining interval, chosen at random, contained stimulation on the other (test) electrode. The stimuli were separated by silent intervals of 500 ms. The subject was asked to identify which interval contained stimulation on the test electrode, ignoring any loudness variations. Feedback was not given. The percent-correct responses out of 30 trials was calculated for each comparison. Initially, discrimination was assessed with all stimuli presented at the loudness-balanced *C*-levels. Following this, discrimination ability was assessed when the current level of the stimulation varied randomly for each interval, within the

TABLE II. Summary of the relationship between the amount of speech information perceived and electrode discrimination ability. r is the correlation coefficient, and a p-value <0.05 indicates a significant correlation.

Random level variation range	Frequency band									
	170–570 Hz		570–1170 Hz		1170–1768 Hz		1768–2680 Hz		2680–5744 Hz	
	r	p	r	p	r	p	r	p	r	p
0%	0.42	0.12	0.49	0.07	0.36	0.19	0.38	0.16	0.09	0.75
20%	0.62	0.02	0.52	0.05	0.67	0.007	0.57	0.03	0.42	0.12
40%	0.57	0.03	0.46	0.08	0.65	0.009	0.73	0.002	0.45	0.09
60%	0.58	0.03	0.51	0.05	0.57	0.03	0.81	0.0002	0.42	0.12

upper 20%, 40%, and 60% of the dynamic range of the electrode (in current steps).

B. Results

Statistical analyses with ANOVA and regression were used to analyze the electrode discrimination data. Since a Ryan–Joiner test showed that some of the discrimination scores were not normally distributed, all discrimination data were transformed into arcsine units (e.g., Thornton and Raffin, 1978). The electrode discrimination data were first analyzed using two-factor ANOVA, with subject and electrode position as factors. This analysis showed a significant effect of electrode position for all four conditions ($p=0.03$ for no level variation, $p=0.004$ for 20% level variation, $p=0.0003$ for 40% level variation, and $p=0.0004$ for 60% random level variation). *Post-hoc* analysis (using Tukey's test) suggested that electrode discrimination was significantly worse in the basal region of the cochlea compared with the middle region. This finding of relatively poorer electrode discrimination ability in the basal region of the cochlea is consistent with those of previous studies (McDermott and McKay, 1994; Nelson *et al.*, 1995; Pfingst *et al.*, 1985, 1999; Tong and Clark, 1985; Townshend *et al.*, 1987; Zwolan *et al.*, 1997).

In order to investigate the relationship between the amount of information perceived in each frequency band and the ability to discriminate electrodes in the same band, percent-correct electrode discrimination scores for adjacent electrode positions were averaged across each of the five bands used in experiment 1, for each subject and for each random level variation condition. For the frequency band 170–570 Hz, electrode discrimination data for comparison electrode positions [14,15] and [15,16] were averaged; for the 570–1170 Hz band, data for comparison positions [11,12], [12,13], and [13,14] were averaged; for the 1170–1768 Hz band, data for comparison positions [8,9], [9,10], and [10,11] were averaged; for the 1768–2680 Hz band, data for comparison positions [5,6], [6,7] and [7,8] were averaged; and for the 2680–5744 Hz band, data for comparison positions [1,2], [2,3], [3,4], and [4,5] were averaged.

The results of regression analysis (the correlation coefficient or the r-value, and the p-value) between the amount of speech information perceived (SII) and electrode discrimination ability are summarized in Table II for each of the random level variation conditions. A p-value less than 0.05 indicates that there is a significant relationship between the amount of speech information perceived and electrode discrimination ability. The relationship between the SII and average electrode discrimination in each of the five frequency bands for the random level variation of 60% is also illustrated in Fig. 6, with each subject's data shown as a different symbol. Referring to Table II, the results can be summarized as follows: For the 0% condition, there was no significant correlation between electrode discrimination and the SII in any of the frequency bands ($p>0.05$). For most of the random level variation conditions, there was a significant relationship between the amount of speech information perceived and electrode discrimination in the 170–570, 570–1170, 1170–1768, and 1768–2680 Hz frequency regions. However, there was no significant relationship between the two measures in the 2680–5744 Hz frequency region for any random level variation condition. The highest correlation was in the 1768–2680 Hz region, where the correlation be-

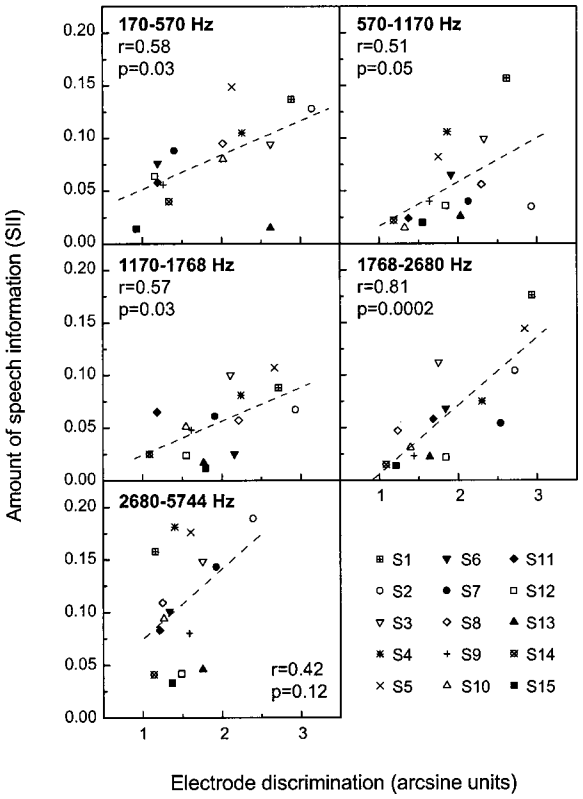


FIG. 6. The relationship between the amount of speech information perceived and electrode discrimination for the 60% random level variation condition. r is the correlation coefficient between the amount of speech information perceived and electrode discrimination, and a p-value <0.05 indicates a significant correlation.

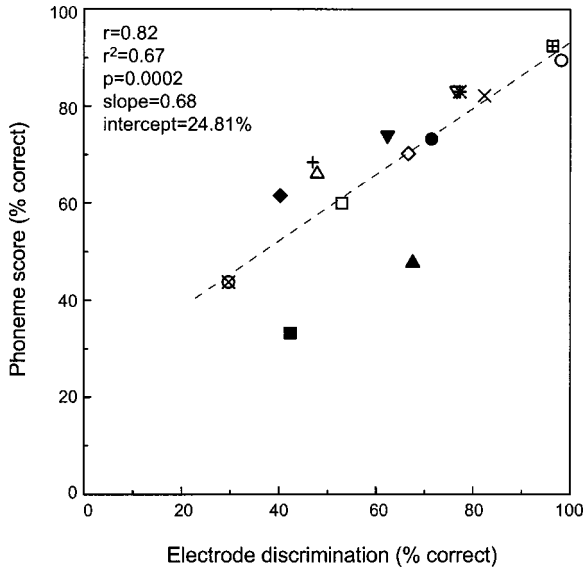


FIG. 7. The relationship between speech perception and electrode discrimination (random level variation of 60%) for the 15 subjects. Speech perception ability (percent of phonemes correctly identified), is with the full set of 16 electrodes included in the subject's map. Electrode discrimination (percent correct units) was averaged in the region 170–2680 Hz. Each subject's data set is denoted by a unique symbol (see Fig. 6).

came stronger as the random level variation range increased (the *r*-value increased and the *p*-value decreased).

The finding of a relationship between electrode discrimination and SII may have important clinical applications. In order to illustrate the predictive value of these results, Fig. 7 shows the relationship between overall speech perception ability (phoneme score for the full band condition) and electrode discrimination ability (percent correct) for the 60% random level variation. Electrode discrimination ability was averaged over the bands for which a relationship was previously found in the 60% condition (i.e., 170–570, 570–1170, 1170–1768, and 1768–2680 Hz, but not 2680–5744 Hz). Regression analysis showed a significant relationship between speech perception and average electrode discrimination ability ($p=0.0002$, $r=0.82$). The r^2 value indicates that 67% of the variability in phoneme scores is accounted for by electrode discrimination ability. Therefore, the ability to discriminate adjacent electrodes with random level variation in the 170–2680 Hz region is one major factor associated with good speech perception performance. The question of whether the relationship between speech perception and electrode discrimination is causal is discussed below. The possible clinical applications of these results are also discussed below.

III. DISCUSSION

The significant relationship between electrode discrimination with random level variation and the amount of speech information perceived in the regions 170–570, 570–1170, 1170–1768, and 1768–2680 Hz indicates that, on average, subjects who are better at discriminating adjacent electrodes corresponding to these frequencies in the presence of randomly varying levels are able to extract more information from the speech signal. It is plausible that the relationship between electrode discrimination and speech perception would result from the fact that the perception of important speech cues in the regions between 170 and 2680 Hz requires good spectral discrimination, for example, to accurately perceive the vowel formant frequencies, as well as cues to consonant place of articulation, including formant transitions.

The pattern of information reduction across frequency (Fig. 3) may be related both to population factors, such as the speech processing, and subject-dependent factors, such as individual patterns of neural survival (see the Introduction). The fact that the information in the 2680–5744 Hz region was significantly less reduced than the information in the 170–2680 Hz region for implantees as a group suggests that the implant and speech processor may not convey information from the 170–2680 Hz region as well as from the regions above 2680 Hz, or that implantees as a group may find it more difficult to discriminate information in the low-to mid-frequencies. It is interesting that while the information in the 2680–5744 Hz region was reduced the least, there was no relationship between electrode discrimination and the amount of speech information perceived in this region. These results, along with the finding that electrode discrimination ability was generally poorer for electrode positions corresponding to the higher frequencies, may imply that fine spectral discrimination is not as important for the perception of speech information in the 2680–5744 Hz region. This may be because much of the information in these frequencies is related to the wide-band noise bursts of consonants (e.g., the frequency range of the turbulent noise of fricatives). In the 1768–2680 Hz region, the poorer performers had relatively greater difficulty perceiving information than their overall performance deficit would predict (Fig. 4), and the correlation between electrode discrimination and the amount of speech information perceived was greater when more extensive level variation was applied (Table II). These results are difficult to interpret, but might indicate that the speech cues in this band are particularly vulnerable to poor spectral discrimination.

There was no significant relationship between the amount of speech information perceived and electrode discrimination ability when the stimuli were presented at loudness-balanced *C*-levels. This finding is consistent with the results of Zwolan *et al.* (1997). The lack of a correlation may have occurred if the electrode discrimination results for the 0% random level variation were limited by a ''ceiling effect'' in some frequency regions. A ceiling effect may have also limited the finding of a relationship between electrode discrimination and speech perception in the Zwolan *et al.* (1997) study, although the presence of such an effect cannot be deduced conclusively from the published data. In contrast to the results presented here and those of Zwolan *et al.* (1997), the results of Collins *et al.* (1997), Nelson *et al.* (1995), and Throckmorton and Collins (1999) have shown a correlation between pitch ranking/pitch scaling/electrode discrimination measures and speech perception ability at loudness-balanced levels. It is possible that the correlation between electrode discrimination at loudness-balanced levels and speech perception is relatively weak, and varies both with the particular number of subjects included in an experi-

ment, and the presence or absence of a ceiling effect.

The use of random level variation may have been an important factor in producing the significant correlation in this study. It is possible that the ability to discriminate electrodes when stimuli are presented at loudness-balanced *C*-levels is a less accurate predictor of the ability to perceive speech, because speech is represented in a cochlear implant using multiple stimuli that are not presented only at *C*-level, but vary widely in level. Therefore, it is necessary for implantees to be able to discriminate electrodes in the presence of random loudness variations for optimal understanding of speech. In the SPEAK strategy, the spectral shape of the acoustic signal is represented by concurrent stimulation of a number of electrodes. To enable a speech sound to be recognized, features of the spectral shape (such as the formant frequencies and relative intensities) must be represented consistently in the neural excitation pattern. If there were idiosyncratic interactions between place of stimulation percepts and stimulation level, this may cause speech phonemes to be less consistently perceived, and may also cause poorer performance in the electrode discrimination tasks with random level variation.

An alternative explanation for the positive relationship between electrode discrimination with randomized levels and speech perception is that there may be an underlying factor which affects both electrode discrimination ability and speech perception ability. Implantees who are better at discriminating electrodes with random level variation may also be better at discriminating other aspects of the electrical signal, and therefore electrode discrimination may not be the primary factor affecting speech perception ability. For example, subjects who have good neural survival and function may perform better on tests of both speech perception and electrode discrimination. It is possible that the better speech perception of these subjects may be related to the better perception of, for example, temporal cues, rather than, or in addition to, spectral shape cues. However, a more causal relationship between electrode discrimination and speech perception is supported by the finding that the relationship is confined to the frequency regions where speech cues generally require good spectral discrimination. Further research is needed to investigate whether subjects with both good speech perception and good electrode discrimination are perceiving fine spectral information from the speech signal, or whether they are better at perceiving other aspects of the electrical stimulation.

It should be noted that although the ability to discriminate 16 adjacent electrode positions was tested, the results do not necessarily imply that 16 discriminable electrodes are necessary for good speech perception. First, if the correlation between electrode discrimination and speech perception is not causal (as discussed above), then there is no reason to postulate that reducing the number of analysis channels or stimulating electrodes would cause a reduction in speech perception. Second, if the correlation with electrode discrimination was causal, then it is possible that fewer analysis channels or electrodes may still provide adequate information about the spectral shape of the speech signal. This possibility is supported by recent research, which suggests that, although speech perception ability generally increases with the number of stimulation channels, as few as seven or eight channels may provide close to the optimum speech perception for some subjects with current speech processing strategies, at least in quiet listening conditions (Brill *et al.*, 1997; Dorman and Loizou, 1997, 1998; Fishman *et al.*, 1997). However, when listening in background noise, a larger number of channels may be required for optimum speech perception (Dorman *et al.*, 1998; Fu *et al.*, 1998). Perhaps the important factor in the relationship between electrode discrimination and speech perception is not only the number of discriminable electrodes, but also the presence or absence of an interaction between place of stimulation perception and stimulation level, and therefore whether features of the acoustic spectral shape can be recognized consistently regardless of changes in intensity levels.

The clinical relevance of these findings has three aspects. First, the fact that 67% of the variability in phoneme scores was accounted for by the regression of phoneme score and electrode discrimination score (Fig. 7) indicates that speech perception may be predicted reasonably accurately by measuring electrode discrimination ability with random level variations encompassing the upper 60% of the dynamic range of each electrode. This predictive ability may provide some indication of the potential speech perception ability of cochlear implant users and may aid in appropriate speech processing strategy selection. However, in order to use the electrode discrimination measure in the clinical setting, a less time-consuming procedure for determining electrode discrimination ability would need to be devised. Second, the finding of a relationship between speech perception and electrode discrimination in the low- to mid-frequency regions gives rise to the hypothesis that speech perception might be improved if the discriminability of electrodes allocated to the vowel formant ($F1$ and $F2$) regions were improved. This may be achieved, for instance, by optimizing the selection of electrodes, which has previously been shown to yield some positive results (e.g., Skinner *et al.*, 1995; Zwolan *et al.*, 1997). Third, speech perception might be improved by increasing the audibility of lower and higher frequency signals (see Fig. 2).

IV. CONCLUSIONS

The main findings of these experiments were,

(1) Relative to normal-hearing listeners, speech information was significantly more reduced, on average, in the low- to mid-frequency regions (between 170 and 2680 Hz) compared to the higher frequency region (2680–5744 Hz).
(2) There was a significant relationship between the amount of speech information perceived and the ability to discriminate electrodes in the presence of random level variation in the low- to mid-frequency regions (between 170 and 2680 Hz), but not in the high-frequency regions (2680–5744 Hz).
(3) There was no significant relationship between the amount of speech information perceived and electrode

discrimination ability in any frequency region when the stimuli were presented at loudness-balanced levels.

(4) The pattern of information reduction across frequency varied with overall performance, with poorer performers having relatively greater difficulty perceiving information in the region 1768–2680 Hz than their overall performance deficit would predict. Also, the amount of speech information perceived and electrode discrimination ability were generally correlated most strongly in this band, and the correlation became stronger as the range over which the levels randomly varied was increased.

ACKNOWLEDGMENTS

We are grateful to all of the subjects for their time and effort in participating in this research. This research was supported by an Australian National Health and Medical Research Council project grant, ''Speech perception by cochlear implantees: Perceptual and related psychophysical studies,'' the Cooperative Research Center for Cochlear Implant, Speech and Hearing Research, and a Garnett Passe and Rodney Williams Memorial Foundation Senior Research Fellowship. The first author was supported by an Australian Postgraduate Award and a University of Melbourne Faculty of Medicine Scholarship. The speech material was recorded by the Cooperative Research Center Combionic Aid Program, and we would particularly like to thank Dr. Chris James. We also thank the three anonymous reviewers for their valuable comments on this paper.

Battmer, R. D., Gupta, S. P., Allum-Mecklenburg, D. J., and Lenarz, T. (**1995**). ''Factors influencing cochlear implant perceptual performance in 132 adults,'' Ann. Otol. Rhinol. Laryngol. **104** (Suppl. 166), 185–187.

Blamey, P. J., Arndt, P., Bergeron, F., Bredberg, G., Brimacombe, J., Facer, G., Larky, J., Lindstrom, B., Nedzelski, J., Peterson, A., Shipp, D., Staller, S., and Whitford, L. (**1996**). ''Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants,'' Audiol. Neuro-Otol. **1**, 293–306.

Blamey, P. J., Pyman, B. C., Gordon, M., Clark, G. M., Brown, A. M., Dowell, R. C., and Hollow, R. D. (**1992**). ''Factors predicting postoperative sentence scores in postlinguistically deaf adult cochlear implant patients,'' Ann. Otol. Rhinol. Laryngol. **101**, 342–348.

Brill, S., Gstöttner, W., Helms, J., v. Ilberg, C., Baumgartener, W., Müller, J., and Keifer, J. (**1997**). ''Optimization of channel number and stimulation rate for the fast continuous interleaved sampling strategy in the COMBI 40+,'' Am. J. Otolaryngol. **18** (Suppl. 6), S104–S106.

Ching, T., Dillon, H., and Byrne, D. (**1997**). ''Prediction of speech recognition from audibility and psychoacoustic abilities of hearing impaired listeners,'' in *Modeling Sensorineural Hearing Loss*, edited by W. Jestaedt (Erlbaum, Mahwah, NJ), pp. 433–445.

Ching, T., Dillon, H., and Byrne, D. (**1998**). ''Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification,'' J. Acoust. Soc. Am. **103**, 1128–1140.

Collins, L. M., Zwolan, T. A., and Wakefield, G. H. (**1997**). ''Comparison of electrode discrimination, pitch ranking, and pitch scaling data in postlingually deafened adult cochlear implant subjects,'' J. Acoust. Soc. Am. **101**, 440–455.

Dirks, D. D., Bell, T. S., Rossman, R. N., and Kincaid, G. E. (**1986**). ''Articulation index predictions of contextually dependent words,'' J. Acoust. Soc. Am. **80**, 82–92.

Dorman, M. F., Dankowski, K., McCandless, G., Parkin, J. L., and Smith, L. (**1990**). ''Longitudinal changes in word recognition by patients who use the Ineraid cochlear implant,'' Ear Hear. **11**, 455–459.

Dorman, M. F., and Loizou, P. C. (**1997**). ''Speech intelligibility as a function of the number of channels of stimulation for normal-hearing listeners and patients with cochlear implants,'' Am. J. Otolaryngol. **18** (Suppl. 6), S113–S114.

Dorman, M. F., and Loizou, P. C. (**1998**). ''The identification of consonants and vowels by cochlear implant patients using a 6-channel Continuous Interleaved Sampling processor and by normal-hearing subjects using simulations of processors with two to nine channels,'' Ear Hear. **19**, 162–166.

Dorman, M. F., Loizou, P. C., Fitzke, J., and Tu, Z. (**1998**). ''The recognition of sentences in noise by normal-hearing listeners with simulations of cochlear-implant signal processors with 6–20 channels,'' J. Acoust. Soc. Am. **104**, 3583–3585.

Fishman, K., Shannon, R. V., and Slattery, W. H. (**1997**). ''Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor,'' J. Speech Lang. Hear. Res. **40**, 1201–1215.

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

Fu, Q., Shannon, R. V., and Wang, X. (**1998**). ''Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing,'' J. Acoust. Soc. Am. **104**, 3586–3596.

Gantz, B. J., Woodworth, G. G., Knutson, J. F., Abbas, P. J., and Tyler, R. S. (**1993**). ''Multivariate predictors of audiological success with multichannel cochlear implants,'' Ann. Otol. Rhinol. Laryngol. **102**, 909–916.

Henry, B. A., McDermott, H. J., McKay, C. M., James, C. J., and Clark, G. M. (**1998**). ''A frequency importance function for a new monosyllabic word test,'' Aust. J. Audiol. **20**, 79–86.

Kamm, C. A., Dirks, D. D., and Bell, T. S. (**1985**). ''Speech recognition and the Articulation Index for normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **77**, 281–288.

Ludvigsen, C. (**1987**). ''Prediction of speech intelligibility for normal-hearing and cochlearly hearing-impaired listeners,'' J. Acoust. Soc. Am. **82**, 1162–1171.

McDermott, H. J., and McKay, C. M. (**1994**). ''Pitch ranking with nonsimultaneous dual-electrode electrical stimulation of the cochlea,'' J. Acoust. Soc. Am. **96**, 155–162.

McKay, C. M., O'Brien, A., and James, C. J. (**1999**). ''Effect of current level on electrode discrimination in electrical stimulation,'' Hear. Res. **136**, 159–164.

Nelson, D. A., Van Tassell, D. J., Schroder, A. C., Soli, S., and Levine, S. (**1995**). ''Electrode ranking of 'place pitch' and speech recognition in electrical hearing,'' J. Acoust. Soc. Am. **98**, 1987–1999.

Parkin, J. L., Stewart, B. E., Dankowski, K., and Haas, L. J. (**1989**). ''Prognosticating speech performance in multichannel cochlear implant patients,'' Otolaryngol.-Head Neck Surg. **101**, 314–319.

Pavlovic, C. V. (**1984**). ''Use of the articulation index for assessing residual auditory function in listeners with sensorineural hearing impairment,'' J. Acoust. Soc. Am. **75**, 1253–1258.

Pavlovic, C. V., Studebaker, G. A., and Sherbecoe, R. L. (**1986**). ''An articulation index based procedure for predicting the speech recognition performance of hearing-impaired individuals,'' J. Acoust. Soc. Am. **80**, 50–57.

Pfingst, B. E., Glass, I., Spelman, F. A., and Sutton, D. (**1985**). ''Psychophysical studies of cochlear implants in monkeys: Clinical implications,'' in *Cochlear Implants*, edited by R. A. Schindler and M. M. Merzenich (Raven, New York), pp. 305–321.

Pfingst, B. E., Holloway, L. A., Zwolan, T. A., and Collins, L. M. (**1999**). ''Effects of stimulus level on electrode-place discrimination in human subjects with cochlear implants,'' Hear. Res. **134**, 105–115.

Schum, D. J., Matthews, L. J., and Lee, F. (**1991**). ''Actual and predicted word-recognition performance of elderly hearing-impaired listeners,'' J. Speech Hear. Res. **34**, 636–642.

Seligman, P. M., and McDermott, H. J. (**1995**). ''Architecture of the Spectra 22 speech processor,'' Ann. Otol. Rhinol. Laryngol. **104** (Suppl. 166), 139–141.

Shannon, R. V. (**1983**). ''Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics,'' Hear. Res. **11**, 157–189.

Shannon, R. V. (**1993**). ''Psychophysics,'' in *Cochlear Implants: Audiological Foundations* (Singular, San Deigo, CA), pp. 357–388.

Skinner, M. W., Clark, G. M., Whitford, L. A., Seligman, P. M., Staller, S. J., Shipp, D. B., Shallop, J. K., Everingham, C., Menapace, C. M., Arndt, P. L., Antogenelli, T., Brimacombe, J. A., Pijl, S., Daniels, P., George, C. R., McDermott, H. J., and Beiter, A. L. (**1994**). ''Evaluation of a new

spectral peak coding strategy for the Nucleus 22 channel cochlear implant system,'' Am. J. Otolaryngol. **15** (Suppl. 2), 15–27.
Skinner, M. W., Holden, L. K., and Holden, T. A. (**1995**). ''Effect of frequency boundary assignment on speech recognition with the SPEAK speech coding strategy,'' Ann. Otol. Rhinol. Laryngol. **104** (Suppl. 166), 307–311.
Skinner, M. W., Karstaedt, M. M., and Miller, J. D. (**1982**). ''Amplification bandwidth and speech intelligibility for two listeners with a sensorineural hearing loss,'' Audiology **21**, 251–268.
Thornton, A. R., and Raffin, M. J. M. (**1978**). ''Speech-discrimination scores modeled as a binomial variable,'' J. Speech Hear. Res. **21**, 507–518.
Throckmorton, C. S., and Collins, L. M. (**1999**). ''Investigation of the effects of temporal and spatial interactions on speech-recognition skills in cochlear-implant subjects,'' J. Acoust. Soc. Am. **105**, 861–873.
Tong, Y. C., Clark, G. M., Blamey, P. J., Busby, P. A., and Dowell, R. C. (**1982**). ''Psychophysical studies for two multiple-channel cochlear implant patients,'' J. Acoust. Soc. Am. **71**, 153–160.
Tong, Y. C., and Clark, G. M. (**1985**). ''Absolute identification of electric pulse rates and electrode positions by cochlear implant patients,'' J. Acoust. Soc. Am. **77**, 1881–1888.
Townshend, B., Cotter, N., Van Compernolle, D., and White, R. L. (**1987**). ''Pitch perception by cochlear implant subjects,'' J. Acoust. Soc. Am. **82**, 106–115.
Zwolan, T. A., Collins, L. M., and Wakefield, G. H. (**1997**). ''Electrode discrimination and speech recognition in postlingually deafened adult cochlear implant subjects,'' J. Acoust. Soc. Am. **102**, 3673–3685.

In vitro measurement of the frequency-dependent attenuation in cancellous bone between 0.2 and 2 MHz

Sana Chaffaï, Fréderic Padilla, Geneviève Berger, and Pascal Laugier[a)]
Laboratoire d'Imagerie Paramétrique UMR CNRS 7623, Université Pierre et Marie Curie, 15 rue de l'Ecole de Médecine 75006, Paris, France

(Received 29 December 1999; accepted for publication 19 June 2000)

Our goal was to evaluate the frequency dependence of the ultrasonic attenuation coefficient in cancellous bone. Estimates were obtained in immersion, using a substitution method in the through-transmit mode, by scanning 14 human bone specimens (calcaneus). Measurements were performed with three pairs of focused transducers with a center frequency of 0.5, 1.0, and 2.25 MHz, respectively in order to cover an extended frequency bandwidth (0.2–1.7 MHz). When the experimental attenuation coefficient values were modeled with a nonlinear power fit $\alpha(f)=\alpha_0+\alpha_1 f^n$, the attenuation coefficient was found to increase as $f^{1.09\pm0.3}$ over the measurement bandwidth. However, a substantial variation of the exponent n (0.4–2.2) within specimens and also between specimens was observed. The acoustical parameters were compared to bone mineral density. A highly significant relationship was noted between α_1 and BMD ($r^2=0.75$, $p<10^{-4}$). No correlation was found between n and BMD. Several attenuation mechanisms are discussed as well as the potential impact these results may have in *in vivo* quantitative measurements. © *2000 Acoustical Society of America.* [S0001-4966(00)04009-1]

PACS numbers: 43.80.Ev, 43.80.Jz, 43.80.Qf, 43.80.Vj [FD]

I. INTRODUCTION

The development of quantitative ultrasound (QUS) techniques for skeletal assessment has been very successful during the past five to ten years.[1] A number of studies have demonstrated that ultrasonic velocity and attenuation represent promising means of evaluating skeletal status[2–5] and several clinical devices have been introduced.[6] The goal of QUS is to have a tool for an accurate diagnosis of osteoporosis, for correct identification of patients with a high fracture risk, and that is sensitive for monitoring skeletal changes over time. Toward this goal, further efforts to systematically investigate the ultrasonic bone properties and to determine the dependence of ultrasonic parameters on specific bone properties, such as bone mineral density (BMD), characteristics of bone micro-architecture, and material properties of bone tissue or bone strength are required. For the time being, BMD represents the gold standard for *in vivo* assessment of fracture risk and is currently measured using x-ray absorptiometric techniques. Bone strength is a function of BMD; however, only approximately 70%–80% of the variance of bone strength is accounted for by bone density, and this measurement alone may be insufficient in accurately predicting the risk of osteoporotic fracture. The fact that ultrasound is a mechanical wave makes it the natural candidate for a technique to measure mechanical properties.

Bone QUS is still in its infancy. Today QUS measurements mostly rely on the assessment of two fundamental parameters: broadband ultrasound attenuation (BUA) and Speed of Sound (SOS). BUA is the rate of change of attenuation with frequency, is measured in dB/MHz, and is determined by linearly fitting the attenuation values across a range of frequencies usually comprised in the interval 200–600 kHz. At this point in time the potential of these QUS approaches to capture some aspects of bone strength that cannot be measured by BMD appears to be limited.[7–10] The introduction of new QUS parameters should allow this field to advance much farther. One development that is of particular interest in this regard is the measurement of the frequency dependence of the attenuation. The magnitude and frequency dependence of ultrasonic attenuation is a complicated function of the composition and micro-architecture of the propagating medium. However, one needs to note a lack of theoretical model that describes how the attenuation is influenced by properties of bone. The BUA approach, implying a simple linear relationship between attenuation and frequency, does not rely upon any physical basis. While applicable over restricted frequency ranges, this assumption of a simple linear relationship between attenuation and frequency may not be universally true and may result in a loss of relevant information. In the present study we extend our systematic investigation of ultrasonic properties of human cancellous bone[8,11,12] to the measurement of the frequency dependence of the attenuation from human calcaneus and to clarify the relationship between the frequency variation of attenuation and BMD.

In Sec. II of this paper we describe bone specimens, the general approach to measure the attenuation coefficient, as well as characteristics of the data acquisition and data analysis. In Sec. III we present the results. In Sec. IV the results are discussed and compared to the state-of-knowledge of attenuation models in cancellous bone. Also, some of the potential limitations of this study are identified.

a)Electronic mail: laugier@lip.bhdc.jussieu.fr

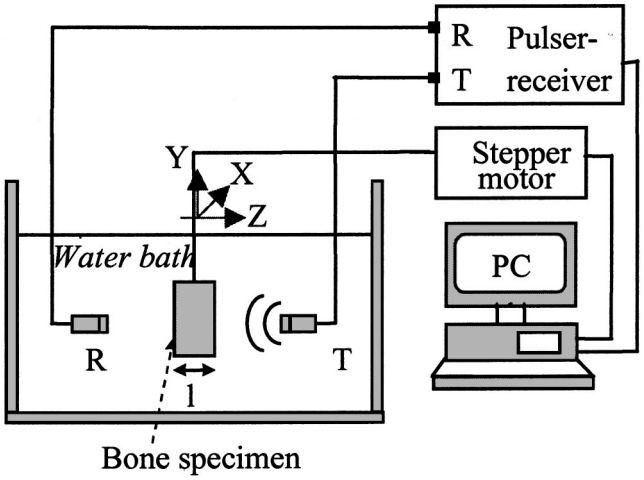


FIG. 1. Experimental setup. T: transmitter; R: Receiver.

II. MATERIAL AND METHODS

A. Bone specimens

Fourteen human calcaneus were obtained from embalmed cadavers (9 women and 5 men, ages ranging from 69 to 89 years) and stored at $-20\,^{\circ}\mathrm{C}$. The bone specimens, while unfrozen, were cut by sawing the entire calcaneus with a rotary saw under a water jet, achieving slices of approximately 1 cm in thickness of cancellous bone in the sagittal plane with parallel plane faces. One slice only was obtained per bone specimen, having the longitudinal and lateral dimensions of the whole calcaneus in the sagittal plane and a uniform thickness (± 1 mm). The samples were then defatted, using methanol, ether, and petroleum ether. The specimens were kept exposed to air at room temperature. They were simply refilled with water under vacuum just before the measurements.

B. Ultrasound measurement

A schematic diagram of the ultrasonic system is shown in Fig. 1. The system consisted of a pair of ultrasonic broadband transducers mounted coaxially, separated by twice the focal length. Transducer alignment was adjusted to maximize the amplitude of a received reference signal transmitted through water. The bone specimens were immersed in a water bath at room temperature ($20\,^{\circ}\mathrm{C}$ approximately), suspended by a clamp which could be rotated around the vertical axis (Y-axis). The bone specimens were simply attached to the clamp and placed half way between the two transducers. They were aligned parallel to the transducer faces by maximizing the amplitude of the signal reflected from their planar front surface. This was achieved by slightly rotating the clamp around the vertical axis. Bone specimens were tested with the ultrasound beam axis (Z-axis) parallel to the medio-lateral axis (along the same path as it would be for *in-vivo* heel measurements with available clinical devices). With this measurement configuration, the beam axis was perpendicular to the direction of main orientation of the bone trabecular network.[13] The temperature of the bath was measured with a precision of $\pm 0.1\,^{\circ}\mathrm{C}$. The supporting electronics consisted of a broadband pulser/receiver (CUS0), amplifier, and digitization card (all furnished by, Contrôle US S.A., Orsay, France). Each received signal was amplified in the pulser/receiver unit, averaged temporally 20 times, digitized at a rate of 12 MHz or 24 MHz and then transferred to the host computer for further processing. 512 samples per trace were digitized into 8 bits.

Ultrasonic measurements were conducted using a substitution method, in a through-transmission normal incidence configuration. A broadband ultrasound pulse was received first without and then with the sample interposed between the transducers and their Fourier transform was computed. With no intervening tissue, a calibration signal is obtained whose amplitude spectrum is:

$$A^r(f)=A_0(f)A_d^r(f); \tag{1}$$

$A_0(f)$ is the instrumentation transfer function including the amplitude spectrum of the input electrical signal, the transfer functions of transmitting and receiving transducers and electronics. $A_d^r(f)$ is a transfer function (diffraction) which characterizes the effect of propagation on the ultrasound pulse for the water path only. This term is dependent upon the geometrical characteristics and frequency of the transmitting and receiving transducers, and upon the acoustic properties of the intervening medium.

With the sample of thickness l placed in the path of the ultrasonic beam, the voltage as function of frequency is

$$A(f)=A_0(f)T(f)A_d(f)e^{-\alpha(f)l}, \tag{2}$$

where $T(f)$ is the combined transmission coefficients associated with the single crossing of a wave at a water to sample interface and a sample to water interface, $A_d(f)e^{-\alpha(f)l}$ is the transfer function that characterizes the effect of propagation on the ultrasound pulse for the water–sample–water propagation path, $\alpha(f)$ being the frequency-dependent attenuation coefficient of the bone specimen. One defines the apparent attenuation (i.e., signal loss), on a logarithmic scale, as follows:

$$\hat{\alpha}(f)l=\mathrm{Ln}\frac{|A^r(f)|}{|A(f)|}, \tag{3}$$

where $\hat{\alpha}(f)$ is the measured apparent attenuation coefficient. Therefore following Eqs. (1) and (2), the apparent attenuation can be written as

$$\hat{\alpha}(f)l=\alpha(f)l+\mathrm{Ln}\frac{|A_d^r(f)|}{|A_d(f)|}-\mathrm{Ln}|T(f)|. \tag{4}$$

Extraction of the attenuation coefficient $\alpha(f)$ from the empirically determined signal loss [as described by the term on the left hand side of Eq. (4)] assumes that the diffraction and transmission losses [as described by the second and third terms on the right hand side of Eq. (4), respectively] be either negligible or accounted for. Provided the phase velocities of the sample and water are reasonably well matched, a requirement comfortably met for measurements on trabecular bone,[8] the effect of diffraction is small and can be neglected.[12] The transmission losses and mode conversions at the interface between a fluid and a fluid-saturated porous medium have been studied by Wu *et al.*[14] We have followed their approach in which the interface conditions at the

boundary between a fluid and a porous medium have been derived directly from Biot's theory for poroelasticity. Using the input parameters of cancellous bone given by Hosokawa and Otani[15] with a mean porosity of 90%,[13] we have found that 99% of the incident energy was transmitted as a fast longitudinal wave, and that 1% only of the energy was lost at the interface between the fluid and the fluid-saturated porous medium (i.e., by reflection or mode conversion). This result strongly suggests that transmission losses at the interface may be neglected.

The specimens were measured (at room temperature) using three pairs of focused broadband transducers (Panametrics Inc., Waltham, MA) all with a diameter of 29 mm: one matched pair of 0.5 MHz-center frequency transducers (focal length of 40 mm), one matched pair of 1.0 MHz-center frequency transducers (focal length of 50 mm), and a pair of unmatched transducers with center frequencies of 2.25 MHz (transmitter, focal length of 50 mm) and 1 MHz (receiver, focal length of 50 mm) (Fig. 2). Passage through bone with its highly frequency-dependent attenuation reduced the frequency maxima and the width of the usable bandwidth. In the third measurement configuration (transmitter of 2.25 MHz), the frequency bandwidth of a 1 MHz-center frequency transducer was found to be better adapted to the bandwidth of the strongly filtered received signal. Measuring the frequency dependence over a wide frequency bandwidth (i.e., combining three partially overlapping bandwidths) was achieved to increase the reliability of estimates of the frequency dependence of attenuation. Two stepping motors mounted on a crossed slide assembly moved the specimen in the ultrasonic beam in a raster fashion in two perpendicular XY directions. The signals transmitted through bone were recorded with each pair of transducers in turn across a 60 $\times 60$ mm scan plane with a displacement of 1 mm between each acquisition. At each of the $(60)^2$ measurement points, the apparent attenuation coefficient $\hat{\alpha}(f)$ was estimated according to Eq. (3).

Bone mineral density (BMD) was measured using x-ray quantitative computed tomography (QCT) following the procedure described in Ref. 8.

C. Analysis of data

The data acquired from an ultrasonic scan represent the attenuation through a volume of the bone specimen (60 mm longitudinal dimension $\times$ 60 mm lateral dimension $\times$ thickness of the specimen). By analyzing the data from a windowed region of a small number of adjacent lines forming a small volume, a local estimation of the attenuation can be obtained. This procedure was repeated on small neighboring volumes within the scanned region. For these measurements, the choice was made to divide the data from each scan into subgroups of 7×7 neighboring lines. The apparent attenuation coefficients were averaged from the corresponding 49 neighboring measurement points in the region of interest (ROI). A variable number (ranging from 16 to 29) of adjacent nonoverlapping regions were analyzed according to the size of each specimen. For each region of bone, the average apparent attenuation coefficient $\langle \hat{\alpha}(f) \rangle$ was determined for each pair of transducers over a fixed frequency bandwidth. For the first pair of transducers (center frequency 0.5 MHz), the usable frequency bandwidth was 0.2–0.7 MHz. It was 0.4–1.2 MHz for the second pair of transducers (center frequency 1 MHz) and 0.5–1.7 MHz for the third pair of transducers (center frequency 1 MHz–2.25 MHz). The criterion used to restrict the frequency range bandwidths used in the calculation was the -10 dB bandwidth of the signals transmitted through the most attenuating regions of bone specimens. This criterion provided a good signal-to-noise ratio. A relatively small bandwidth of 0.5–1.7 MHz for the pair of 1 MHz–2.25 MHz transducers was selected to reduce the effect of noise at low signal levels particularly at the upper frequency limit, due to a considerable level of attenuation reached with some of the specimens.

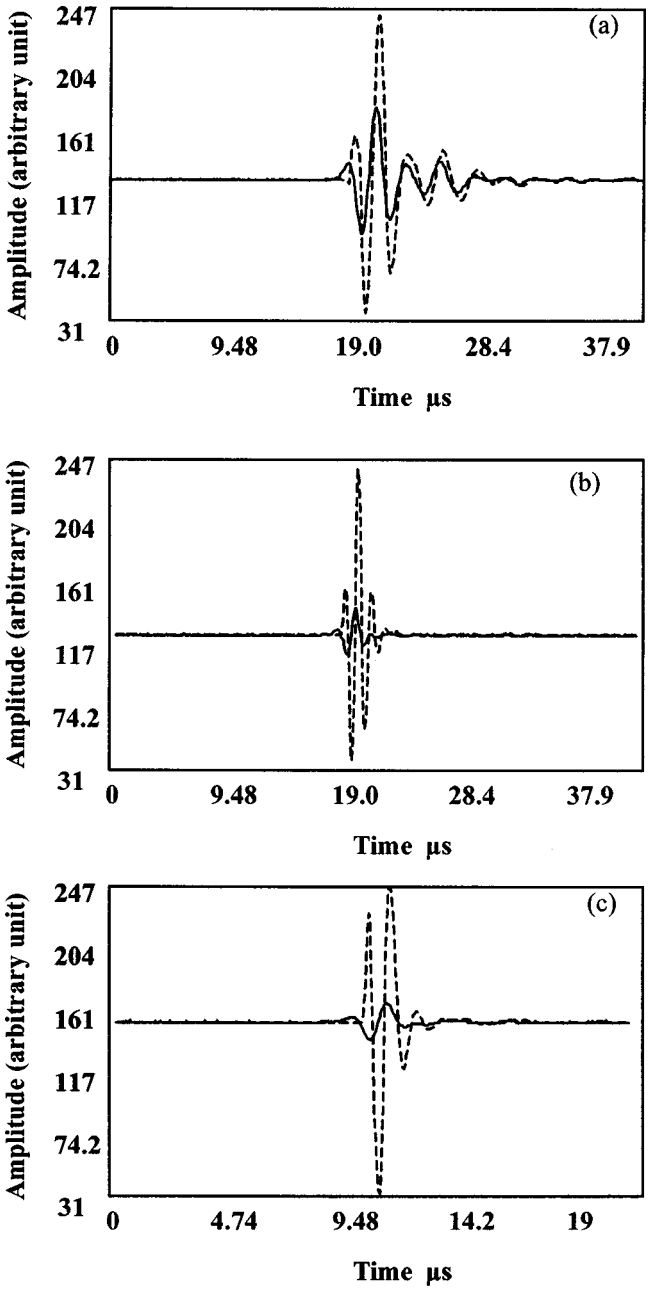


FIG. 2. Reference signals (dashed line) and signals transmitted through bone (full line) recorded with (a) 0.5 MHz center frequency pairs of transducers; (b) 1 MHz center frequency pairs of transducers; (c) 1 MHz–2.25 MHz center frequency pairs of transducers.

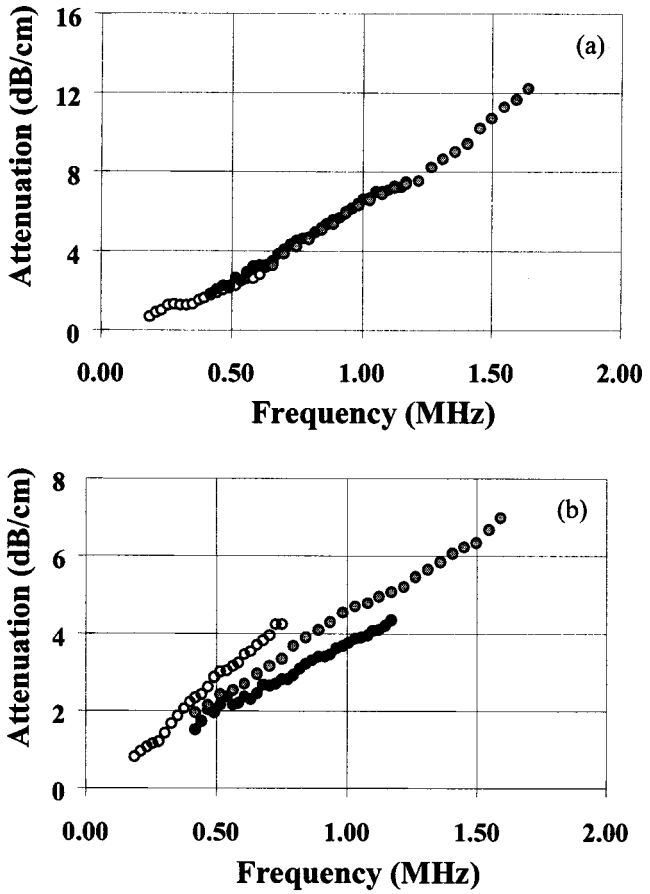


FIG. 3. Representative values of attenuation coefficient $\langle \hat{\alpha}(f) \rangle$ versus frequency using three different pairs of transducers for two bone specimens. (a) Acceptable data (average RMSE less than 2 times the technique's precision); (b) nonacceptable data (average RMSE greater than 2 times the technique's precision). ○: 0.5 MHz center frequency pairs of transducers; ●: 1 MHz center frequency pairs of transducers; gray dots: 1 MHz–2.25 MHz center frequency pairs of transducers.

The precision of estimates of $\langle \hat{\alpha}(f_i) \rangle$ at any discrete frequency f_i was defined as the coefficient of variation (CV) of repeated measurements on a given region of analysis of a given specimen with intermediate repositioning:

$$\mathrm{CV}(f_i)\% = 100 * \frac{\mathrm{SD}(f_i)}{\langle\langle \hat{\alpha}(f_i) \rangle\rangle}, \tag{5}$$

where $\mathrm{SD}(f_i)$ and $\langle\langle \hat{\alpha}(f_i) \rangle\rangle$ are the standard deviation and the mean value of estimates of $\langle \hat{\alpha}(f_i) \rangle$ over repeated measurements. Measurements of precision on a single specimen or region of analysis may not be representative of the performance of the technique in general. A representative group of specimens was assessed and the technique precision was given by the RMS average of the CVs of all the individual regions analyzed for these specimens. Precision was assessed in the frequency range 0.4–1.2 MHz on 4 bone specimens using the pair of 1 MHz-center frequency transducers.

The three attenuation curves obtained with the three pairs of transducers for each region of analysis were assembled in an attenuation-frequency plot similar to Fig. 3. Because system-dependent factors or uncertainties may affect estimates of the magnitude and frequency-dependence of the attenuation coefficient, a data acceptance criterion was

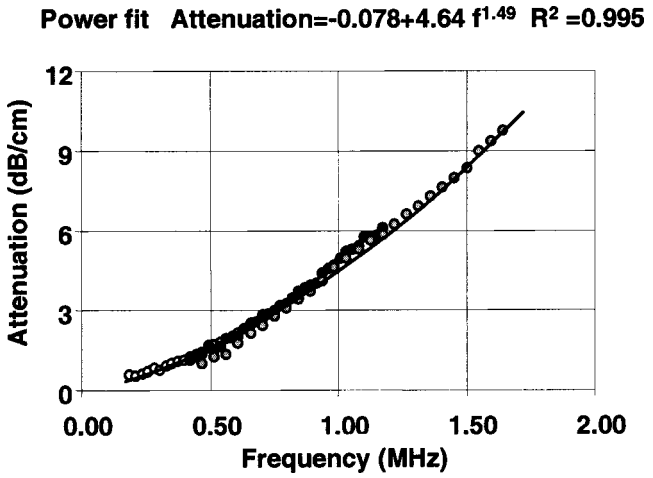


FIG. 4. Example of nonlinear power fit for a representative specimen. ○: 0.5 MHz center frequency pairs of transducers; ●: 1 MHz center frequency pairs of transducers; gray dots: 1 MHz–2.25 MHz center frequency pairs of transducers; — power nonlinear fit.

established for the present study based on the goodness of the agreement between the frequency-matched estimates of $\langle \hat{\alpha}(f) \rangle$ performed under various experimental conditions (i.e., using different pairs of transducers). As the frequency bandwidth of each pair of transducers partially overlapped, a root mean square error (RMSE) was computed from the measurement of $\langle \hat{\alpha}(f) \rangle$ in the overlapping frequency bandwidth ($f_i = f_1, \ldots, f_n$) of two different pairs of transducers following:

$$\begin{aligned} \mathrm{RMSE}_{1-2} &= \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\langle \hat{\alpha}_2(f_i) \rangle - \langle \hat{\alpha}_1(f_i) \rangle)^2}, \\ \mathrm{RMSE}_{1-2}(\%) &= 100 * \frac{\mathrm{RMSE}_{1-2}}{\frac{\langle \hat{\alpha}_2(f_c) \rangle + \langle \hat{\alpha}_1(f_c) \rangle}{2}}, \end{aligned} \tag{6}$$

where subscripts 1 and 2 refer to the first and second pairs of transducers, $\langle \hat{\alpha}_1(f_i) \rangle$ and $\langle \hat{\alpha}_2(f_i) \rangle$ are the measured attenuation coefficients in the first and second frequency bandwidths, respectively (e.g., 0.2–0.7 MHz and 0.4–1.2 MHz), f_1 and f_n the lower and upper frequency limits of the overlapping frequency bandwidth, and f_c the center frequency of the overlapping frequency bandwidth. Large values of RMSE would be expected whenever either system contributions to the attenuation differed substantially with different transducers or substantial error occurred in specimen repositioning when changing the pair of measuring transducers. A single RMSE value was obtained for a given specimen by averaging RMSE values from of all the regions of analysis. An objective criterion was used to segregate acceptable data (average RMSE$<$2 times the technique's precision) from nonacceptable data (average RMSE$>$2 times the technique's precision), as illustrated in Fig. 3. Values of RMSE less than two times the technique precision would mean that systematic errors, if any, cannot be distinguished from measurement uncertainties. Whenever this acceptance criterion was met, the attenuation coefficient from the different frequency ranges were pooled together for further processing. Three specimens out of the 14 measured specimens did not meet

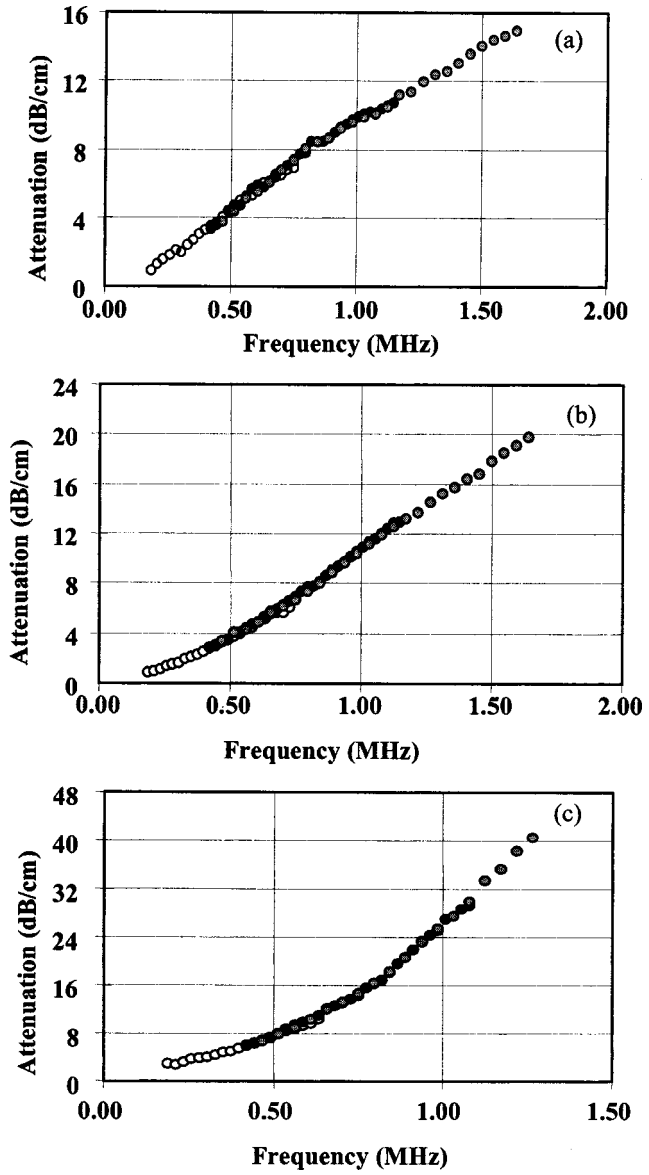


FIG. 5. Attenuation coefficient versus frequency for three representative bone specimens with $n=0.8$ (a), $n=1.1$ (b), and $n=2.2$ (c). ○: 0.5 MHz center frequency pairs of transducers; ●: 1 MHz center frequency pairs of transducers; gray dots: 1 MHz–2.25 MHz center frequency pairs of transducers.

the acceptance criterion, and thus were excluded from the analysis. In total, 215 different regions were analyzed from the 11 bone specimens.

Then, the frequency variation of the attenuation was characterized by the exponent n of nonlinear power fit least-squares calculations with respect to frequency over the total bandwidth 0.2–1.7 MHz (resulting from the combination of the individual frequency bandwidth of each pair of transducers). That is, the apparent attenuation coefficient was assumed to obey a simple power law, $\langle \hat{\alpha}(f) \rangle = \alpha_0 + \alpha_1 f^n$, where α_0 is a constant. The fit was performed using the Levenberg–Marquardt method, a standard of nonlinear least-squares routines.[16] The nonlinear fit proceeds iteratively and must be fed with an initial set of parameters. The choice of these initial values is critical, since the error function must be convex in the domain of variation of the parameters

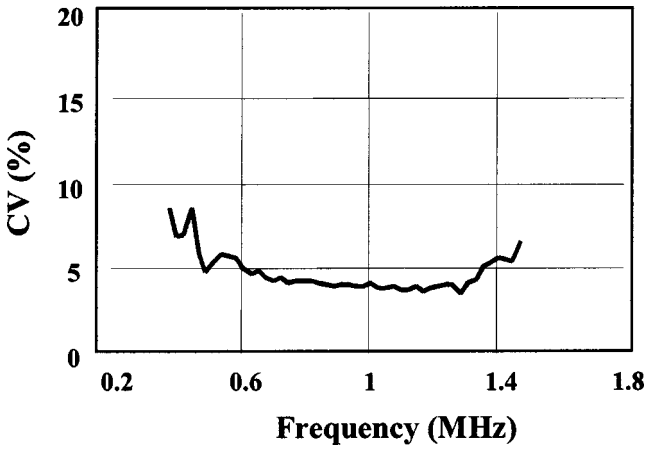


FIG. 6. Technique precision plotted as a function of frequency.

(α_0, α_1). Initial input parameters were derived from a linear fit over the total bandwidth. This choice was relevant since the linear fit led to reasonable predictions of experimental data in the frequency range. In evaluating estimates of frequency variation, the coefficient of determination (r^2), a measure of the goodness of the fit of data to theoretical model, was used. An example of nonlinear power fit is displayed in Fig. 4. The correlation coefficients between ultrasonic parameters n and α_1 *and* site-matched BMD values were calculated. The site-matching procedure has been extensively described previously in Ref. 8.

III. RESULTS

Figure 5 depicts examples of attenuation-frequency plots obtained for three representative specimens included in the study. Good agreement between estimates of $\langle \hat{\alpha}(f) \rangle$ is noted in overlapping frequency bandwidths from different pairs of transducers. Precision was found to be nearly constant (5%) over the frequency range 0.5–1.3 MHz (Fig. 6). RMSE values were relatively low in 11 of the 14 specimens and were consistent with technique precision errors. In these 11 specimens, RMSE between apparent attenuation coefficients measured with the 0.5 MHz and 1 MHz pairs of transducers was, on average, $(\pm \mathrm{SD})\ 7 \pm 4\%$. It was $5 \pm 4\%$ when the measurements were performed with the 1.0 MHz and the 1 MHz–2.25 MHz pairs of transducers. A significant deviation between RMSE and expected precision error was found in 3 specimens out of 14, suggesting inaccurate repositioning for these specimens. Small changes in the orientation of the ultrasonic beam with respect to the sample during each experiment may have induced marked changes in the attenuation-frequency plots.

Descriptive statistics of the results of the non linear power fit are shown in Table I. The mean value, standard deviation and range of α_1 and n are indicated along with the determination coefficient (r^2) for each specimens and for the pooled data. Between specimen statistics (mean, standard deviation) were calculated from considering all 11 mean values obtained on the 11 specimens. The standard deviation of the 11 mean values represented the between specimen variability. We observed substantial variation of the parameter within specimens (i.e., between ROIs of a given specimen)

TABLE I. Descriptive statistics of non-linear power fit for the 11 specimens and for the pooled data.

Sample #	Number of ROI	α_1 (dB/cm)			n			r^2		
		Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
1	17	13.7	7.3	4–29	0.94	0.22	0.7–1.3	0.991	0.006	0.977–0.997
2	19	9.6	3.2	5–16	1.06	0.22	1.7–1.5	0.994	0.004	0.984–0.999
3	16	6.1	2.6	3–11	1.37	0.14	1.1–1.5	0.994	0.002	0.991–0.997
4	22	19.0	10.6	4–40	0.95	0.24	0.5–1.4	0.987	0.006	0.976–0.996
5	18	20.6	8.1	10–38	0.95	0.25	0.6–1.5	0.976	0.019	0.915–0.993
6	20	15.8	10.3	4–35	1.15	0.32	0.5–1.8	0.991	0.005	0.979–0.997
7	19	14.8	6.1	6–26	0.90	0.22	0.5–1.3	0.990	0.004	0.982–0.997
8	29	20.1	7.8	8–37	1.19	0.44	0.4–2.2	0.992	0.007	0.967–0.999
9	18	6.2	3.3	2–14	1.26	0.19	1.0–1.7	0.993	0.003	0.986–0.999
10	20	11.0	6.8	3–26	1.19	0.28	0.7–1.7	0.994	0.004	0.984–0.998
11	17	8.9	4.3	4–19	0.99	0.19	0.7–1.4	0.991	0.009	0.966–0.997
All data pooled	215	13.8	8.4	2–40	1.09	0.30	0.4–2.2	0.990	0.009	0.915–0.999
Between sample		13.3	5.3	6–21	1.09	0.16	0.9–1.4	0.990	0.005	0.976–0.994

and also between specimens. The plot of the distribution of all measured n values is given in Fig. 7. Exponent n ranged between 0.4 and 2.2. Variations of mean n between specimens fell in the range 0.9 to 1.4. On average, the attenuation coefficient was found to increase as $f^{1.09\pm0.3}$ (mean$\pm$SD). When pooling the data, a highly significant linear relationship was noted between α_1 and BMD ($r^2=0.75$, $p<10^{-4}$). No correlation was found between n and BMD (Fig. 8).

IV. DISCUSSION

Quantitative ultrasound measurements have been successfully used recently to predict fracture risk in osteoporosis. This technique is now incorporated into commercial devices, enabling attenuation to be measured. Attenuation is typically characterized in the clinical practice by BUA, i.e., the slope of a linear fit to the measured frequency-dependent attenuation. Measurement of the frequency dependence in a relatively narrow frequency range 200–800 KHz of the attenuation in human cancellous bone was initially described by Strelitzki and Evans.[17] This former study indicated that the attenuation fits better to a second order polynomial function of frequency, than to the linear fit.

The reliability of estimates of frequency dependence of the attenuation increases with the frequency bandwidth used in the calculations. For this reason, our bone specimens were measured with three different pairs of transducer to cover a wider frequency range. In this study we document the frequency variation of the attenuation of human cancellous bone over the frequency range 0.2–1.7 MHz. On the other hand, we are aware that the accuracy may be impaired by nonconstant system-dependent factors varying with the different transducers used in this study. The accuracy of the measurements techniques has been tested by preliminary measurements on test samples and has been reported elsewhere.[12] The overall continuity of the curves and the relatively low RMSE values offer persuasive results showing (a) that the response of the measurement system, including the diffraction effect, either has been removed or is negligible and (b) that the specimens were correctly repositioned between successive measurements with different pairs of transducers.

Another possible source of artifact in the apparent acoustic properties, besides the diffraction effect, may be due to the phase-cancellation effect associated with geometrical distortion of ultrasonic wave fronts by inhomogeneous or irregularly shaped specimens, or even by the roughness of the specimen surface. The results obtained in this study are based on the assumption that phase-cancellation effects are insignificant. This is a valid assumption when the sample is homogeneous, or weakly inhomogeneous. The assessment of phase cancellation by Strelitzki *et al.* with a multielement receiver[18] suggested that measurements were not strongly affected by phase cancellation. In addition, we have found that the transmission losses at the fluid/highly porous cancellous bone interface could reasonably be neglected in a first approximation.

The frequency variation of the attenuation coefficient between 0.2–1.7 MHz has been characterized with a nonlinear power law $\langle \hat{\alpha}(f)\rangle = \alpha_0 + \alpha_1 f^n$. A nonlinear power fit has been used in a first approach, since it may potentially provide helpful information to describe the attenuation mechanisms in trabecular bone. Some investigators have proposed to fit their data with a power law $\alpha(f)=\alpha_1 f^n$ (Ref. 19) which indicates no attenuation at zero-frequency but implicitly assumes (a) negligible transmission loss and (b) that the fit can be extrapolated outside the measured frequency range, in particular in the low frequency range. In contrast with these investigators, we have introduced the constant α_0 in the

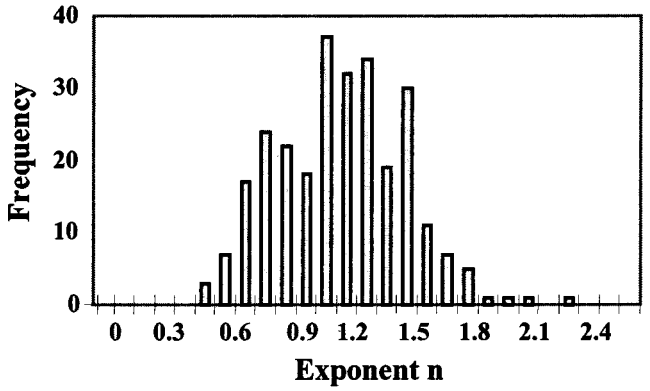


FIG. 7. Distribution of measured n values.

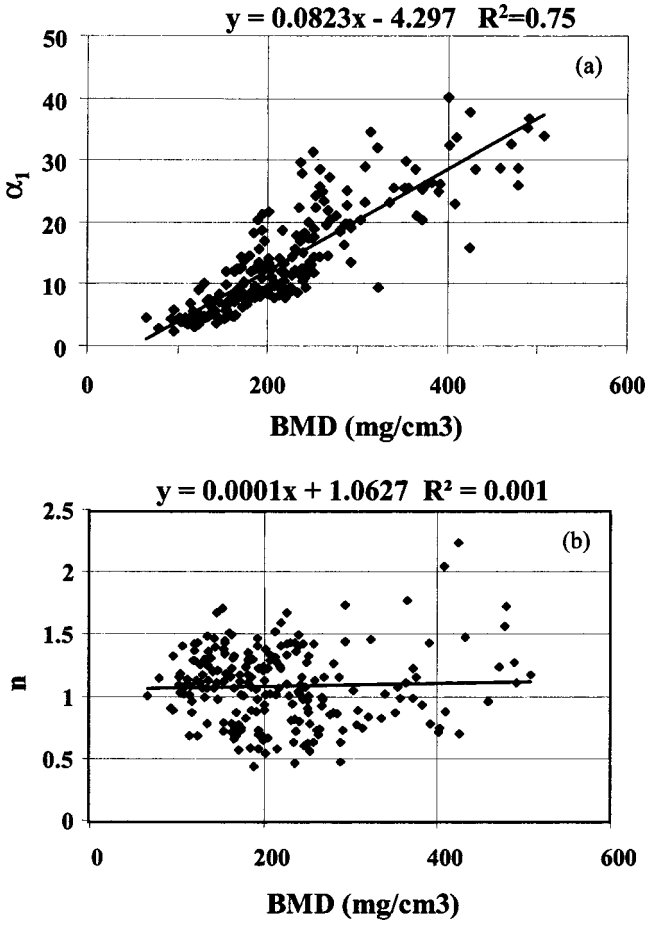


FIG. 8. The relationship between QCT determined BMD and the constants α_1 (a) and n (b). The regression line corresponds to least squares linear fit to all 215 points.

model in order to avoid any assumption about the frequency variation of the attenuation outside the measured frequency bandwidth. Nonlinear power fits were very accurate in predicting experimental data with square correlation coefficients higher than 0.975. In our study, a general trend emerges, showing that the average dependence of the attenuation on frequency ($f^{1.1\pm0.3}$) may reasonably be approximated by a linear fit, which offers some persuasive experimental basis for the current clinical approach. On the other hand, the mean value of n was derived by pooling the data of several local estimates in specimens coming from several donors. As shown in Fig. 7, n demonstrated a wide range of variation between 0.4 and 2.2, reflecting both between specimen variability and spatial heterogeneity (i.e., within specimen variability). The spatial heterogeneity of the calcaneus has been investigated in several previous studies using imaging techniques such as x-ray quantitative computed tomography, quantitative ultrasound,[8] or micro-MRI.[20] Coefficients of variation within specimens (i.e., between ROIs of a given specimen) as high as several tenths of a percent for the slope of the attenuation coefficient and bone mineral density[8] or for structural parameters[20] are in correspondence with the within specimen variability of n reported in this study.

These results we obtained with our bone specimens are important. The fact that the frequency dependence of the attenuation differed from a linear relationship indicates that

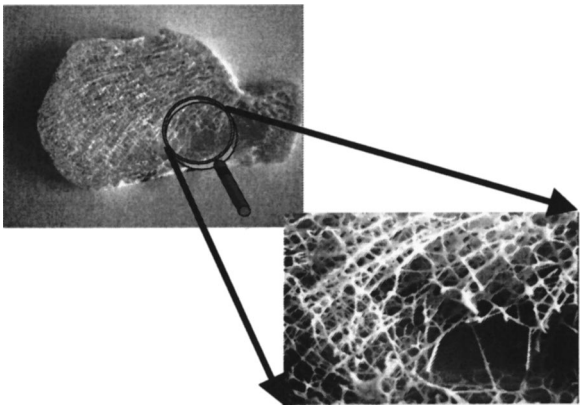


FIG. 9. This figure shows an example of human calcaneus and displays a detail of its internal micro-architecture.

an analysis which assumes such linear behavior *a priori* cannot be applied to all situations. If attenuation measured over a range of frequencies can be described more accurately by a power law curve, then two parameters are now required for bone characterization instead of BUA alone: α_1 the attenuation coefficient at 1 MHz and n the power law dependence. This experimental work was undertaken in order to provide accurate additional data in a broader range of frequencies than the clinically usable bandwidth. We believe that these important additional data may be used to develop appropriate acoustical bone model, which in turn may be used to have a better insight into the interaction between bone and ultrasound.

Different attenuation mechanisms can be invoked including scattering, resonance, visco-elastic absorption, and mode conversion, however, no convincing theoretical approach has been advanced so far. Cancellous bone is a highly porous medium composed of a solid matrix (mineralized collagen) of interconnected plates and rods (trabeculas) filled with fluid-like marrow (Fig. 9). Trabecular elements of average size ranging between 50 μm and 150 μm are separated by an average distance of 0.5 mm to 2 mm. Interpretation of attenuation in cancellous bone is usually based on Biot theory.[21] This theory assumes that the solid frame and the saturating liquid may have relative motion. Within this framework, the attenuation is caused by the viscous friction occurring within the boundary layer at the interface between fluid and solid. Biot[22,23] has defined a low and a high frequency range, depending on the value of the ultrasonic frequency compared to the characteristic frequency of the porous material given below:

$$f_c = \frac{\mu\beta}{2k\pi\rho_f}, \tag{7}$$

where μ is the viscosity of the fluid, β the porosity, k is the permeability, and ρ_f the mass density of the fluid. Using the values for these parameters taken from the works of Hosokawa and Otani,[15] Mc Kelvie and Palmer,[24] and Williams[25] the characteristic frequency for water-saturated cancellous bones is lower than a few tenths of Hz, a value much lower than the characteristic frequency ranging from several kHz to a few MHz commonly found in geophysics.[26] The difference can be explained by the permeability of can-

cellous bone which is much higher than that of rocks. Thus, the frequencies used to characterize bone fall in the very high frequency domain of the Biot theory. Consequently, based on the results of Biot theory in the high frequency domain, the frequencydependent attenuation should be proportional to $\sqrt{f}$.[22] This is in contrast with the results observed for most of the tested ROIs.

Modeling the acoustical behavior of cancellous bone using the Biot theory remains highly controversial. Haire and Langton recently reviewed various available experimental and numerical data.[21] Biot's theory has been found to be successful in predicting the correct values of the speed of sound for both fast and slow longitudinal waves.[27] Given this relatively good agreement with experimental data, we have computed the transmission coefficient at the fluid–bone interface, following the approach of Wu *et al.*[14] assuming a macroscopically homogeneous, isotropic dissipationless, and fluid-saturated medium based on a Biot model. In contrast, predictions of attenuation values in cancellous bone based on Biot theory have been found to disagree with observed values by several investigators: the experimental values were considerably higher than those predicted by Biot theory.[15] The Biot model is very complex, requiring many input parameters that are difficult to measure or are sometimes unknown for cancellous bone. Input of ill-estimated parameters in the Biot model may in turn cause systematical errors on the prediction of the acoustical parameters. The question also arises as to whether the Biot theory is valid for cancellous bone. The model is strictly valid only for porous isotropic and macroscopically homogeneous media. Cancellous bone is definitely a very complicated, anisotropic, and heterogeneous medium, and neither of these conditions are respected. For example, considering the usable frequency range of 0.2–1.7 MHz, the corresponding wavelengths in water ranging from 1 to 7.5 mm are comparable to the size of mean trabecular separation of 0.5–2 mm. This indicates that the assumption of macroscopically homogeneous medium required in the Biot model is not respected.

Nevertheless, in the framework of the Biot theory, the linear frequency dependence of attenuation or the excessive attenuation observed by several investigators compared to numerical predictions based on Biot theory were attributed to several additional dissipation mechanisms. Stoll *et al.* suggested that the linear variation with frequency of attenuation observed in sediments could be explained by viscous attenuation in the solid frame.[28] When sound waves propagate in sands, the relative motion of grains causes friction at their contact surfaces. The solid frame of sands is no longer a perfect elastic medium and complex elastic constants have been introduced by Stoll *et al.* in the model.[28] With this approach, a good agreement was obtained with experimental results. In porous rocks, Gist[29] has revealed the role of microcracks. However, he has reported that this effect is predominant at low porosity, which does not make it a good candidate to account for the high attenuation measured in bones with porosity of 80 to 90% at least. Also, the roughness of the solid frame surface has been shown to play a role in rocks by increasing the surface of friction between fluid and solid.[29] The magnitude of this effect increases when the viscous skin depth becomes small compared to the characteristic dimension of the roughness. In our frequency range, the viscous skin depth in water is 1–0.5 μm. A further investigation into these effects is required to elucidate the contribution of these mechanisms on attenuation in trabecular bone.

Other mechanisms, not included in the Biot theory, may be involved, such as bulk viscous attenuation in the fluid or scattering by inhomogeneities. Camus *et al.* for bones[30] and Gist for sandstone[29] have shown that viscosity of the fluid does not influence the attenuation of high porosity material. Scattering represents an alternative model to the interaction between ultrasound and heterogeneous propagation medium. Wear has recently developed a scattering model in which the trabeculas are approximated with long thin cylinders with a diameter small compared with the ultrasonic wavelength.[31] He reported experimental results in good agreement with theoretical predictions in which scattering should lead to a frequency-dependent attenuation varying as f^3. None of the above attenuation mechanism in itself seems to be able to account for the observed magnitude and frequency variation of attenuation in cancellous bone. One noticeable outcome of this study is the wide range of variation of the power law dependence n. Considering the dramatic spatial heterogeneity of our bone specimens, it is feasible that different mechanisms (or variable mechanisms combination) might explain the attenuation in different parts of bone.

The strong correlation between α_1 and BMD is very similar to that reported previously by our group between BUA (measured in the frequency range 0.2–0.6 MHz) and BMD.[8] In the present study we demonstrate that there was no relationship between n and BMD. This suggests that variations in BMD could not explain alone the observed between specimen and regional variations of n. In contrast, Strelitzki and Evans[18] reported that the departure of the attenuation frequency variation from the linearity seemed to be greatest for low attenuating specimens (i.e., low BMD). The apparent disparity with our results remains unresolved but differences in methods make comparisons somewhat difficult: these include transducer configuration, frequency bandwidth used in the calculations, differences in data analysis performed with a polynomial fit rather than a frequency power law.

In this study we did not directly assess the possibility that the cause of variation in n might be variations in bone microarchitecture. Whether the frequency variation of the attenuation coefficient are attributable to noisy measurements or to variations in micro-architecture (or other bone properties) unrelated to BMD remains to be investigated. If additional information about bone status could be extracted from more complex (than simply linear) curve fitting of the attenuation data, a multivariate approach of QUS that makes use of both attenuation magnitude (i.e., α_1) and frequency dependence (i.e., n) information would lead to further improvement of fracture assessment. Clearly such multi-parametric approaches represent areas that warrant further investigations.

Such a technique would be important to the extent that it can be applied in clinical practice. The problem of transmis-

sion loss factor has to considered carefully for *in vivo* assessment of the frequency dependence of the attenuation. The effect of transmission loss is more complex in *in vivo* measurements than in this experimental study: *in vivo*, many interfaces have to be taken into account, including the interface between water and skin, those between different layers of intervening soft tissues, and the interface between soft tissue and bone. If the component due to transmission loss is frequency independent then it does not contribute to any frequency dependence of apparent attenuation. Its only impact is a positive offset of the apparent attenuation curve versus frequency. However, the assumption that the transmission loss factor is frequency independent is strictly true only if (a) the beam is perpendicular to the interface, (b) the different propagating media are not dispersive, and (c) there is no roughness at the interface. If one or several of these conditions are not true, then the transmission loss factor is frequencydependent and, therefore, influences the frequency dependence of attenuation. Future *in vivo* studies characterizing changes in the frequency dependence of the attenuation arising from biological changes should incorporate considerations of potential frequency dependence of the transmission loss factor.

In addition, as stated above, the reliability of estimates of frequency dependence increases with the frequency bandwidth used for calculations. But, it should be noted that cancellous bone is extremely attenuating and that the limiting factors to these measurements are the high attenuation of cancellous bone which is such that there would be a significant signal-to-noise ratio problem in increasing the frequency bandwidth. Such *in vitro* investigations point to the need to focus in future on progress in the design of very broadband transducers, which, if combined with significant improvement of signal-to-noise ratio, would be useful for selective accurate assessment of the frequency variation of acoustic properties of bone. Thus, this technique may be amenable to *in vivo* measurements to gain additional information which is not measured by established QUS techniques.

[1] C. C. Glüer, ''Quantitative ultrasound techniques for the assessment of osteoporosis: expert agreement on current status,'' J. Bone Miner. Res. **12**, 1280–1288 (1997).

[2] D. Hans, P. Dargent-Moline, A. M. Schott, J. L. Sebert, C. Cormier, P. O. Kotski, P. D. Delmas, J. M. Pouilles, G. Breart, and P. J. Meunier, ''Ultrasonographic heel measurements to predict hip fracture in elderly women: the Epidos prospective study,'' Lancet **348**, 511–514 (1996).

[3] D. C. Bauer, C. C. Glüer, J. A. Cauley, T. M. Vogt, K. E. Ensrud, H. K. Genant, and D. M. Black, ''Broadband ultrasound attenuation predicts fractures strongly and independently of densitometry in older women,'' Arch. Int. Med. **157**, 629–634 (1997).

[4] P. W. Thomson, J. Taylor, R. Oliver, and A. Fisher, ''Quantitative ultrasound (QUS) of the heel predicts wrist and osteoporosis-related fractures in women age 45-75 years,'' J. Clin. Densitometry **1**, 219–225 (1998).

[5] E. W. Gregg, A. M. Kriska, L. M. Salamone, M. M. Roberts, S. J. Anderson, and R. E. Ferrell, ''The epidemiology of quantitative ultrasound: a review of the relationships with bone mass, osteoporosis and fracture risk,'' Osteoporosis Int. **7**, 89–99 (1997).

[6] *Quantitative Ultrasound: Assessment of Osteoporosis and Bone Status*, edited by C. F. Njeh, D. Hans, T. Fuerst, C. C. Glüer, and H. K. Genant (Martin Dunitz, London, 1999).

[7] P. H. F. Nicholson, M. Haddaway, and M. W. J. Davie, ''The dependence of ultrasonic properties on orientation in human vertebral bone,'' Phys. Med. Biol. **39**, 1013–1024 (1994).

[8] P. Laugier, P. Droin, A. M. Laval-Jeantet, and G. Berger, ''*In vitro* assessment of the relationship between acoustic properties and bone mass density of the calcaneus by comparison of ultrasound parametric imaging and QCT,'' Bone (N.Y.) **20**, 157–165 (1997).

[9] P. H. F. Nicholson, R. Müller, G. Lowet *et al.*, ''Do quantitative ultrasound measurement reflect structure independently of density in human vertebral cancellous bone?,'' Bone (N.Y.) **23**, 425–431 (1998).

[10] D. Hans, C. Wu, C. F. Njeh, S. Zhao, P. Augat, D. Newitt, T. Link, Y. Lu, S. Majumdar, and H. K. Genant, ''Ultrasound velocity of trabecular cubes reflects mainly bone density and elasticity,'' Calcif. Tissue Int. **64**, 18–23 (1999).

[11] V. Roberjot, P. Laugier, P. Droin, and G. Berger, ''Measurement of the integrated backscatter coefficient of trabecular bone,'' in *1996 IEEE Ultrasonics Symposium Proceedings*, edited by S. C. Schneider, M. Levy, and B. R. McAvoy (Institute of Electrical and Electronic Engineers, New York, 1996), pp. 1123–1126.

[12] P. Droin, G. Berger, and P. Laugier, ''Dispersion of acoustic waves in cancellous bone,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 581–592 (1998).

[13] S. Chaffaî, F. Peyrin, G. Berger, and P. Laugier, ''Variation of human cancellous bone ultrasonic properties with density and microarchitecture,'' in *1999 IEEE Ultrasonics Symposium Proceedings*, edited by S. C. Schneider, M. Levy, and B. R. McAvoy (Institute of Electrical and Electronic Engineers, New York, in press).

[14] K. Wu, X. Qiang, and L. Adler, ''Reflection and transmission of elastic waves from a fluid-saturated porous solid boundary,'' J. Acoust. Soc. Am. **87**, 2349–2358 (1990).

[15] A. Hosokawa and T. Otani, ''Acoustic anisotropy in bovine cancellous bone,'' J. Acoust. Soc. Am. **103**, 2718–2722 (1998).

[16] *Numerical Recipes in C*, edited by W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (Cambridge University Press, London, 1992).

[17] R. Strelitzki and J. A. Evans, ''An investigation of the measurement of broadband ultrasonic attenuation in trabecular bone,'' Ultrasonics **34**, 785–791 (1996).

[18] R. Strelitzki, S. C. Metcalfe, P. H. F. Nicholson, J. A. Evans, and V. Paech, ''On the ultrasonic attenuation and its frequency dependence in the os calcis assessed with a multielement receiver,'' Ultrasound Med. Biol. **25**, 133–141 (1999).

[19] K. J. Parker, M. S. Asztely, R. M. Lerner, E. A. Schenk, and R. C. Waag, ''*In vivo* measurement of ultrasound attenuation in normal or diseased liver,'' Ultrasound Med. Biol. **14**, 127–136 (1988).

[20] J. C. Lin, M. Amling, D. C. Newitt, K. Selby, S. K. Srivatav, G. Delling, and H. K. Genant, ''Heterogenity of trabecular bone structure in the calcaneus using magnetic resonance imaging,'' Osteoporosis Int. **8**, 16–24 (1998).

[21] T. J. Haire and C. M. Langton, ''Biot theory: A review of its application to ultrasound propagation through cancellous bone,'' Bone (N.Y.) **24**, 291–295 (1999).

[22] M. A. Biot, ''Theory of propagation of elastic waves in a fluid-saturated porous solid. I Low-frequency range,'' J. Acoust. Soc. Am. **28**, 168–178 (1956).

[23] M. A. Biot, ''Theory of propagation of elastic waves in a fluid-saturated porous solid. II Higher frequency range,'' J. Acoust. Soc. Am. **28**, 178–191 (1956).

[24] M. L. Mac Kelvie and S. B. Palmer, ''The interaction of ultrasound with cancellous bone,'' Phys. Med. Biol. **36**, 1331–1340 (1991).

[25] J. L. Williams, ''Ultrasonic wave propagation in cancellous and cortical bone: Prediction of some experimental results by Biot's theory,'' J. Acoust. Soc. Am. **91**, 1106–1112 (1992).

[26] *Acoustics of Porous Media*, edited by T. Bourbié, O. Coussy, and B. Zinszner (Gulf, Houston, TX, 1987).

[27] A. Hosokawa and T. Otani, ''Ultrasonic wave propagation in bovine cancellous bone,'' J. Acoust. Soc. Am. **101**, 1–5 (1997).

[28] R. D. Stoll, *Lecture Notes in Earth Sciences, Vol. 26: Sediments Acoustics* (Springer-Verlag, New York, 1989).

[29] G. A. Gist, ''Fluid effects on velocity and attenuation in sandstones,'' J. Acoust. Soc. Am. **96**, 1158–1173 (1994).

[30] E. Camus, K. Mukuda, T. Otani, and Y. Watanabe, ''Influence of fluid viscosity on the propagation of ultrasonic waves in bovine cancellous bone,'' Technical Report of IEICE, US97-85, EA97-81, January 1998.

[31] K. A. Wear, ''Frequency dependence of ultrasonic backscatter from human trabecular bone: theory and experiment,'' J. Acoust. Soc. Am. **106**, 3659–3664 (1999).

Ultrasound-induced lung hemorrhage is not caused by inertial cavitation

William D. O'Brien, Jr.[a] and Leon A. Frizzell
Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering, University of Illinois, 405 North Mathews, Urbana, Illinois 61801

Ronald M. Weigel and James F. Zachary
Department of Veterinary Pathobiology, University of Illinois, 2001 South Lincoln Avenue, Urbana, Illinois 61802

(Received 15 February 2000; accepted for publication 21 May 2000)

In animal experiments, the pathogenesis of lung hemorrhage due to exposure to clinical diagnostic levels of ultrasound has been attributed to an inertial cavitation mechanism. The purpose of this article is to report the results of two experiments that directly contradict the hypothesis that ultrasound-induced lung hemorrhage is caused by inertial cavitation. Elevated hydrostatic pressure was used to suppress the involvement of inertial cavitation. In experiment one, 160 adult mice were equally divided into two hydrostatic pressure groups (0.1 or 1.1 MPa), and were randomly exposed to pulsed ultrasound (2.8-MHz center frequency, 1-kHz PRF, 1.42-μs pulse duration, 10-s exposure duration). For the two hydrostatic pressure groups (80 mice each), 8 *in situ* peak rarefactional pressure levels were used that ranged between 2.82 and 11.8 MPa (10 mice/group). No effect of hydrostatic pressure on the probability of hemorrhage was observed. These data lead to the conclusion that lung hemorrhage is not caused by inertial cavitation. Also, the higher hydrostatic pressure enhanced rather than inhibited the impact of ultrasonic pressure on the severity (hemorrhage area, depth, and volume) of lesions. These counterintuitive findings were confirmed in a second experiment using a 2×5 factorial design that consisted of two ultrasonic pressure levels and five hydrostatic pressure levels (100 mice, 10 mice/group). If inertial cavitation were the mechanism responsible for lung hemorrhage, then elevated hydrostatic pressures should have resulted in less rather than more tissue damage at each ultrasonic pressure level. This further supports the conclusion that the pathogenesis of ultrasound-induced lung hemorrhage is not caused by inertial cavitation. © *2000 Acoustical Society of America.* [S0001-4966(00)01309-6]

PACS numbers: 43.80.Gx, 43.25.Yw [FD]

INTRODUCTION

The clinical use of diagnostic ultrasound has had a remarkable safety record with no reported adverse effects in human beings. However, concerns for its safety have been raised recently following the publication of experimental findings documenting lung hemorrhage in mice,[1–5] rats,[6] rabbits,[4,5] monkeys,[7] and pigs[8–10] at levels of ultrasound exposure and pulsing conditions consistent with those used for ultrasonography in human beings. Thus there are questions of fundamental clinical significance regarding the safe use of ultrasonography. Can diagnostic ultrasound produce lung damage in human beings and, if so, under what exposure conditions does the damage affect pulmonary function? Credible answers to these questions must come from *in vivo* laboratory animal studies that are focused on the mechanical and biological mechanisms responsible for ultrasound-induced lung hemorrhage. Comparisons of *in vivo* studies on the safety of ultrasound, using animal species with lung structure and function similar and dissimilar to human beings, can provide the scientific basis for extrapolation to human beings.

Exposure to ultrasound can produce thermal and nonthermal effects that alter tissue integrity and cellular function. Heating results from absorption, the transfer of mechanical energy from the ultrasound wave to tissue, and the failure of tissue to dissipate that energy efficiently. A rise in the temperature of several degrees Celsius can result in cell injury ranging from altered function of enzyme systems to coagulation of cellular proteins and cell death. However, several studies have indicated that heating is not responsible for ultrasound-induced lung hemorrhage.[1,2]

Nonthermal effects in lung could occur through both cavitational and noncavitational mechanisms. Inertial cavitation (cavitational mechanism) involves the growth and collapse of small microbubbles (and requires the presence of cavitation nuclei-stabilized gas microbubbles) at ultrasonic pressures typical of medical ultrasound. Other mechanical effects (noncavitational mechanisms) can occur in the presence of larger gas bodies within tissues or in tissues lacking cavitation nuclei. Noncavitational mechanisms include radiation force, radiation torque, and microstreaming and can cause tissue injury due to stress and shearing.[11]

Studies on ultrasound exposure conducted over the past decade have provided indirect evidence implicating cavitation in the pathogenesis of lung hemorrhage.[1,6,7,11] Aeration of the lung is necessary to produce hemorrhage.[2] Because of air in alveoli, it has been hypothesized that small bubbles

[a]Electronic mail: wdo@uiuc.edu

TABLE I. Experiment 1 ultrasonic exposure levels for the 0.1- and 1.1-MPa hydrostatic pressure study. The means of the axial maximum water-based values of the peak rarefactional $p_{r(in\ vitro)}$ and peak compressional $p_{c(in\ vitro)}$ ultrasonic pressures are reported for a hydrostatic pressure of 0.1 MPa. The estimated *in situ* (at the pleural surface) values of the peak rarefactional $p_{r(in\ situ)}$ and peak compressional $p_{c(in\ situ)}$ ultrasonic pressures are reported for both hydrostatic pressure conditions. The Mechanical Index is determined according to the ODS procedures (Ref. 28). The pulse repetition frequency was 1 kHz for all exposure conditions except the sham-exposure conditions for which the PRF was 10 Hz.

Number of mice (0.1-MPa Hydrostatic pressure)	Number of mice (1.1-MPa Hydrostatic pressure)	$p_{r(in\ vitro)}$ (0.1-MPa Hydrostatic pressure) (MPa)	$p_{c(in\ vitro)}$ (0.1-MPa Hydrostatic pressure) (MPa)	$p_{r(in\ situ)}$ (0.1-MPa Hydrostatic pressure) (MPa)	$p_{c(in\ situ)}$ (0.1-MPa Hydrostatic pressure) (MPa)	Mechanical index (0.1-MPa hydrostatic pressure)	$p_{r(in\ situ)}$ (1.1-MPa Hydrostatic pressure) (MPa)	$p_{c(in\ situ)}$ (1.1-MPa Hydrostatic pressure) (MPa)
15 (sham)	15 (sham)	0.28	0.28	0.25	0.26	0.14	[a]	[a]
10	-	3.13	3.46	2.82	3.12	1.54	-	-
10	10	4.19	4.46	3.78	4.02	2.06	3.42	3.64
10	-	5.33	6.24	4.99	5.63	2.72	-	-
10	10	6.42	7.20	5.79	6.50	3.15	5.24	5.88
10	10	7.49	8.35	6.76	7.53	3.67	6.11	6.81
10	10	9.02	10.8	8.14	9.77	4.40	7.37	8.84
10	10	10.2	13.0	9.21	11.8	4.97	8.33	10.6
10	10	11.4	15.1	10.2	13.6	5.51	9.26	12.3
-	10	12.8	19.9	-	-	6.16	10.5	16.3
-	10	14.5	22.9	-	-	6.96	11.8	18.7

[a]Positioning of the ultrasound focal region on the lung surface was done at a hydrostatic pressure of 0.1 MPa prior to the mouse being placed in the hyperbaric chamber. This procedure was required to maintain blinded exposures so that the mouse handler did not know the hydrostatic pressure condition.

(1–5 μm in diameter) may exist in the surfactant layer and serve as cavitation nuclei.[7] If cavitation nuclei are present, then exposure of these gas bubbles to high ultrasonic stresses may result in their violent oscillation and rapid collapse, a process known as inertial cavitation. Under these conditions, the motion of the gas–liquid interface may reach supersonic speeds (producing shock waves) and bubble collapse may generate chemically reactive free radicals, extremely high temperatures, and microjets, which could easily damage the air–blood barrier resulting in lung hemorrhage.[11] As a result, investigators have suggested that inertial cavitation is the mechanism responsible for lung hemorrhage in at least mice,[1] rats,[6] and monkeys,[7] and it is thus important to determine conclusively if inertial cavitation is the mechanism responsible for lung hemorrhage.

The purpose of this article is to report the results of a series of experiments that directly contradict the view that ultrasound-induced lung hemorrhage is caused by inertial cavitation. To test whether inertial cavitation was responsible for lung damage, this study was designed to determine if overpressure (i.e., increased hydrostatic pressure) could suppress lung damage. Increased hydrostatic pressure has been used to suppress cavitation in studies of the biological effects of ultrasound,[3,12–18] as well as to suppress cavitation associated with extracorporeal shock waves.[19–21] The effect of the overpressure may be to reduce or eliminate the negative total pressure during the pulse, which will increase the threshold for inertial cavitation.[22,23] However, overpressures that were small compared to the peak negative pressure associated with lithotripsy pulses have been shown to reduce or eliminate inertial cavitation *in vitro*,[19–21] and it has been suggested that this may be associated with the elimination of cavitation nuclei.[20,21] Regardless of the mechanism, these studies utilized the fact that elevated hydrostatic pressure increases the cavitation threshold, or suppresses cavitation.[24–27]

I. MATERIALS AND METHODS

A. Animal handling and experimental design

Two experiments were conducted. The second experiment was designed to test counterintuitive findings from the first experiment. The experimental protocols were approved by the campus' Laboratory Animal Care Advisory Committee and satisfied all campus and NIH rules for the humane use of laboratory animals. Animals were housed in an AAALAC approved animal facility, placed in groups of four in polycarbonate cages, and provided food and water *ad libitum*.

For experiment 1 (Table I[28]) there were a total of 190 six-to-seven-week-old 27.8 ± 2.1-g female ICR mice (Harlen Sprague Dawley Laboratories, Indianapolis, IN). One hundred and sixty mice were divided ($n = 80$/group) into two hydrostatic pressure groups [0.1 MPa (1 atm) and 1.1 MPa (11 atm)]. Each hydrostatic pressure group was subdivided into eight-ultrasonic pressure groups consisting of ten mice per group. The other 30 mice were divided ($n = 15$/group) into the 2 hydrostatic pressure groups and were the sham-exposed animals.

For experiment 2 (Table II) there were a total of 115 six-to-seven-week-old 26.8 ± 3.1-g female ICR mice (Harlen Sprague Dawley Laboratories, Indianapolis, IN). One hundred mice were divided ($n = 10$/group) into ten groups based on a 2×5 factorial design. There were two *in situ* peak rarefactional pressure groups (6.0 and 10.9 MPa) and five hydrostatic pressure groups (0.1, 0.4, 0.7, 1.0, and 1.3 MPa). The other 15 mice were divided ($n = 3$/group) into the 5 hydrostatic pressure groups and were the sham-exposed animals.

For both experiments, animals were assigned to each hydrostatic pressure group and ultrasonic pressure group at random. The sham-exposed animals were incorporated into the randomization. The individuals involved in animal han-

TABLE II. Experiment 2 ultrasonic exposure levels at the five hydrostatic pressure levels. The means of the axial maximum water-based values of the peak rarefactional $p_{r(in\ vitro)}$ and peak compressional $p_{c(in\ vitro)}$ ultrasonic pressures are reported for a hydrostatic pressure of 0.1 MPa. The estimated *in situ* (at the pleural surface) values of the peak rarefactional $p_{r(in\ situ)}$ and peak compressional $p_{c(in\ situ)}$ ultrasonic pressures are reported at the indicated hydrostatic pressure levels. The Mechanical Index is determined according to the ODS procedures (Ref. 28). The pulse repetition frequency was 1 kHz for all exposure conditions except the sham exposure conditions for which the PRF was 10 Hz.

Number of mice	Hydrostatic pressure (MPa)	$p_{r(in\ vitro)}$[a] (MPa)	$p_{c(in\ vitro)}$[a] (MPa)	$p_{r(in\ situ)}$ (MPa)	$p_{c(in\ situ)}$ (MPa)	Mechanical index[a]
15 (sham)	0.1	0.28	0.28	0.25[b]	0.25[b]	0.14
10	0.1	6.45	7.26	5.95	6.69	3.09
10	0.4	6.70	7.79	5.99	6.96	3.21
10	0.7	7.09	8.60	6.05	7.34	3.39
10	1.0	7.22	8.87	6.07	7.48	3.45
10	1.3	7.22	8.87	6.07	7.48	3.45
10	0.1	11.7	18.3	10.8	16.8	5.52
10	0.4	12.2	19.2	10.9	17.1	5.72
10	0.7	12.8	20.5	10.9	17.5	6.02
10	1.0	13.0	20.9	11.0	17.7	6.11
10	1.3	13.0	20.9	11.0	17.7	6.11

[a]$p_{r(in\ vitro)}$, $p_{c(in\ vitro)}$, and Mechanical Index are reported at 0.1-MPa hydrostatic pressure.
[b]Positioning of the ultrasound focal region on the lung surface was done at a hydrostatic pressure of 0.1 MPa prior to the mouse being placed in the hyperbaric chamber. This procedure was required to maintain blinded exposures so that the mouse handler did not know the hydrostatic pressure condition.

dling, exposure, and lesion scoring were blinded to the exposure conditions. The exposure conditions were revealed only after the final results were tabulated.

Mice were weighed and then anesthetized with ketamine hydrochloride (125.0 mg/kg) and xylazine (25.0 mg/kg) administered intraperitoneally. For each animal, the skin of the left thorax was exposed by removing the hair with an electric clipper, followed by a depilatory agent (Nair®, Carter-Wallace, Inc., New York, NY) to maximize sound transmission. A black dot was placed on the skin at approximately the sixth to ninth rib. The anesthetized animal was placed in a specially designed holder, to which the ultrasonic transducer was attached. A removable pointer, attached to the transducer, was used to position the ultrasonic beam perpendicular to the skin at the position of the black dot and in the beam's focal region.

The holder was placed in a separate degassed, temperature-controlled (30 °C) water tank for transducer positioning. The transducer was not in direct contact with the mouse. Water served as the coupling medium between the ultrasound transducer and the mouse in order to align the transducer's focal region on the pleural surface. The low-power pulse-echo signal of the exposure system (RAM 5000, Ritec, Inc., Warwick, RI) displayed on an oscilloscope was used to adjust the transducer's axial position so that the focal region was within 1 mm of the lung surface. The low-power exposure conditions were as follows: 2.8-MHz center frequency, 10-Hz pulse repetition frequency, 1.42-μs pulse duration, 5–20-s exposure duration (see sham in Tables I and II for these low-level ultrasonic pressure levels). The lung surface provided distinctive echo characteristics that allowed for the precise positioning of the focal region of the ultrasound beam on the pleural surface. The holder was then removed from this separate water tank and placed in a specially fabricated hyperbaric chamber,[12,16,17] and the animal was exposed to pulsed ultrasound (2.8-MHz center frequency, 1-kHz pulse repetition frequency, 1.42-μs pulse duration, 10-s exposure duration). For animals exposed at the elevated hydrostatic pressure, the hydrostatic pressure was increased to its maximum level in about 6 min and then decreased back to 0.1 MPa in about 4 min, a procedure that was adhered to for all animals. Following the ultrasound exposure procedure, the animal was removed from the chamber and holder, and euthanized under anesthesia by cervical dislocation.

The 2.8-MHz center frequency was used since it was within the lower end of the diagnostic ultrasound frequency range. The 10-s exposure duration was used to simulate incidental exposure to lung tissue since, in clinical practice, the lung is generally not intentionally exposed to diagnostic ultrasound.

The thorax was opened, and the thickness of each left thoracic wall (skin, rib cage, and parietal pleura) at the point of exposure was measured (experiment 1: 2.89 ± 0.90 mm for all 190 mice; experiment 2: 2.32 ± 0.36 mm for all 115 mice) using a digital micrometer (accuracy: 10 μm). These chest wall measurements were used for later calculation of the *in situ* ultrasonic pressures at the visceral pleural surface (Tables I and II). The lungs were removed from each animal, and the left lung lobe was scored for the presence or absence of hemorrhage. Lung hemorrhage formed along the pathway of the ultrasound beam, and the lesion assumed a conical shape. The base of the lesion originated at the visceral pleural surface and was elliptical in shape. The lesion extended into lung parenchyma to form its apex at varied depths within the lung. The left lung was fixed by immersion in 10% neutral-buffered formalin for a minimum of 24 h. After fixation, the elliptical dimensions of each lung lesion at the visceral pleural surface were measured using a digital micrometer where ''*a*'' is the semi-major axis and ''*b*'' is the semi-minor axis. The lesion was then bisected and the depth

''*d*'' of the lesion within the pulmonary parenchyma was also measured. The surface area (πab) and volume ($\pi abd/3$) of the lesion were calculated for each animal. Each half of the bisected lesion was embedded in paraffin, sectioned at 5 μm, stained with hematoxylin and eosin, and evaluated microscopically.

B. Transducer characteristics and ultrasound field calibrations

Ultrasonic exposures were conducted for both experiments using a focused, 19-mm-diameter, $f/1$, lithium niobate ultrasonic transducer (Valpey Fisher, Hopkinton, MA). Pulse-echo field distribution measurements[29] in degassed water yielded the following transducer characteristics: 2.8-MHz center frequency, 11.6% fractional bandwidth, 18.9-mm focal length, 466-μm, 6-dB beamwidth at the focus, 2.73-mm, 6-dB depth of focus.

A special procedure was developed to routinely calibrate the ultrasound fields in degassed water. This procedure was based on accepted national standards.[28,30] The source transducer's drive voltage was supplied by Ritec's RAM5000 that has the capability to deliver up to a 5-kW single-cycle pulse into a 50-ohm load. An automated search procedure was developed to determine the beam axis at each drive voltage.[31] This search procedure was based on the maximum value of the pulse intensity integral (PII). The maximum PII value was determined at various axial positions using a computer-controlled micropositioning system (three orthogonal axes, each with a linear accuracy of 2 μm). The coordinates of these maximum PII values were used to determine the beam axis using a linear regression procedure. The same calibrated hydrophone used to determine the beam axis was scanned along the beam axis and through the focal region at 50-μm intervals. At each interval, the rf hydrophone waveform was digitized (500 Ms/s, LeCroy Model 9354TM, Chestnut Ridge, NY). These digitized hydrophone waveforms were processed off-line (Matlab®, Natick, MA) on a Sun UltraSparc workstation to yield the following axial profiles: rarefactional pressure, compressional pressure, pulse intensity integral, and their respective derated (0.3-dB/cm MHz) profiles. Twelve independent axial calibrations were conducted over the two-month period of the experiments using two calibrated PVDF membrane hydrophones (Sonic Technologies Model 804-010, Hatboro, PA and Marconi Model Y-34-6543, Chelmsford, U.K.); standard deviations were less than 15% of their respective mean values. The water-based exposure quantities determined from the calibrated axial profiles were the maximum values of the peak rarefactional $p_{r(in\ vitro)}$ and peak compressional $p_{c(in\ vitro)}$ ultrasonic pressures (Tables I and II). The axial locations of $p_{c(in\ vitro)}$ and $p_{r(in\ vitro)}$ were determined, and their axial differences ranged from about 10 μm for the sham-exposure level to about 800 μm for the highest exposure level. The Mechanical Index was calculated using the ODS procedure[28] from $p_{r.3}/\sqrt{f_c}$, where $p_{r.3}$ is the peak rarefactional pressure derated by 0.3 dB/cm MHz at the location where the derated pulse intensity integral $PII_{0.3}$ is a maximum and f_c is the center frequency of 2.8 MHz.

The *in situ* (at the pleural surface) peak rarefactional pressure levels were estimated from

$$p_{r(in\ situ)} = T \cdot p_{r(in\ vitro)} e^{-(A \cdot x)}, \tag{1}$$

where $p_{r(in\ vitro)}$ is the maximum peak rarefactional pressure at 0.1-MPa hydrostatic pressure, ''*A*'' is the mean attenuation coefficient of the chest wall (3.10 ± 0.86 dB/cm at 2.8 MHz; Ref. 32) measured from 35 separate chest walls using a broadband through-transmission insertion loss technique,[33] and ''*x*'' is the mean chest wall thickness (2.89 ± 0.90 mm). The transducer's relative pressure efficiency ''*T*'' was determined by operating the Ritec's RAM5000 in its pulse-echo mode. The 2.8-MHz transducer was placed in the hyperbaric chamber, and the echo amplitude reflected from a stainless steel reflector (oriented normal to the beam axis with the focus at the reflector surface) was measured as a function of hydrostatic pressure (from 0.1 to 1.45 MPa in 0.15-MPa increments). Five Ritec drive voltage levels were used that bracketed the ultrasonic pressure levels of both animal experiments. The pressure efficiency experiment was conducted on two different days, and there was no apparent difference between the two data sets. Therefore, the pulse-echo amplitude results at each drive voltage were averaged for each hydrostatic pressure level. The pulse-echo amplitude results as a function of hydrostatic pressure were normalized to that at a hydrostatic pressure of 0.1 MPa to yield the squared relative pressure efficiency T^2. For experiment 1, two hydrostatic pressures were used for which $T = 1.00$ (0.1 MPa) and 0.91 (1.1 MPa). For experiment 2, five hydrostatic pressures were used for which $T = 1.00$ (0.1 MPa), 0.97 (0.4 MPa), 0.93 (0.7 MPa), 0.91 (1.0 MPa), and 0.91 (1.3 MPa). The results reported herein are in terms of $p_{r(in\ situ)}$ (Tables I and II).

II. RESULTS

Sham-exposed mice were included in the randomized designs of both experiments. These mice received low-level ultrasonic exposure (see Tables I and II) during the beam alignment procedure. Thirty sham-exposed mice were included in the first experiment, fifteen for each hydrostatic pressure condition. Fifteen sham-exposed mice were included in the second experiment, three for each hydrostatic pressure condition. In all cases, none of the sham-exposed mice demonstrated lung hemorrhages.

Data analyzed for each animal included the presence or absence of a lesion, the lesion surface area, and the lesion depth. Lesion volume was not statistically analyzed as an outcome because it did not provide information independent of lesion area and depth. The effects of $p_{r(in\ situ)}$ and hydrostatic pressure, and the interaction of these factors, upon the presence or absence of a lesion (percentage of animals with lung lesions) were analyzed using multiple logistic regression analysis[34] (Fig. 1). There was no significant difference in percentage of lesions between the two hydrostatic pressure groups.

The effects of $p_{r(in\ situ)}$ and hydrostatic pressure, and the interaction of these factors, upon lesion surface area and depth were analyzed using multiple linear regression analysis[35] (Fig. 2). At low ultrasonic pressures, lesions were

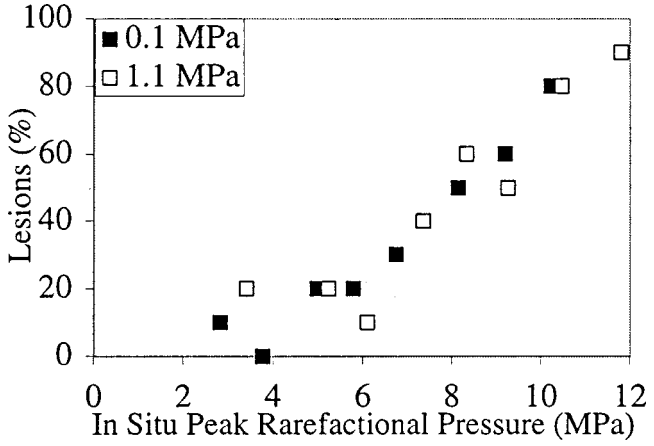


FIG. 1. Percentage of lesions ($n=10$/group) as a function of $p_{r(in\ situ)}$ at hydrostatic pressures of 0.1 and 1.1 MPa for experiment 1. The multiple logistic regression analysis (model $\chi^2=101.4$; $p<0.0001$) indicates there was no apparent effect of hydrostatic pressure on whether or not a lesion occurred. However, increased $p_{r(in\ situ)}$ was associated with an increased probability of a lesion occurring ($p<0.0001$). There was no apparent interaction between the experimental factors.

nonexistent or small, and there were no significant effects due to hydrostatic pressure. As $p_{r(in\ situ)}$ increased, lesion area and depth increased. Increase in lesion area with increased $p_{r(in\ situ)}$ occurred at a greater rate at 1.1-MPa hydrostatic pressure than at 0.1-MPa hydrostatic pressure. With respect to lesion depth, there was no significant effect of hydrostatic pressure or an interaction of $p_{r(in\ situ)}$ with hydrostatic pressure. These data indicate that increases in hydrostatic pressure enhance (rather than inhibit) lesion surface area; however, hydrostatic pressure did not appear to modify the effect of $p_{r(in\ situ)}$ in penetrating lung tissue.

Based on the results of experiment 1, a second experiment was conducted to confirm the counterintuitive findings. The second experiment had a 2×5 factorial design (two $p_{r(in\ situ)}$ levels and five hydrostatic pressure levels). One of the $p_{r(in\ situ)}$ levels (6.0 MPa) was selected to be near but slightly greater than the 5.5-MPa $p_{r(in\ situ)}$ crossover point of the two regression lines from the first study [Figs. 2(A) and 2(B)], where no substantial hydrostatic pressure effect was noted. The other $p_{r(in\ situ)}$ level (10.9 MPa) was selected to be much higher, where the hydrostatic pressure level was observed to affect lesion size. The water-based calibrated value of $p_{r(in\ vitro)}$ was varied in order to keep $p_{r(in\ situ)}$ constant at either 6.0 MPa or 10.9 MPa for each of the hydrostatic pressure levels.

If the relationships identified in experiment 1 were replicated, then lesion depth and surface area would increase with higher $p_{r(in\ situ)}$, and for lesion area these differences would be predicted to be greater as hydrostatic pressures increased. This prediction was supported in part. The results of experiment 2 confirmed an increase in the probability of a lesion, lesion area, and lesion depth with increasing ultrasonic pressure. There were new findings in the second experiment. Increased hydrostatic pressure was associated with increased probability of a lesion (Fig. 3) and increased lesion depth (Fig. 4). However, in contrast to the first experiment, there was no interaction between $p_{r(in\ situ)}$ and hydrostatic

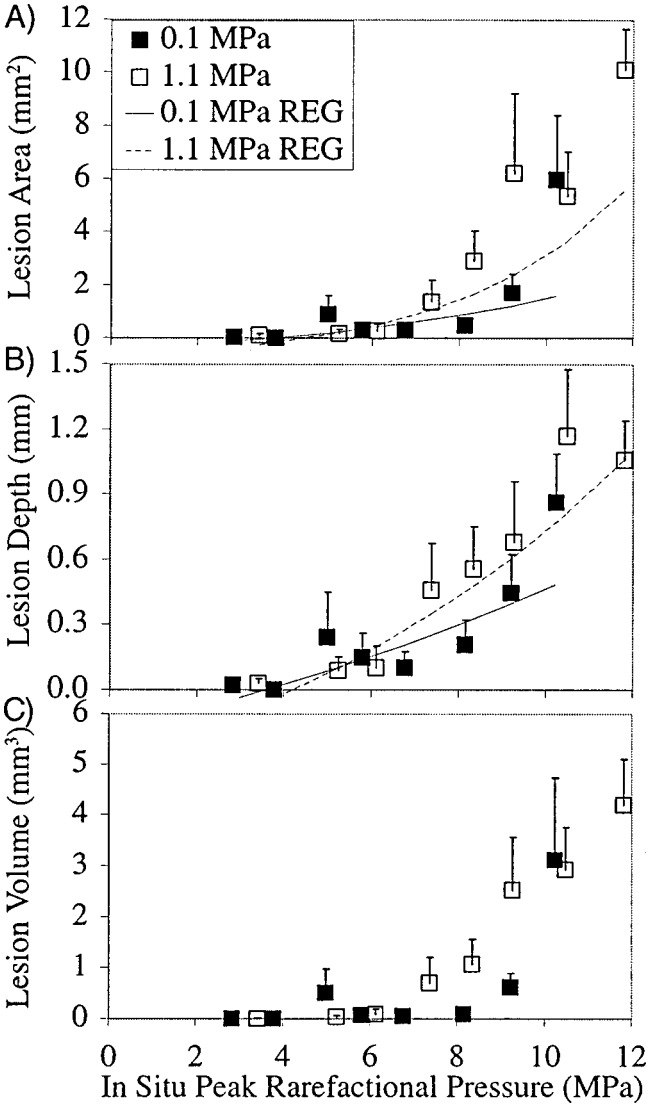


FIG. 2. Mean lesion area (A), depth (B), and volume (C) as a function of $p_{r(in\ situ)}$ at hydrostatic pressures of 0.1 and 1.1 MPa for experiment 1. Error bars (shown in only one direction) represent standard error. The fitted lines are regression functions determined from natural log transformation [ln(measure+1)] of lesion area and depth as a function of $p_{r(in\ situ)}$ at hydrostatic pressures of 0.1 (solid line) and 1.1 (dashed line) MPa for all 160 mouse lungs (16 groups with 10 mice/group). The logarithmic transformation of lesion area and depth was conducted to correct for nonnormality and heterogeneity of residual variance. The linear regression analysis indicated there was an interaction between $p_{r(in\ situ)}$ and hydrostatic pressure in affecting lesion surface area ($p=0.017$). There was no significant interaction between $p_{r(in\ situ)}$ and hydrostatic pressure in affecting lesion depth ($p=0.12$). Lesion depth increased with increasing ultrasonic pressure ($p<0.0001$). [Model $R^2=0.39$ for surface area; $=0.31$ for depth.]

pressure, although the main effect of $p_{r(in\ situ)}$ in increasing lesion area was still apparent (Fig. 4). This lack of interaction between $p_{r(in\ situ)}$ and hydrostatic pressure suggests that the effects are additive from the individual observations ($p_{r(in\ situ)}$ and/or hydrostatic pressure).

III. DISCUSSION

A considerable amount of work has been published regarding lung hemorrhage caused by ultrasound.[11] There is agreement that gas in the lung plays a role in the ultrasound-induced damage mechanism, and that the mechanism is non-

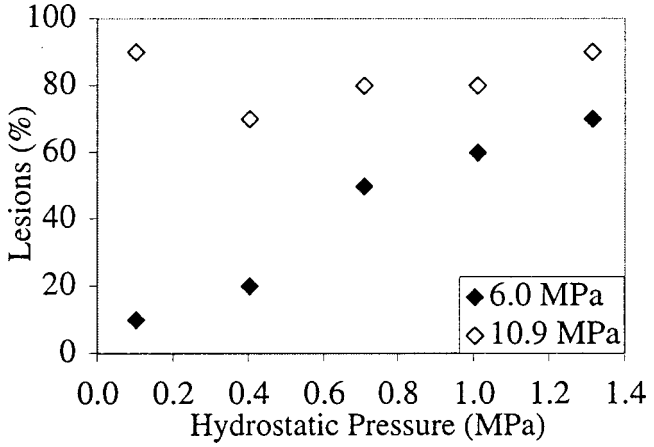


FIG. 3. Percentage of lesions ($n=10$/group) as a function of hydrostatic pressure at $p_{r(in\ situ)}$ values of 6.0 and 10.9 MPa for experiment 2. The multiple logistic regression analysis (model $\chi^2=35.7$; $p<0.0001$) indicates that the probability of a lesion increased with increasing $p_{r(in\ situ)}$ ($p<0.0001$) and with increasing hydrostatic pressure ($p=0.004$), but there was no significant interaction effect for these experimental factors ($p=0.066$).

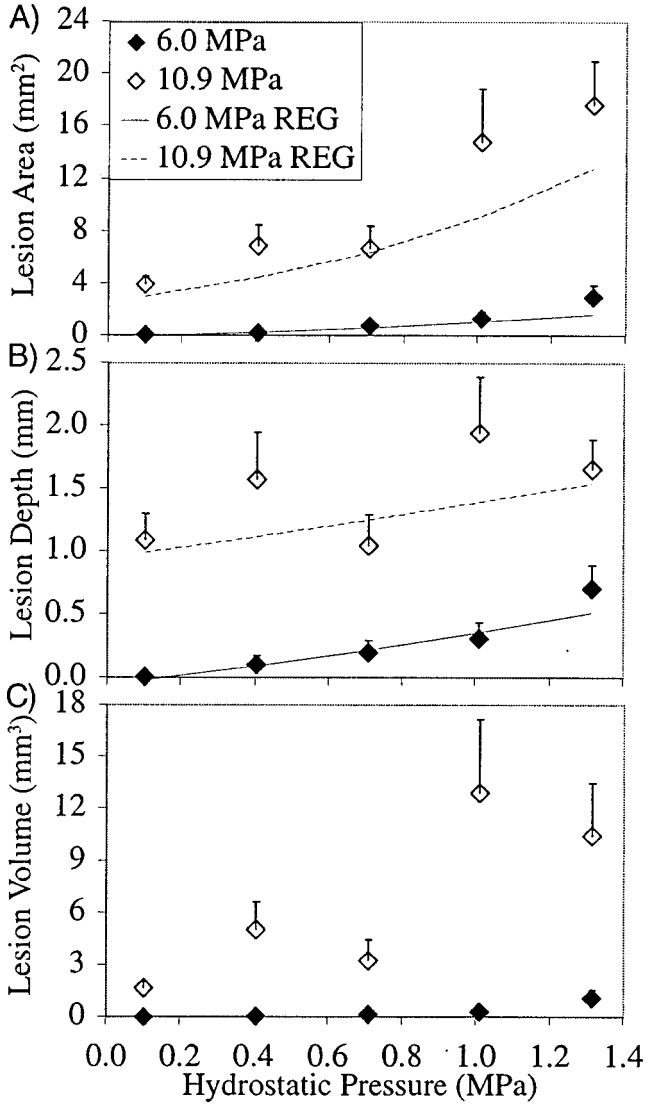


FIG. 4. Mean lesion area (A), depth (B), and volume (C) as a function of hydrostatic pressure at $p_{r(in\ situ)}$ values of 6.0 and 10.9 MPa for experiment 2. Error bars (shown in only one direction) represent standard error. The fitted lines are regression functions determined from natural log transformation [ln(measure+1)] of lesion area and depth as a function of hydrostatic pressures at $p_{r(in\ situ)}$ values of 6.0 (solid line) and 10.9 (dashed line) MPa for all 100 mouse lungs (10 groups with 10 mice/group). The logarithmic transformation of lesion area and depth was conducted to correct for non-normality and heterogeneity of residual variance. The linear regression analysis indicated that both increasing $p_{r(in\ situ)}$ and hydrostatic pressure increased the lesion area [$R^2=0.57$], and lesion depth [$R^2=0.47$] [both p values <0.0001]. However, the interaction effect between $p_{r(in\ situ)}$ and hydrostatic pressure in affecting lesion surface area was not apparent. Ultrasound exposure contributed most to the lesion surface area (accounting for 45% of the variance in surface area and 41% of the variance in lesion depth).

thermal. However, a distinction must be made between mechanisms involving large gas bodies, such as gas in the alveoli of the lung (38–49 μm; Refs. 36–38), and classical inertial cavitation that involves small microbubbles as nuclei (radii on the order of 1 μm or less; Ref. 39). Evidence has been slowly accumulating that suggests that the mechanism of damage in the lung is not inertial cavitation. There seems to be no dependence on whether the negative or positive pressure components of the ultrasonic pulse cause lithotripter-induced lung damage whereas inertial cavitation is associated with the negative pressure.[40] The frequency dependence may not be the same as that associated with effects due to the presence of contrast agents that quite clearly nucleate inertial cavitation.[41]

These two experiments using overpressure to suppress inertial cavitation indicate quite conclusively that inertial cavitation is not responsible for the lung damage. If inertial cavitation were the responsible mechanism, then the higher hydrostatic pressure results would have resulted in lower rather than higher tissue damage at each $p_{r(in\ situ)}$ level. The interaction identified in this study indicates that the mechanism is more complex. At low $p_{r(in\ situ)}$ (around 5.5 MPa), there was no apparent effect of hydrostatic pressure on lung hemorrhage. As $p_{r(in\ situ)}$ increased, lesions became more common and larger in area and depth, and hence volume. At these higher ultrasonic pressures, a modifying effect of hydrostatic pressure became apparent. However, instead of inhibiting lung hemorrhage, higher hydrostatic pressure enhanced damage (as indicated by higher lesion areas, depths, and volumes) (Figs. 2 and 4), with this enhancement of hemorrhage increasing with increasing $p_{r(in\ situ)}$.

The absence of lung damage at low $p_{r(in\ situ)}$ levels for both hydrostatic pressures from experiment 1 (Figs. 1 and 2) suggests there is a threshold for ultrasound-induced lung hemorrhage that is independent of hydrostatic pressure. The imperceptible differences in the level of damage at the lowest $p_{r(in\ situ)}$ levels, however, makes the determination of a threshold (ultrasonic pressure level at which there is no effect) difficult. In any case, an approximate range of thresholds can be estimated from the results reported herein. It must be emphasized, however, that the basic experimental design was not intended to identify definitive thresholds. Best-fit straight lines for the two hydrostatic pressure data sets (using eight values each of Fig. 1) yielded $p_{r(in\ situ)}$ intercepts of 3.1 and 3.0 MPa for 0.1- and 1.1-MPa hydrostatic pressures, respectively. It is clear from Fig. 1 that the $p_{r(in\ situ)}$ threshold is less than 3 MPa and therefore an esti-

mated threshold range around 2.0 MPa seems more reasonable. This value is slightly greater than the values reported by others[1,42] with adult mice near this ultrasonic frequency using longer exposure durations.

The results shown in Fig. 1 indicate that the $p_{r(in\ situ)}$ threshold for hemorrhage occurrence was approximately the same at hydrostatic pressures of 0.1 and 1.1 MPa. This finding leads to the conclusion that the initiation of lung hemorrhage is not caused by inertial cavitation. Even though the hydrostatic pressures used were not large enough to completely eliminate a negative total pressure, based on previous work *in vitro*[19] and *in vivo*,[3] these levels were sufficient to significantly change the threshold for effects of inertial cavitation, should it occur.

It is interesting that the hemorrhage areas, depths, and volumes were greater under increased hydrostatic pressure, which initially seems counterintuitive. Recent studies[20,21] have shown that increased hydrostatic pressure can increase the damage caused by cavitation on metal foils where cavitation nuclei are trapped in crevices. However, these same studies showed that inertial cavitation effects on cells were eliminated by overpressure in the absence of the metal foils. These data lead us to suggest that some other phenomenon must be responsible for the enhanced effects on hemorrhage at the higher hydrostatic pressure.

The change in acoustic impedance difference between intercostal tissue and lung was evaluated as a possible explanation for the enhanced lung damage with increased hydrostatic pressure. A planar boundary was assumed between intercostal tissue and lung with the incident ultrasonic field in the intercostal tissue and the ultrasonic beam axis normal to the boundary. Lung was modeled as two components consisting of air and parenchyma where density

$$\rho_{\text{lung}} = x_{\text{air}}\rho_{\text{air}} + x_{\text{parenchyma}}\rho_{\text{parenchyama}} \tag{2}$$

and adiabatic bulk modulus

$$B_{\text{lung}} = x_{\text{air}}B_{\text{air}} + x_{\text{parenchyama}}^{3.5}B_{\text{parenchyama}}, \tag{3}$$

and where the volume fractions are $x_{\text{air}} + x_{\text{parenchyma}} = 1$. This model was selected because it fit the experimental measurements of reflection coefficient versus lung inflation in the fixed lungs of dogs at 2.4 MHz[43] and the experimental measurements of propagation speed at one lung inflation in fresh lungs of dogs at 2.25 MHz.[44] These published measurements, of course, were all conducted at atmospheric pressure (0.1 MPa). The fit to these published measurements yielded $\rho_{\text{air}} = 1.21\ \text{kg/m}^3$, $\rho_{\text{parenchyma}} = 600\ \text{kg/m}^3$, $B_{\text{air}} = 142\ \text{kPa}$, and $B_{\text{parenchyma}} = 1\ \text{GPa}$ for $x_{\text{air}} = 0.31$, and, in turn, yielded $\rho_{\text{lung}} = 414\ \text{kg/m}^3$ and $B_{\text{lung}} = 273\ \text{MPa}$. Propagation speed is determined from $c_{\text{lung}} = \sqrt{B_{\text{lung}}/\rho_{\text{lung}}}$, and impedance from $z_{\text{lung}} = \sqrt{\rho_{\text{lung}}B_{\text{lung}}}$. These fit values agreed well with published results. The fit reflection coefficient value relative to 1.5 Mrayl was -7.7 dB; the published values were between -2 and -4 dB.[43] The fit and published[44] lung propagation speed values were 812 m/s. As hydrostatic pressure increased from 0.1 to 1.3 MPa, ρ_{air} increased from 1.21 to 15.7 kg/m^3 and B_{air} from 0.142 to 1.85 MPa (ideal gas law), but ρ_{lung} increased slightly from 414 to 419 kg/m^3 and c_{lung} decreased slightly from 812 to 808 m/s which resulted in only a slight increase in z_{lung} from 336 to 338 krayl. Since the intercostal tissue is reasonably modeled as incompressible, it is shown that there is essentially no change in the reflection and transmission coefficients as a function of hydrostatic pressure at a constant lung inflation (constant $x_{\text{air}} = 0.31$). It is interesting to note that under these conditions, the sound power transmission coefficient is 41%, indicating that 41% of the incident power at the intercostal tissue-lung boundary is transmitted into lung. Thus as a function of hydrostatic pressure, this impedance difference alone cannot explain the enhanced effects on hemorrhage at the higher hydrostatic pressure.

If the mouse's breathing pattern was somehow altered as a function of hydrostatic pressure, and this alteration affected the volume of air inspired and expired, a supposition, then an increase in the power transmitted into lung might occur. From the above two-component lung model at 0.1-MPa hydrostatic pressure, if the volume fraction of air x_{air} varied between 0.25 and 0.40, then the respective sound power transmission coefficient would vary between 66% and 48%. If the increased hydrostatic pressure caused the mouse to expire more air, and hence result in $x_{\text{air}} < 0.25$, then the sound power transmission coefficient would be greater than 66% at maximum expiration, i.e., at $x_{\text{air}} = 0.20$, 72% of the incident power at the intercostal tissue-lung boundary would be transmitted into lung. For this supposition to account for the enhanced lung damage at increased hydrostatic pressure, the argument would also have to include a causal relationship between the sound power transmission coefficient and lung damage, one that is not yet available.

In summary, the results of these two experiments indicate that the pathogenesis of ultrasound-induced lung hemorrhage is not caused by inertial cavitation, because lung damage is not inversely correlated with hydrostatic pressure. Also, there is no significant change of the acoustic properties of lung tissue as a function of hydrostatic pressure to account for the enhanced effects of hemorrhage at elevated hydrostatic pressure.

ACKNOWLEDGMENTS

We thank G. Altmeyer, J. Blue, K. Clemens, B. McNeill, R. McQuinn, R. Miller, B. Qiao, J. Sempsrott, K. Norrell, G. Teotico, and A. Wunderlich for technical contributions. This work was supported by NIH Grant No. HL58218 awarded to WDO and JFZ.

[1] S. Z. Child, C. L. Hartman, L. A. Schery, and E. L. Carstensen, ''Lung damage from exposure to pulsed ultrasound,'' Ultrasound Med. Biol. **16**, 817–825 (1990).
[2] C. L. Hartman, S. Z. Child, R. Mayer, E. Schenk, and E. L. Carstensen, ''Lung damage from exposure to the fields of an electrohydraulic lithotripter,'' Ultrasound Med. Biol. **16**, 675–683 (1990).
[3] L. A. Frizzell, E. Chen, and C. Lee, ''Effects of pulsed ultrasound on the mouse neonate: Hind limb paralysis and lung hemorrhage,'' Ultrasound Med. Biol. **20**, 53–63 (1994).
[4] J. F. Zachary and W. D. O'Brien, Jr., ''Lung hemorrhage induced by continuous and pulse wave ultrasound in mice, rabbits, and pigs,'' Vet. Pathol. **32**, 43–54 (1995).
[5] W. D. O'Brien, Jr. and J. F. Zachary, ''Lung damage assessment from exposure to pulsed-wave ultrasound in rabbit, mouse, and pig,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 473–485 (1997).

[6] C. K. Holland, C. X. Deng, R. E. Apfel, J. L. Alderman, L. A. Fernandez, and K. J. Taylor, ''Direct evidence of cavitation in vivo from diagnostic ultrasound'' Ultrasound Med. Biol. **22**, 917–925 (1996).
[7] A. F. Tarantal and D. R. Canfield, ''Ultrasound-Induced lung hemorrhage in the monkey,'' Ultrasound Med. Biol. **20**, 65–72 (1994).
[8] G. H. Harrison, H. A. Eddy, J.-P. Wang, and F. Z. Liberman, ''Microscopic lung alterations and reduction of respiration rate in insonated anesthetized swing,'' Ultrasound Med. Biol. **21**, 981–983 (1995).
[9] R. Baggs, D. P. Penney, C. Cox, S. Z. Child, C. H. Raeman, D. Dalecki, and E. L. Carstensen, ''Thresholds for ultrasonically induced lung hemorrhage in neonatal swine,'' Ultrasound Med. Biol. **22**, 119–128 (1996).
[10] D. Dalecki, S. Z. Child, C. H. Raeman, C. Cox, and E. L. Carstensen ''Ultrasonically induced lung hemorrhage in young swine,'' Ultrasound Med. Biol. **23**, 777–781 (1997).
[11] *Mechanical Bioeffects from Diagnostic Ultrasound: AIUM Consensus Statements* (American Institute of Ultrasound in Medicine, Laurel, MD, 2000). Also, J. Ultrasound Med. **19**, 67–168 (2000).
[12] W. J. Fry, D. Tucker, F. J. Fry, and V. J. Wulff, ''Physical factors involved in ultrasonically induced change in living systems: II. Amplitude duration relations and the effect of hydrostatic pressure for nerve tissue,'' J. Acoust. Soc. Am. **23**, 364–368 (1951).
[13] C. R. Hill, ''Ultrasonic exposure thresholds for changes in cells and tissues,'' J. Acoust. Soc. Am. **52**, 667–672 (1972).
[14] V. Ciaravino, M. W. Miller, and E. L. Carstensen, ''Pressure-mediated reduction of ultrasonically induced cell lysis,'' Radiat. Res. **88**, 209–213 (1981).
[15] K. I. Morton, G. R. terHaar, I. J. Statford, and C. R. Hill, ''Subharmonic emission as an indicator of ultrasonically-induced biological damage,'' Ultrasound Med. Biol. **9**, 629–633 (1983).
[16] C. S. Lee and L. A. Frizzell, ''Exposure levels for ultrasonic cavitation in the mouse neonate,'' Ultrasound Med. Biol. **14**, 735–742 (1988).
[17] L. A. Frizzell, C. S. Lee, P. D. Aschenbach, M. J. Borrelli, R. S. Morimoto, and F. Dunn, ''Involvement of ultrasonically induced cavitation in the production of hind limb paralysis of the mouse neonate,'' J. Acoust. Soc. Am. **74**, 1062–1065 (1983).
[18] E. C. Everbach, I. R. Makin, M. Azadniv, and R. S. Meltzer, ''Correlation of ultrasound-induced hemolysis with cavitation detector output in vitro,'' Ultrasound Med. Biol. **23**, 619–624 (1997).
[19] M. Delius, ''Minimal static excess pressure minimizes the effect of extracorporeal shock waves on cells and reduces it on gallstones,'' Ultrasound Med. Biol. **23**, 611–617 (1997).
[20] J. A. McAteer, M. A. Stonehill, K. Colmenares, J. C. Williams, A. P. Evan, R. O. Cleveland, M. R. Bailey, and L. A. Crum, ''SWL cavitation damage *in vitro*: Pressurization unmasks a differential response of foil targets and isolated cells,'' Proc. 16th Intl. Congress on Acoustics, pp. 2497–2498 (1998).
[21] R. O. Cleveland, M. R. Bailey, L. A. Crum, M. A. Stonehill, J. C. Williams, and J. A. McAteer, ''Effect of overpressure on dissolution and cavitation of bubbles stabilized on a metal surface,'' Proc. 16th Intl. Congress on Acoustics, pp. 2499–2500 (1998).
[22] R. E. Apfel, ''Possibility of microcavitation from diagnostic ultrasound,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **UFFC-33**, 139–142 (1986).
[23] C. K. Holland and R. E. Apfel, ''An improved theory for the prediction of microcavitation due to pulsed ultrasound,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **UFFC-36**, 204–208 (1989).
[24] E. N. Harvey, ''Sonoluminscence and sonic chemiluminescence,'' J. Am. Chem. Soc. **61**, 2392–2398 (1939).
[25] R. D. Finch, ''The dependence of sonoluminescence on static pressure,'' Br. J. Appl. Phys. **16**, 1543–1553 (1965).
[26] P. K. Chendke and H. S. Fogler, ''Sonoluminescence and sonochemical reactions of aqueous carbon tetrachloride solutions,'' J. Phys. Chem. **87**, 1362–1369 (1983).
[27] J. Sponer, ''Dependence of ultrasonic cavitation threshold on raised ambient pressure,'' Stud. Biophys. **137**, 91–97 (1990).
[28] *Standard for the Real-Time Display of Thermal and Mechanical Acoustic Output Indices on Diagnostic Ultrasound Equipment* (American Institute of Ultrasound in Medicine, Laurel, MD, 1992, rev. 1996).
[29] K. Raum and W. D. O'Brien, Jr., ''Pulse-echo field distribution measurement technique of high-frequency ultrasound sources,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 810–815 (1997).
[30] *Acoustic Output Measurement and Labeling Standard for Diagnostic Ultrasound Equipment* (American Institute of Ultrasound in Medicine, Laurel, MD, 1992).
[31] J. M. Sempsrott and W. D. O'Brien, Jr., ''Experimental verification of acoustic saturation,'' Proceedings of the 1999 IEEE Ultrasonics Symposium, pp. 1287–1290 (1999).
[32] G. A. Teotico, R. J. Miller, L. A. Frizzell, J. F. Zachary, and W. D. O'Brien, Jr., ''Attenuation coefficient estimates of mouse and rat chest wall,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control (in press).
[33] E. L. Madsen, F. Dong, G. R. Frank, B. S. Gara, K. A. Wear, T. Wilson, J. A. Zagzebski, H. L. Miller, K. K. Shung, S. H. Wang, E. J. Feleppa, T. Liu, W. D. O'Brien, Jr., K. A. Topp, N. T. Sanghvi, A. V. Zaitsen, T. J. Hall, J. B. Fowlkes, O. D. Kripfgans, and J. G. Miller, ''Interlaboratory comparison of ultrasonic backscatter, attenuation and speed measurements,'' J. Ultrasound Med. **18**, 615–631 (1999).
[34] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression* (Wiley, New York, 1989).
[35] J. Cohen and P. Cohen, *Applied Linear Regression/Analysis for the Behavioral Sciences* (Erlbaum, Hillsdale, NJ, 1983).
[36] M. L. Crosfill and J. G. Widdicombe, ''Physical characteristics of the chest and lungs and the work of breathing in different mammalian species,'' J. Physiol. (London) **158**, 1–14 (1961).
[37] S. M. Tenney and J. E. Remmers, ''Comparative quantitative morphology of the mammalian lung: Diffusing area,'' Nature (London) **197**, 54–56 (1963).
[38] E. R. Weibel, ''Dimensions of the tracheobronchial tree and alveoli,'' in *Biological Handbooks: Respiration and Circulation*, edited by P. L. Altman and D. S. Dittmer (Federation of American Societies for Experimental Biology, Bethedsa, MD, 1971), pp. 930–939.
[39] H. G. Flynn and C. C. Church, ''Transient pulsations of small gas bubbles in water,'' J. Acoust. Soc. Am. **84**, 985–998 (1988).
[40] M. R. Bailey, D. Dalecki, S. Z. Child, C. H. Raeman, D. P. Penney, D. T. Blackstock, and E. L. Carstensen, ''Bioeffects of positive and negative acoustic pressures *in vivo*,'' J. Acoust. Soc. Am. **100**, 3941–3946 (1996).
[41] E. L. Carstensen, D. Dalecki, S. M. Gracewski, and T. Christopher, ''Nonlinear propagation of the output indices,'' J. Ultrasound Med. **18**, 69–80 (1999).
[42] C. H. Raeman, S. Z. Child, D. Dalecki, C. Cox, and E. L. Carstensen, ''Exposure-time dependence of the threshold for ultrasonically induced murine lung hemorrhage,'' Ultrasound Med. Biol. **22**, 139–141 (1996).
[43] T. J. Bauld and H. P. Schwan, ''Attenuation and reflection of ultrasound in canine lung tissue,'' J. Acoust. Soc. Am. **56**, 1630–1637 (1974).
[44] F. Dunn, ''Attenuation and speed of ultrasound in lung,'' J. Acoust. Soc. Am. **56**, 1638–1639 (1974).

Acoustic flow perception in cf-bats: Extraction of parameters

Rolf Müller[a] and Hans-Ulrich Schnitzler
Animal Physiology, Zoological Institute, University of Tübingen, Morgenstelle 28, D-72076 Tübingen, Germany

(Received 23 April 1999; accepted for publication 5 June 2000)

The narrow-band portions of the echolocation pulses seen in cf-bats are a hypothetical substrate for target localization. This localization could be based on estimates of echo envelope amplitude and carrier frequency together with their derivatives. Evaluation of these parameters is referred to as ''acoustic flow'' in loose analogy to optic flow. It is assessed whether the requirements for this task may be reconciled with known principles of auditory function. For the evaluation of a single echo, this seems to be the case: auditory filter shapes provide sufficient frequency resolution; at the same time envelopes are preserved well and some noise removal is achieved. Nevertheless, should bats not be endowed with additional capabilities for noise removal, analysis of acoustic flow would be limited to favorable signal-to-noise ratios. Multiple, temporally overlapping echoes are probable in any realistic echolocation scenario. In this case, additional auditory processing steps have to be postulated, which allow simultaneous estimation of multiple carrier frequencies and reduction of demodulation distortions. © *2000 Acoustical Society of America.* [S0001-4966(00)01709-4]

PACS numbers: 43.80.Ka [WA]

INTRODUCTION

Although bats are highly mobile animals, the possible evaluation of genuinely egomotion-related echo properties (summarily referred to as ''acoustic flow'' in loose analogy to optic flow) for obtaining spatial information has received little attention so far (exceptions are Lee, 1990; Lee *et al.*, 1992). In particular, the group of so-called ''cf-bats'' has not been considered at all. The echolocation pulse design in these animals is dominated by prolonged ''cf-portions'' with highly constant carrier frequencies and shallow-sloping envelopes. Frequency-modulated signal portions, which are most readily evaluated for target localization, are also present, but contain little energy at times (Tian and Schnitzler, 1997; Müller and Schnitzler, 1999). Under such circumstances, target localization could be based on the cf-portions by monitoring the changes in Doppler shift and sound-pressure amplitude of the echoes along a trajectory (Müller and Schnitzler, 1999).

Any echo feature underlying the acquisition of spatial knowledge has to fulfill two conditions: Its information content has to meet the requirements of the tasks bats are known to perform and feature extraction must be feasible in a biological signal-processing framework. Together, both conditions suffice to demonstrate feasibility, the topic of this work, but not to prove flow-field usage occurring in nature.

According to the companion paper to this report (Müller and Schnitzler, 1999), the first condition (sufficient information content) is met at least for idealized scenes (see Sec. I for the involved assumptions). The present work addresses the second condition, i.e., whether the hypothesis of acoustic flow-field perception in cf-bats can be reconciled with the known properties of auditory function in mammals in general and cf-bats in particular. Furthermore, it is attempted to identify critical elements of the auditory signal-processing cascade, the layout of which would be pivotal to an eventual flow perception. Hypotheses regarding such elements could be employed in devising experiments designed to assess whether cf-bats are endowed with the perceptual competences required for perceiving acoustic flow in the way it is defined here. Such experiments are a further intermediate step towards eventually establishing flow-field usage in cf-bats.

For the protracted narrow-band pulse portions seen in cf-bats, the echoes bounced off the various reflectors in a richly structured scene are returning simultaneously for all practical purposes. Therefore, it has to be postulated as a minimum requirement for acoustic flow usage that one signal of interest can be evaluated in the presence of distractor echoes. The notion of configural knowledge, defined as obtaining the configuration of several targets from the returns of a single pulse, would constitute an even tougher requirement, because more than one signal of interest has to be evaluated at a time. No compelling experimental evidence demonstrating this capability in bats is available. Certainly, configural knowledge would constitute a powerful basis for many spatial skills (e.g., path planning).

Extracting flow-field information from echoes originating at a single target constitutes a first step towards more realistic scenarios which involve multiple, simultaneous echoes. Therefore, the plan of this work is first to examine the general building blocks of auditory signal analysis in the context of acoustic flow perception (Sec. II), then evaluate the performance of a model for single echoes (Sec. III), and finally ask how this framework might be extended to accommodate the requirements of multiple-target situations (Sec. IV). The present work resorts to *ad hoc* estimation techniques only. This is sufficient for a first characterization of the estimation problem and arriving at a lower bound on estimation error (not the smallest lower bound). If the estimation problem can be reconciled with auditory function and

[a]Electronic mail: rolf.mueller@uni-tuebingen.de

a lower bound on the estimation error is commensurate with the requirements of the task, the inner validity of the hypothesis is strengthened.

I. ACOUSTIC FLOW—A BRIEF SUMMARY OF THE VARIABLES

A detailed description and assessment of the cues under consideration is given in Müller and Schnitzler (1999); this section provides only a brief summary: The echolocation scene is assumed to consist of point scatterers in a plane with position (d,h), where h is the heading distance always aligned with the animal's velocity and d is the passing distance orthogonal to h (see Fig. 2 in Müller and Schnitzler, 1999). The Doppler effect is described according to a simplified narrow-band model as a frequency shift f_d (Skolnik, 1980)

$$f_d = 2 f_e \frac{v}{c} \cos\varphi, \tag{1}$$

where f_e is the emitted frequency, v the bat's velocity, and c the speed of sound. The bearing angle φ is related to d and h by

$$\cos\varphi = h/\sqrt{d^2+h^2}. \tag{2}$$

The first acoustic flow observable is the proportional change in Doppler shift along the heading, denoted by $\Psi(d,h)$. From Eqs. (1) and (2), it follows that

$$\Psi(d,h) = \frac{\partial f_d}{\partial h} \Big/ f_d = \frac{d^2}{h(d^2+h^2)} \ [\mathrm{m}^{-1}]. \tag{3}$$

The second observable is the proportional change in sound-pressure amplitude P along the heading, denoted by $\Theta(d,h)$. Its evaluation is based on a simplified propagation model which combines a spherical spreading assumption (Kinsler and Frey, 1962) with a scalar-valued directivity function (e.g., Freedman, 1962). The directivity function $\Phi(d,h)$ combines the effects of emitter and receiver. According to this model, the received sound-pressure amplitude $P(d,h,f)$ is

$$P(d,h,f) = \frac{\sqrt{2\rho_0 c I_0}\sigma}{4\pi(d^2+h^2)} m(f,d,h)\Phi(d,h), \tag{4}$$

where f is the frequency, ρ_0 the density of the medium, c the respective speed of sound, I_0 the intensity at some reference distance, σ the scattering coefficient, and $m(f,d,h) = 10^{-\sqrt{d^2+h^2}\alpha(f)/10}$ the absorption term containing the absorption coefficient $\alpha(f)$ [dB/m]. It follows from Eq. (4) that

$$\begin{aligned}\Theta(d,h) &= \frac{\partial P}{\partial h} \Big/ P \\ &= \frac{\partial \ln(P)}{\partial h} \\ &= -\frac{2h}{d^2+h^2} - \frac{\alpha(f)/10\ln(10)h}{\sqrt{d^2+h^2}} + \frac{\partial\Phi}{\partial h} \Big/ \Phi \ [\mathrm{m}^{-1}].\end{aligned} \tag{5}$$

Considering proportional changes in Eq. (3), Eq. (5) cancels all nontrivial constants (e.g., σ), which are hardly known to the animal, and allows Ψ and Θ to be interpreted readily as a curve in the d,h plane. Depending on the used directivity function Φ, Eqs. (3) and (5) may be solved for d and h either analytically (e.g., in case of a spherical directivity) or numerically. The latter was done in the case of the ''sum-of-Gaussians'' directivity employed here. This directivity function is a descriptive model for experimental data (Müller and Schnitzler, 1999) and given by

$$\Phi(d,h) = A e^{[-\arctan^2(d/h)]/2s_1^2} + (1-A) e^{[-\arctan^2(d/h)]/2s_2^2}, \tag{6}$$

where $A<1$ is the amplitude of the first Gaussian (the amplitudes of both Gaussians sum to unity gain at zero bearing) and s_1, $s_2 > s_1$ are the standard deviations of the first and second Gaussian, respectively.

II. BIOLOGICALLY MOTIVATED SIGNAL-PROCESSING FRAMEWORK

Given the hypothetical status of acoustic flow perception, elaborating on the eventual neural mechanisms in any detail would be premature. Therefore, the design of the signal-processing model was guided by parsimony and compatibility with the established principles of auditory function in cf-bats or mammals in general. The model was employed in experiments evaluating the localization accuracy in single-target situations, particularly with respect to the influence of noise (Sec. III). Ways to extend this framework in order to cope with more than one signal at a time are discussed in Sec. IV.

A. Generating synthetic echoes

The performance of all signal-processing steps was evaluated by supplying synthetic input signals. The pulses had a constant carrier frequency adjusted for echoes from straight ahead to fall into the center of the auditory fovea, where maximum filter qualities occur (see Sec. II C 1). The flanks of the pulse envelope $\tilde{x}(t)$ were described by a logistic function

$$\tilde{x}(t) = 1 - \frac{1}{1+(t/c)^b}. \tag{7}$$

The steepness parameter b and the inflection point c were chosen arbitrarily as $b=8$ and $c=$pulse duration/8. Echoes were generated under the assumptions given in Sec. I; the scattering coefficient of the targets was set to unity for convenience, since it is being canceled in the expression for Θ. Doppler effects on echo carrier and envelope were modeled taking into account the cumulative phase history along the sound path. White Gaussian noise was added to the echoes by setting a fixed in-band pulse-to-noise ratio (PNR) defined by

$$\mathrm{PNR} = 10\log_{10}\frac{\int_{f_{\min}}^{f_{\max}} S_{\mathrm{pulse}}(f)\,df}{\int_{f_{\min}}^{f_{\max}} S_{\mathrm{noise}}(f)\,df}, \tag{8}$$

where $S_{\mathrm{pulse}}(f)$, $S_{\mathrm{noise}}(f)$ are the power spectra of pulse and noise, respectively. The integration limits $f_{\min}=82$ kHz and

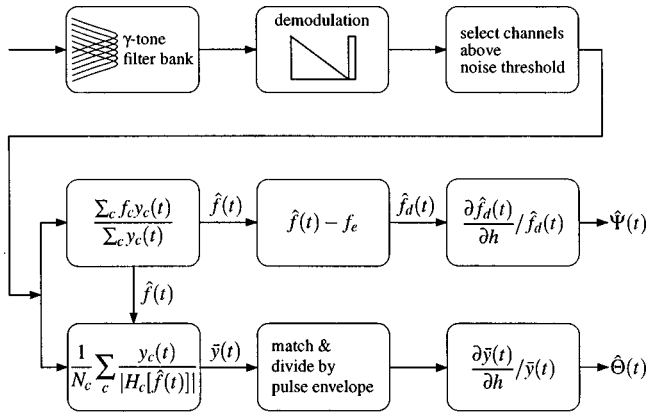


FIG. 1. Flowchart depicting the extraction procedure used for obtaining estimates $\hat{\Psi}$ and $\hat{\Theta}$ from a synthetic echo originating at a single target. Equations are explained in the text [see Eqs. (9), (10), (3), (5)].

$f_{\max}=84$ kHz are, respectively, the minimum and maximum center frequency in the filter bank used to assess model performance (Sec. III). The in-band signal-to-noise ratio for the echoes is calculated in the same manner by substituting $S_{\text{echo}}(f)$ for $S_{\text{pulse}}(f)$ in Eq. (8). For simplicity, adjustments to source amplitude as a function of target range (Kick and Simmons, 1984; Hartley, 1992a, 1992b; Tian and Schnitzler, 1997) are not considered.

B. Overview of system design

The linear input stage is a gammatone filter bank (Sec. II C). It is followed by a demodulator which discards the phase of the carrier (the envelope is the magnitude of the analytic signal generated by a Hilbert transform). After demodulation, channels are selected by comparing their output to a threshold. This threshold was set by the output level expected for each channel if it were driven by noise only. From the resulting representation, two parameters of interest have to be extracted: the carrier frequency and amplitude of the input signal. Considering only the output of a single, envelope-detected channel, this constitutes an ill-posed problem, because both amplitude and carrier frequency fluctuations of the input will manifest themselves in the output amplitude. However, the filter-bank nature of the auditory basis representation may be used to overcome this equivocation: In case of a single spectral component, the normalized first moment of the excitation pattern $y_c(t)$ (the index c denotes the channel), taken with respect to center frequency f_c at a given instant in time t, may be resorted to as an estimate $\hat{f}(t)$ of the carrier frequency of the input signal (Fig. 1)

$$\hat{f}(t)=\frac{\Sigma_c f_c y_c(t)}{\Sigma_c y_c(t)}. \tag{9}$$

This relationship holds exactly for the Wigner distribution, but not for the spectrogram, which constitutes a smoothed Wigner distribution (Hlawatsch and Boudreaux-Bartels, 1992). The filter-bank representation employed here (Sec. II C) is similar to a spectrogram (Ackroyd, 1971), but for a peculiar dependence of filter bandwidth on center frequency and omission of the envelope squaring usually applied to spectrograms (Hlawatsch and Boudreaux-Bartels, 1992). Estimates of carrier frequency obtained in this way are therefore biased. Nevertheless, they suffice for flow estimation (error <10 Hz within the fovea, Sec. II C). For multiple signals, a set of basis functions can be designed which takes into account these biases (Sec. IV D). Assuming knowledge of the channel transfer function magnitudes $|H_c(f)|$, $\hat{f}(t)$ can be used to correct for the influence of $|H_c(f)|$ on each $y_c(t)$ by dividing by $|H_c[\hat{f}(t)]|$ as long as $|H_c[\hat{f}(t)]|\gg 0$ holds. The latter condition requires the signal spectrum to be located somewhere well inside the filter's passband; this is ensured by the channel selection according to output amplitude (see above). After correcting for $|H_c(f)|$, envelope amplitude is estimated by averaging across channels in order to reduce the variance (Fig. 1)

$$\bar{y}(t)=\frac{1}{N_c}\sum_c \frac{y_c(t)}{|H_c[\hat{f}(t)]|}. \tag{10}$$

The estimated envelope amplitudes $\bar{y}(t)$ constitute a product of the information-bearing channel gain and the pulse envelopes, which were modified (compressed or dilated) by the Doppler effect. In the absence of any *a priori* knowledge regarding the sound-channel properties, the pulse envelope constitutes the best *ad hoc* hypothesis the animal may form about the echo envelopes generated. Since no prior knowledge about echo delay is available either, the echo positions are found by cross correlating the demodulated signal with the envelope of the pulse serving as a template. Having found the position of the best match, the echo envelope is divided by the template in order to separate the influence of the pulse flanks from that of the channel gain. The usability of this approach is confined to regions where the pulse envelope is of reasonable amplitude; otherwise, the noise is amplified by division with a small number. Solutions obtained for the envelope flanks result in unrealistic target speeds, which provide a convenient pruning criterion used here (successive position estimates implying ground speeds more than five times the actual speed of the bat were excluded). In the last step the derivatives of the envelope can be formed in the same way as for the carrier frequency. Quadratic Savitzky–Golay filters have been implemented for smoothing and computing the first derivative of carrier frequency and envelope.

C. *Rhinolophus*-like gammatone filter bank

Gammatone filters (see, e.g., Appendix 1 in Aertsen and Johannesma (1980) for a compilation of properties) have been applied widely in auditory modeling studies. As pointed out by Lyon (1996), advantages of this filter are its easy parametrization in terms of center frequency and bandwidth. Here, the gammatone model will be resorted to as a convenient way of creating an array of bandpass filters reproducing the pronounced dependency of bandwidth on center frequency, which has been found in cf-bats like the greater horseshoe bat, *Rhinolophus ferrumequinum* (Suga, 1976). At the same time, gammatone filter properties will subserve tentative guesses at relevant features of the auditory basis representation which have not yet been determined experimentally.

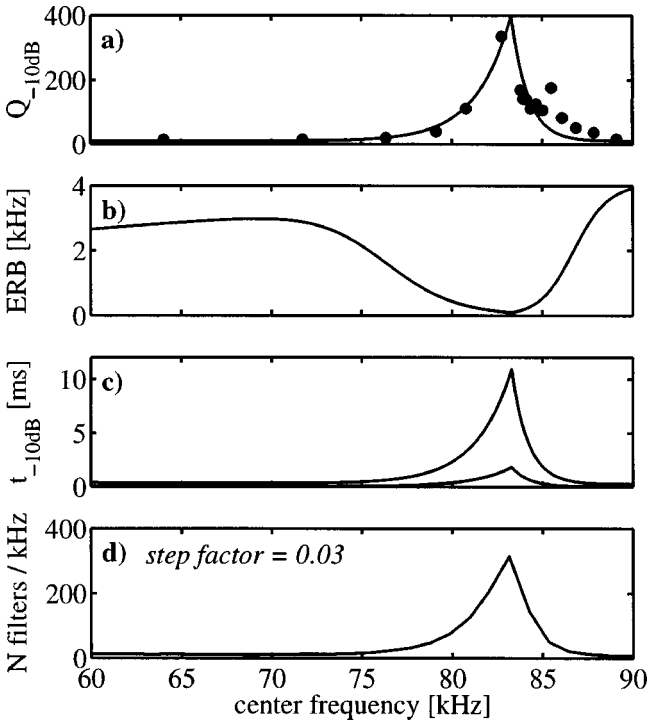


FIG. 2. Properties of the gammatone filter bank model for *Rhinolophus*, with fovea centered around $f_r = 83.3$ kHz. (a) model of filter quality Q as a function of center frequency and data of Suga (1976) (dots shown are the maxima of the $Q_{-10\,\mathrm{dB}}$-values in logarithmically spaced bins); (b) equivalent rectangular bandwidth (ERB), minimum ERB≈92.4 Hz (−10-dB bandwidth ≈166.5 Hz); (c) −10-dB cutoff times for the impulse response envelopes; (d) filter density.

1. Filter placement and auditory fovea

The most striking features of the auditory system in cf-bats are the specializations pertinent to the processing of the narrow-band echo portions. The two best-established aspects of these specializations are over-representation and comparatively small analysis bandwidth (see Schnitzler and Ostwald (1983) for a summary). Both are included in the filter-bank model presented here.

The data of Suga (1976) quantify filter quality Q: The highest values reported ($Q_r \approx 400$ at −10 dB) are assigned to the center of the fovea ($f_r \approx 83.3$ kHz). Above and below f_r, $Q_{-10\,\mathrm{dB}}$ decreases with different slopes to a minimum value of $Q_0 \approx 10$ reached at the borders of the frequency range studied ($f_0 = 70$ kHz and $f_0 = 90$ kHz, respectively). This behavior is modeled here by virtue of the expression

$$Q_{-10\,\mathrm{dB}}(f) = \frac{Q_r - Q_0}{m-1}\left(m^{(f_0-f)/(f_0-f_r)} - 1\right) + Q_0. \qquad (11)$$

A blemish in this description lies in the large magnitude of the parameter m ($m = 3.8505 \times 10^{18}$ was used for $f \leqslant f_r$ and $m = 3.0512 \times 10^3$ for $f > f_r$), but it was found useful in reproducing the salient features of Suga's (1976) data, nevertheless [Fig. 2(a)].

Combining a function describing the relationship between center frequency and filter quality with a rough guess at the number of filters available [3000 filters thought to represent inner hair cells were assumed arbitrarily after inspection of the data presented in Vater (1988)] does not define placement of filters unequivocally. Thus, the assumption was incorporated that filters are spaced for equal overlap in energy between neighboring passbands (Slaney, 1993). The function 1/ERB(f) (ERB is the equivalent rectangular bandwidth of the filters) was integrated numerically to find overlap and center frequencies. The data of Suga (1976) specify the −10-dB bandwidth, which under the assumption of a fourth-order gammatone filter is ~2 the −3-dB bandwidth. The −3-dB bandwidth is turn is 0.887 ERB (Patterson *et al.*, 1992). The resulting filter density function [Fig. 2(d)] is in reasonable qualitative agreement with anatomical results obtained by Bruns and Schmieszek (1980) for the spiral ganglion in *Rhinolophus* [Bruns and Schmieszek (1980) report 21% of the neurons to fall within 80 and 86 kHz; the model places 32% of its filters in this interval].

2. Relevant gammatone filter properties

Improving the resolution of spectral peaks constrained by little *a priori* knowledge about their number and location (Sec. IV A) calls for narrowing of the filter passbands. This may conflict with conservation of the envelopes, which requires constant gain and linear phase over the frequency band, where the signal contains non-negligible energy.

The fourth-order gammatone filters employed (IIR-implementation according to Slaney, 1993) meet the requirements of envelope preservation imperfectly (Fig. 3): The magnitude transfer functions are not flat [Fig. 3(a)] and phase responses deviate more from linear as equivalent rectangular bandwidth ERB decreases [Figs. 3(b), 3(c)]. Together, these effects cause a strong overall dependence of faithfulness in transmission of the signal envelope on ERB [Fig. 3(d)]. For the ERB values in the model filter bank (Sec. II C 1) the errors are readily reconcilable with the requirements of envelope analysis (relative error in envelope 4×10^{-3} for min{ERB}≈92.4 Hz and 6.47×10^{-7} for max{ERB}≈5.4 kHz). Further reduction of bandwidth, however, would eventually lead to intolerable degradation of the signal's time structure. In the presence of noise or distractor echoes, the fact that biological filter bandwidth in *Rhinolophus* is apparently not matched to the spectral shape of the envelopes [Fig. 3(a)] adversely affects flow evaluation (see Secs. III, IV B).

III. MODEL PERFORMANCE IN SINGLE-TARGET SITUATIONS

Any signal analysis based on derivatives is potentially susceptible to noise. Although bats can be expected to encounter favorable signal-to-noise ratios in the transmission channel and the band of operation used by their sonar systems, neural processing as well as the stochastic impulse responses of extended scatterers will most likely limit the overall signal-to-noise ratio. In the absence of any data, which would allow for a guess at the extent of these influences, demonstrating a certain robustness against noise of the proposed extraction procedure is desirable in order to support usage of acoustic flow as a viable hypothesis for biological function.

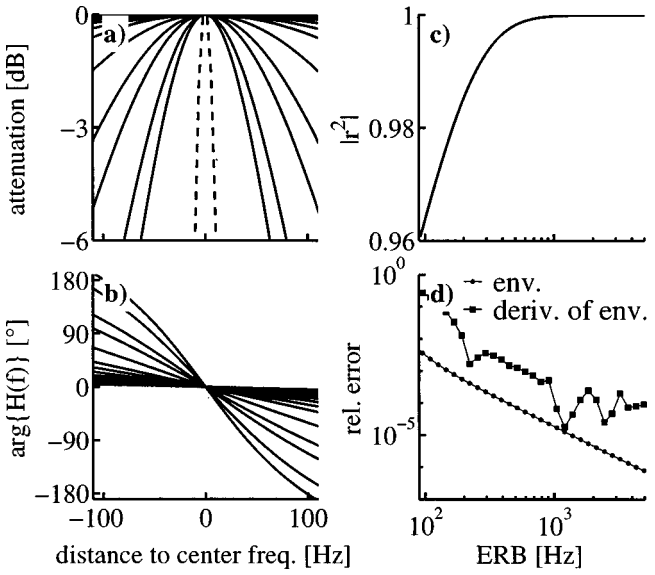


FIG. 3. Suitability of gammatone filters (fourth order IIR implementation according to Slaney, 1993) for conservation of signal envelope and its first derivative: (a) magnitude transfer functions of filters from the model filter bank (see Fig. 2). A sample including every 200th channel is shown. The dashed line shows the power spectral density of a model pulse envelope [logistic slopes, see Eq. (7), duration 80 ms] for comparison; (b) same as (a), but phase responses are shown; (c) the linearity of the phase in an interval of center frequency ±500 Hz as a function of equivalent rectangular bandwidth (ERB). The correlation coefficient of a linear regression of the phase response is plotted on the ordinate; (d) mean relative deviation (error) between envelope and its first derivative when comparing input and output signal. All filters used had a fixed center frequency of 83.3 kHz.

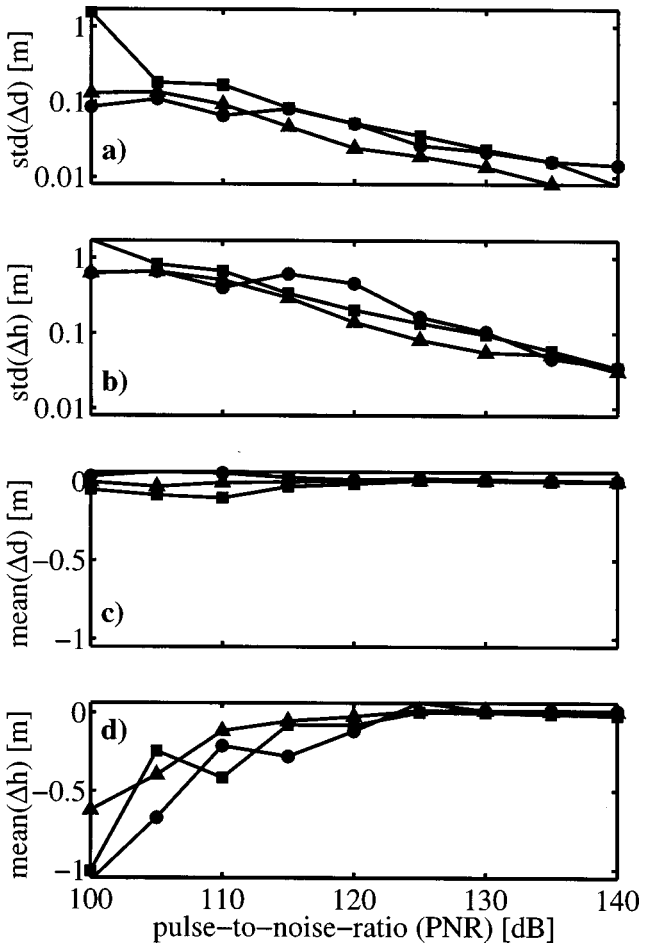


FIG. 5. Localization errors in noise: standard deviations (a), (b) and biases (c), (d) for passing distance *d* (a), (c) and heading distance *h* (b), (d). The symbols denote passing distance as in Fig. 4. For each passing distance, localization was performed with ten different realizations of the noise.

The total noise reduction provided by the initial bandpass filtering and the final smoothing operating on the derivatives (Fig. 1) was assessed by attempting target position estimates at different noise levels. The noise levels are specified in terms of an in-band signal-to-noise ratio for the pulses given by Eq. (8). This provides a constant reference (the pulse amplitude) despite echo amplitudes decaying with distance and bearing (Fig. 4).

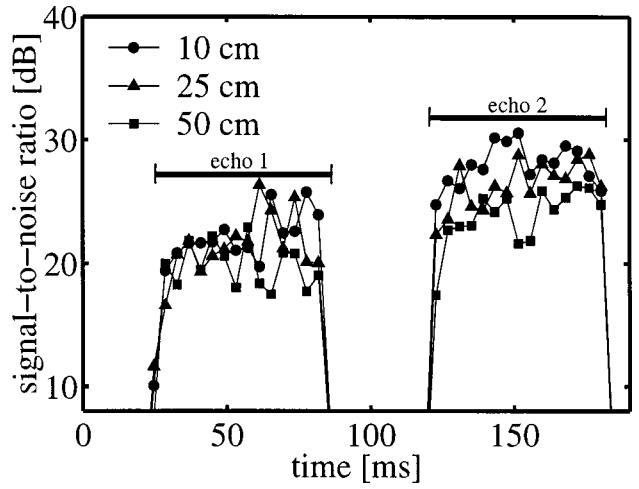


FIG. 4. In-band echo-to-noise ratio for the two echoes (marked ''echo 1'' and ''echo 2''), which were analyzed in Fig. 5. The legend refers to the different passing distances *d*. PNR is 100 dB; the corresponding curves for the other PNR values in Fig. 5 would be shifted by the respective PNR-level differences. Signal segments were selected by a Hanning window of ~4 ms duration. For each segment, the in-band echo-to-noise ratio was calculated according to Eq. (8) with $S_{\mathrm{echo}}(f)$ substituted for $S_{\mathrm{pulse}}(f)$.

Only a subset of all auditory filters in the model filter bank for *Rhinolophus* (Fig. 2) was employed; it comprised 100 filters arranged between 82 and 84 kHz. The emitted frequency was set to 80.92 kHz, so that echoes from straight ahead fell into the fovea ($v = 5\ \mathrm{ms}^{-1}$, the fovea being located at 83.3 kHz). All generated echoes were covered by the bank's center frequencies.

Unfavorable signal-to-noise ratios result in greater variability of the estimates as well as in biases (Fig. 5).

Both types of errors are generally larger for the heading (*h*) than for the passing distance (*d*) estimates. This is particularly true for biases at the lower signal-to-noise ratios. The biases are due to the high-frequency components of the noise which result in large derivatives. The latter are indicative of proximate targets; therefore, biases are mostly negative (underestimation of target distance). Although position estimation involves several nonlinearities (demodulation, determining *d* and *h* from Ψ and Θ, pruning of the results), the relationship between the variances of the noise and the estimate is approximately linear [regression coefficient $|r| > 0.9$ for all *d* but 50 cm, where $r = -0.643$ for *d* and *r*

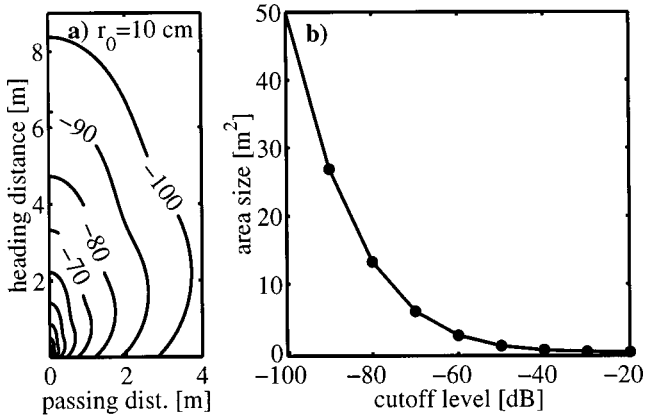


FIG. 6. Modeled sensitivity for the echolocation system of *Rhinolophus ferrumequinum* as a function of position in a ''sensing plane.'' (a) contour lines map cutoff levels for attenuation [dB] in the right, forward hemifield. The reference for the values given is the sound-pressure level at a distance $r_0=10$ cm. (b) size of area (full field) with attenuation≥cutoff level given on the abscissa.

$=-0.863$ for h, Figs. 5(a), 5(b)], as would be expected for a linear estimator.

Performance at the lowest PNR values studied is commensurate only with blurry localization and bats would either be confined to operating in the more favorable parts of the PNR range given in Fig. 5 or require noise removal capabilities beyond those of the model.

IV. COPING WITH MULTIPLE TARGETS

The proposed usage of acoustic flow needs to be evaluated in the presence of multiple targets, because such situations are possible in natural obstacle avoidance scenarios. This section discusses the probability of multiple echoes (Sec. IV A 1), performance requirements (Sec. IV A 2), the composition of the resulting auditory representations (Sec. IV B), and their evaluation (Secs. IV C, IV E).

A. System requirements

1. Distributions for echo amplitude and center frequency

Sequential echo processing in time will be rarely feasibly in cf-bats, because of the long pulse durations (e.g., routinely ≥50 ms in *Rhinolophus*; this corresponds to 8.5 m in target range). Useful echo separation can hence occur either in amplitude or in center frequency. The combination of directivity, geometrical attenuation, and absorption (Sec. I) severely limits the sonar's field of view (Fig. 6).

If this field of view contains distinct, solitary reflectors (like smooth tree trunks or branches), the number of simultaneous echoes may be limited to a few. Echo separation in the frequency domain has to cope with the restriction that all echoes share the frequency band $-2v/cf_e \leqslant f_d \leqslant 2v/cf_e$. Estimating the distribution of carrier frequencies requires knowledge of the reflector distribution with respect to bearing $p_\varphi(\varphi)$, which is presently not available. However, for any $p_\varphi(\varphi)$ $(0\leqslant\varphi\leqslant\pi)$, the functional form of the mapping $\varphi\mapsto f_d$ [Eq. (1)] increases the density of carrier frequencies towards the extreme values $f_d=\pm 2v/cf_e$ in comparison to the density of target bearings close to its limits $(0,\pi)$. This is evident from considering the denominator in the expression for the distribution of f_d

$$p_{f_d}(f_d)=\frac{p_\varphi[\arccos(c/(2vf_e)f_d)]}{\sqrt{4\frac{v^2}{c^2}f_e^2-f_d^2}}, \tag{12}$$

which can be derived from Eq. (1) by virtue of the relationship $p_{f_d}(f_d)=p_\varphi[g^{-1}(f_d)]|dg^{-1}(f_d)/df_d|$ (Papoulis, 1991), where $f_d=g(\varphi)$ and $g^{-1}(f_d)$ is the inverse of $g(\varphi)$. $p_{f_d}(f_d)$ may further constrain the required frequency resolution unless $p_\varphi(\varphi)$ is biased against small target bearings, which may be the case, if the bat always tries to steer into openings.

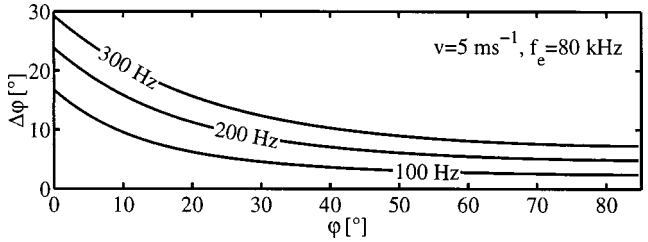


FIG. 7. Angular resolution $\Delta\varphi$ as a function of bearing angle φ; the parameter is frequency resolution Δf.

2. Frequency resolution

A minimum requirement of acoustic flow analysis is the ability to monitor frequency location and amplitude of one spectral peak over time in the presence of other signals. An impression of the constraints imposed on frequency resolution by the spatial task can by gained from the dependence of bearing resolution $\Delta\varphi$ on frequency resolution Δf. From Eq. (1), Δf is given by

$$\Delta f=2f_e\frac{v}{c}[\cos\varphi-\cos(\varphi+\Delta\varphi)]. \tag{13}$$

Solving Eq. (13) for $\Delta\varphi$,

$$\Delta\varphi(\Delta f,\varphi)=\arccos\left(\cos\varphi-\frac{\Delta f}{f_e}\frac{c}{2v}\right)-\varphi;$$
$$\cos\varphi-1\leqslant\frac{\Delta f}{f_e}\frac{c}{2v}\leqslant\cos\varphi+1, \tag{14}$$

is obtained (Fig. 7).

$\Delta\varphi(\Delta f,\varphi)$ has its maximum at $\varphi=0$ for any given Δf. Since the system's point of operation for obstacle avoidance has to be placed close to this direction, small values of Δf (e.g., ~100 Hz) are highly desirable.

B. Interference problem

A filter bank representation shares the spectrogram's interference structure. If $a_n\cos(\omega_n t+\varphi_n)$ is the output of a linear filter to the nth input signal presented in isolation, the envelope-detected output of this filter to the entire ensemble $|\underset{\triangledown}{y}(t)|$ is

$$\nabla |y(t)| = \sqrt{\left(\sum_n a_n \cos(\omega_n t+\varphi_n)\right)^2+\left(\sum_n a_n \sin(\omega_n t+\varphi_n)\right)^2},$$

with $\Delta\omega_{mn}=\omega_m-\omega_n$ and $\Delta\varphi_{mn}=\varphi_m-\varphi_n$

$$= \sqrt{\sum_n a_n^2+2\sum_{n\neq m} a_n a_m \cos(\Delta\omega_{mn}t+\Delta\varphi_{mn})}. \qquad (15)$$

The second sum under the square root corresponds to the cross terms of the spectrogram, N interfering signals give rise to $N(N-1)/2$ cross terms; the cross-term carriers are oscillating with the difference frequency of the two generating signal terms. The oscillatory nature of the cross terms is detrimental to recovery of both Ψ and Θ because the involved derivatives are dominated by the steep slopes of the oscillations. It can be seen readily from Eq. (15) that cross terms are confined to the overlapping region of the autorepresentations $(a_n\neq 0 \wedge a_m\neq 0)$. Spectral separation of the echoes is not sufficient for preventing the formation of cross terms. If an auditory filter's passband is too wide to suppress all but one echo, cross terms will be found in its output.

The high filter qualities found in cf-bats (e.g., up to $Q_{-10\,\mathrm{dB}}\approx 400$ in *Rhinolophus* [Suga (1976)] reflect primarily large center frequencies; the −10-dB bandwidth hardly falls below 200 Hz (Fig. 2). Therefore, ''distractor'' echoes with center frequencies 100 Hz away from the echo of interest could only be suppressed by 10 dB. This relationship between center frequency spacing and attenuation does not render the raw output of the gammatone filter bank a practicable substrate for simultaneous flow analysis of several targets (see Fig. 8 for an example).

Assuming correctness of the current filter bank model, evaluation of acoustic flow in multiple-target situations requires an additional processing stage capable of suppressing the cross terms.

C. An opportunity window for a modulation low pass?

Demodulation in the auditory system is modeled reasonably well by half-wave rectification and subsequent low-pass filtering (Schroeder and Hall, 1974). Demodulation is a nonlinear operation, which introduces undesired modulation products, if the input comprises more than a single frequency. The modulation products depend on the input as well as on the demodulation procedure (Aiken, 1931, 1933; Bennett, 1933). The strongest undesired spectral component generated by full-wave rectifying two superposed amplitude-modulated signals is the ''carrier beat'' at the difference frequency of the two carriers (Aiken, 1931). For narrow-band cf-signals there is a chance that at least the heavy distortions do not coincide spectrally with the desired envelope. For example, the carrier beat is above the envelope spectrum as long as the difference in center frequencies exceeds the envelope's bandwidth. Distortions outside the envelope spectrum can be removed by an additional step of low-pass filtering. Because of the difficulties in analytic treatment of the distortion problem, the effect of a modulation low pass applied to a half-wave rectified output of a gammatone filter was studied experimentally: Two signals were fed into a gammatone bandpass (ERB∼173 Hz, center frequency ∼82 kHz): one signal of interest, its carrier frequency, aligned with the filter's center frequency and an attenuated or amplified distractor with displaced carrier frequency. Low-pass filtering proved to have considerable capacity for reduction of the errors in the estimates of the envelope and its derivative (Fig. 9).

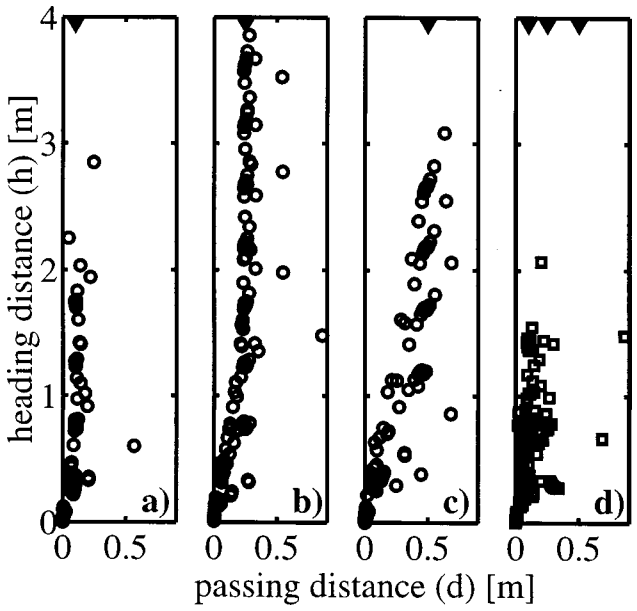


FIG. 8. Example of the inadequacy of flow-field analysis based on the unworked output of the gammatone filter bank for situations with several targets: (a), (b), (c) reconstructions of target positions obtained, when synthetic, noise-free echoes from three targets at (a) $d=10$ cm (initial $h=2$ m), (b) $d=25$ cm (initial $h=4$ m), (c) $d=50$ cm (initial $h=3$ m) are presented in isolation. (d) reconstruction results for a superposition of the echoes analyzed in (a), (b), (c). Passing distances are indicated by triangles at the top of all graphs. Flight speed was set to 5 m/s, eight echoes (pulse duration 80 ms, interpulse interval 20 ms) were received within the displayed interval of heading distances. Reconstructions of target position (circles or squares) were attempted every 1.6 ms.

An estimation error of a few percent is commensurate with the requirements of acoustic flow evaluation (Müller and Schnitzler, 1999). For equal amplitudes of the two signals, this can be achieved at a frequency separation of ∼150 Hz. If the sensitivity of the sonar system or the targets' scattering coefficients provide ∼20 dB attenuation of one signal, the stronger echo's envelope can be estimated with about 1 percent error at ∼80 Hz separation in frequency. Psychoacoustic results in humans are modeled well by a modulation filter bank (Dau *et al.*, 1997), which operates on the demodulated output of the initial auditory filter bank. Assuming cf-bats to share this feature and modulation filters to be low passes matched to the pulse envelopes, this could considerably enhance access to acoustic flow-field information for multiple signals.

D. Excitation patterns: Properties and decomposition

For the simplified auditory model considered (linear bandpass filtering plus demodulation) and under the assumption of negligible demodulation distortions, a compound ex-

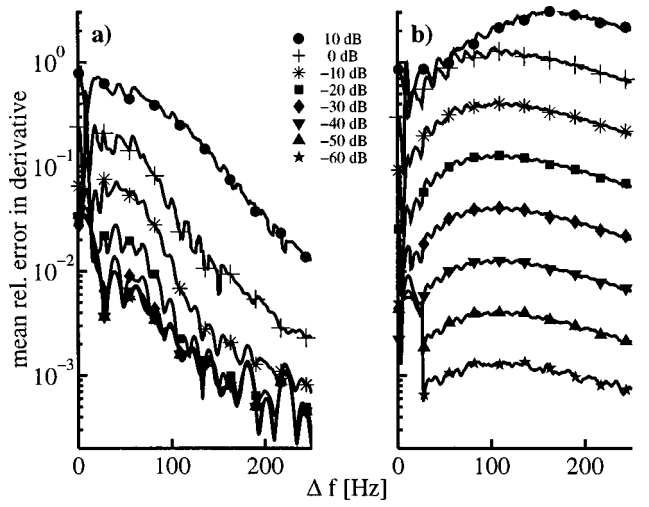


FIG. 9. Effect of low-pass filtering on the error in the derivative estimates for a signal envelope in the presence of an attenuated and carrier-shifted copy of the signal. (a) additional low-pass filtering: A 128-tap FIR filter was designed with a windowed linear-phase method (Hamming window) to match the frequency separation of components with its cutoff frequency. (b) no low-pass filtering. (a) and (b) abscissa: frequency separation between carriers (Δf); ordinate: relative error in derivative (i.e., deviation divided by true value).

citation pattern can be decomposed into its components by a set of basis functions. The basis functions set can be obtained from the filter bank response to single echoes of different carrier frequency. The auditory filters used here are adjusted for equal gain at the maximum of their transfer functions (no data on cf-bats are available). In such a filter bank, asymmetric peak shapes are caused by the change in ERB as a function of center frequency: Filters with center frequencies between echo carrier and foveal frequency have smaller ERBs than those on the opposite site of the carrier and therefore respond with smaller output amplitudes. This leads to excitation patterns having a positive skewness, if their center frequency is higher than the foveal frequency, and a negative skewness, if the reverse is true (Fig. 10).

Frequency location and weights of the basis functions making up a compound excitation pattern are continuous functions of time. An approximation is possible by finding the weights for a discrete set of basis functions, the centers of which are arranged along the frequency axis in an orderly manner. Number and positioning of these basis functions are limited by the sampling of the excitation pattern given by the center frequency locations of the bandpass channels. For a supposedly noisy excitation pattern, the number of basis functions has to be smaller than the number of samples, in order to avoid overfitting of the noise. Taking into account the information that all components of the excitation pattern are positive in sign, the problem can be posed in a general least-squares sense as minimizing

$$\chi^2 = |\mathbf{H_{Bf}} w - g|^2 \quad \text{subject to} \quad w \geqslant 0, \tag{16}$$

where $\mathbf{H_{Bf}}$ is a $m \times n$ design matrix of n basis functions evaluated at m positions, w a vector containing the n unknown weights to be determined, and g the excitation pattern at a certain instant in time comprising m filter outputs. The linear inequality constraint ($w \geqslant 0$) was taken into account by resorting to the non-negative least-squares algorithm stated by Lawson and Hanson (1995). The intermediate least-squares problems in this algorithm were solved using the pseudoinverse computed by singular value decomposition. An example for this approach is presented in Sec. IV E and shown in Fig. 11.

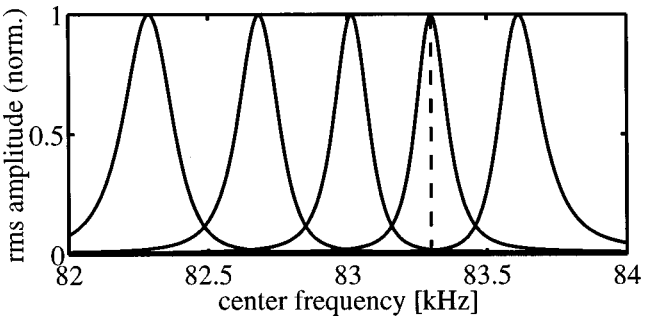


FIG. 10. Examples of excitation patterns generated by echoes with different carrier frequencies. Abscissa: center frequencies of the employed channels. Ordinate: normalized mean square amplitude of the channel output. The frequency location of the auditory fovea's center is marked by the dashed line.

E. Frequency tracking example

In order to hint at the principal feasibility of the approach combining low-pass smoothing of the excitation pattern (Sec. IV C) with a basis function decomposition of the result (Sec. IV D), an example for tracking three frequency components simultaneously is presented (Fig. 11). Frequency sweeps [Fig. 11(a)] were generated according to targets initially positioned at passing distances d of 8.7, 84.5, and 172 cm. The respective initial heading distances h were 1, 1.8, and 2.45 m. The targets were then approached at a speed of $v = 6\ \mathrm{ms}^{-1}$. As evident from Fig. 6, the spatial sensitivity of the bat's echolocation system will in many cases result in a pronounced weighting of the targets. An unbiased assessment of tracking performance for all three components is provided by looking at the tracking of echoes with all amplitudes kept equal. This has been done for the example in Fig. 11. The analysis was based on the output of 500 bandpass channels spaced between 82.9 and 84 kHz [Fig. 11(b)]; this corresponds to a raised sampling rate for the excitation function when compared to the filter bank shown in Fig. 2, which positions only 322 filters in this range, but may nevertheless still be inside the range of what is biologically plausible (Vater, 1988). Basis functions were placed on every third channel's center frequency; thus, their density mirrored the filter density and the degree of overdetermination of the problem stated in Eq. (16) was independent of position on the frequency axis. The errors in frequency estimates obtained by calculating the first moment of each ''ridge'' of weights with respect to frequency [Fig. 11(d)] were found to be ~10 to 20 Hz (onsets excluded), which amounts to ~0.35% to 1% of the Doppler shifts occurring in the example. This is readily compatible with the requirements of acoustic flow evaluation (Müller and Schnitzler, 1999). Optimization of the decomposition system's specifications would still be required in order to yield estimates for Ψ and Θ with sufficient accuracy for instance in the presence of

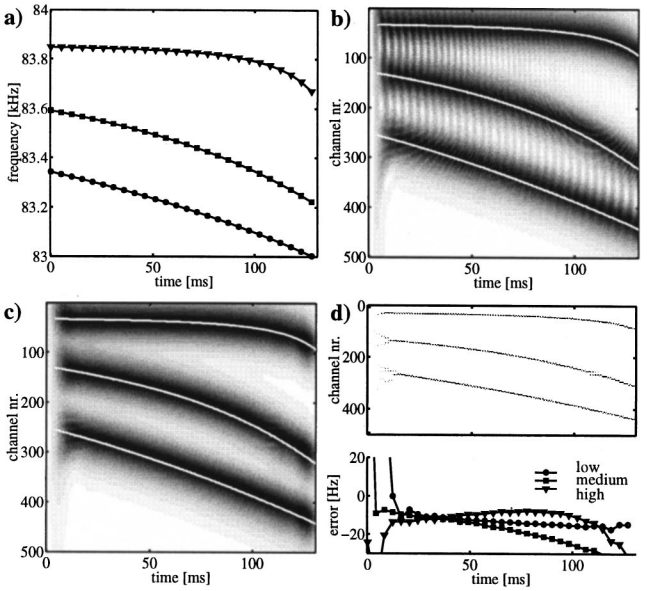


FIG. 11. Example of the reconstruction of multiple carrier frequencies: (a) location of the carrier frequencies for the three component sweeps; the respective amplitudes are equal and constant for all components; (b) output of the bandpass filter stage: although interference phenomena are most visible between the sweeps, they do extend into the maxima of the response pattern; (c) output after low-pass filtering stage, which reduced the interference phenomena. In (b) and (c) white lines mark the mapping of the echo carriers onto the filter bank's frequency axis. (d) upper graph: basis function weights serving as carrier frequency estimates; lower graph: error in estimating the three component carriers, which are referred to as ''low,'' ''medium,'' and ''high.'' In (b), (c), (d) the ordinate is linearly scaled in channel numbers and the mapping to the respective center frequencies is nonlinear; see Fig. 2 for the dependence of sampling the excitation function in the frequency domain on center frequency.

noise. However, the developmental effort for such an optimization of the system is outside the scope of this work, which is restricted at this point to proposing this approach, without fully exploring its limits.

V. CONCLUSIONS

The specifications of the hearing system of cf-bats as mirrored by the current auditory models are compatible with obtaining flow-field information from single targets. Noisiness of the input as well as of internal signal representations is a major concern, however. In the context of multiple targets, the filter bandwidths of the initial bandpass stage appear to be ill-dimensioned, because they are not matched to echo bandwidth, thereby fostering the generation of oscillatory cros terms. However, initial bandpass filtering acts to reduce demodulation distortions, opening up a window of opportunity for subsequent separation of the strongest envelope in each channel from distractors by low-pass filtering. Consequently, in a simplified acoustic environment a likewise simplified auditory system may well be able to gather some crude localization information from acoustic flow. Both major limitations (signal-to-noise ratio and interference) can be expected to effectively shut off meaningful flow-field information at times, and other information sources will be required if localization is critical during these time intervals. The real physical world and the genuine set of functional specifications for the auditory system will doubtlessly introduce numerous additional complications. Examples are extended, textured reflecting surfaces and realistic wave-field geometries for the acoustic scenario and level-dependent nonlinearities, as well as a realistic neural encoding for the auditory model. These properties might impose further restrictions on the available information and on the capability to extract it, but unexplored properties of the hearing system could also function as the building blocks of a better estimator. In any case, these aspects are worth exploring.

Ackroyd, M. H. (**1971**). ''Short-time spectra and time–frequency energy distributions,'' J. Acoust. Soc. Am. **50**, 1229–1231.

Aertsen, A. M. H. J., and Johannesma, P. I. M. (**1980**). ''Spectro-temporal receptive fields of auditory neurons in the grassfrog,'' Biol. Cybern. **38**, 223–234.

Aiken, C. B. (**1931**). ''The detection of two modulated waves which differ slightly in carrier frequency,'' Proc. IRE **19**, 120–137.

Aiken, C. B. (**1933**). ''Theory of the detection of two modulated waves by a linear rectifier,'' Proc. IRE **21**, 601–629.

Bennett, W. R. (**1933**). ''New results in the calculation of modulation products,'' Bell Syst. Tech. J. **12**, 228–243.

Bruns, V., and Schmieszek, E. (**1980**). ''Cochlear innervation in the greater horseshoe bat: Demonstration of an acoustic fovea,'' Hear. Res. **3**, 27–43.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). ''Modeling auditory processing of amplitude modulation. I. detection and masking with narrow-band carriers,'' J. Acoust. Soc. Am. **102**, 2892–2905.

Freedman, A. (**1962**). ''A mechanism of acoustic echo formation,'' Acustica **12**, 10–21.

Hartley, D. J. (**1992a**). ''Stabilization of perceived echo amplitudes in echolocating bats. I. Echo detection and automatic gain control in the big brown bat, *Eptesicus fuscus*, and the fishing bat, *Noctilio leporinus*,'' J. Acoust. Soc. Am. **91**, 1120–1132.

Hartley, D. J. (**1992b**). ''Stabilization of perceived echo amplitudes in echolocating bats. II. The acoustic behavior of the big brown bat, *Eptesicus fuscus*, when tracking moving prey,'' J. Acoust. Soc. Am. **91**, 1133–1149.

Hlawatsch, F., and Boudreaux-Bartels, G. F. (**1992**). ''Linear and quadratic time-frequency signal representations,'' IEEE Signal Process. Mag. **9**, 21–67.

Kick, S. A., and Simmons, J. A. (**1984**). ''Automatic gain control in the bat's sonar receiver and the neuroethology of echolocation,'' J. Neurosci. **4**, 2725–2737.

Kinsler, L. E., and Frey, A. R. (**1962**). *Fundamentals of Acoustics*, 2nd ed. (Wiley, New York).

Lawson, C. L., and Hanson, R. J. (**1995**). *Solving Least Squares Problems*, Classics in Applied Mathematics (SIAM, Philadelphia).

Lee, D. N. (**1990**). ''Getting around with light or sound,'' in *The Perception and Control of Self-motion*, edited by R. Warren and A. H. Wertheim (Erlbaum, Hillsdale, NJ).

Lee, D. N., van der Weel, F. R., Hitchcock, T., Matejowsky, E., and Pettigrew, J. D. (**1992**). ''Common principle of guidance by echolocation and vision,'' J. Comp. Physiol. A **171**, 563–571.

Lyon, R. F. (**1996**). ''The all-pole gammatone filter and auditory models,'' in Computational models of signal processing in the auditory system, Forum Acusticum 96, Antwerp, Belgium (unpublished).

Müller, R., and Schnitzler, H.-U. (**1999**). ''Acoustic flow perception in cf-bats: Properties of the available cues,'' J. Acoust. Soc. Am. **105**, 2958–2966.

Papoulis, A. (**1991**). *Probability, Random Variables, and Stochastic Processes*, 3rd ed. (McGraw-Hill, New York).

Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (**1992**). ''Complex sounds and auditory images,'' in *Auditory Physiology and Perception*, edited by Y. Cazals and L. Demany (Pergamon, Oxford), pp. 429–446.

Schnitzler, H.-U., and Ostwald, J. (**1983**). ''Adaptation for the detection of fluttering insects by echolocation in Horseshoe Bats, in *Advances in Vertebrate Neuroethology*, edited by J. P. Ewert, R. R. Capranica, and D. J. Ingle (Plenum, New York), pp. 801–827.

Schroeder, M. R., and Hall, J. L. (**1974**). ''Model for mechanical to neural

transduction in the auditory receptor,'' J. Acoust. Soc. Am. **55**, 1055–1060.
Skolnik, M. E. (**1980**). *Introduction to Radar Systems*, 2nd ed. (McGraw-Hill Kogakusha, Tokyo).
Slaney, M. (**1993**). ''An efficient implementation of the patterson-holdsworth auditory filter bank,'' Technical Report 35, Apple Computer Inc.
Suga, N. (**1976**). ''Peripheral auditory tuning for fine frequency analysis by the cf-fm bat, *Rhinolophus ferrumequinum*. IV. Properties of peripheral auditory neurons,'' J. Comp. Physiol. **106**, 111–125.
Tian, B., and Schnitzler, H.-U. (**1997**). ''Echolocation signals of the greater horseshoe bat (*Rhinolophus ferrumequinum*) in transfer flight and during landing,'' J. Acoust. Soc. Am. **101**, 2347–64.
Vater, M. (**1988**). ''Cochlear physiology and anatomy in bats,'' in *Animal Sonar Processes and Performance*, edited by P. E. Nachtigall and P. Moore (Plenum, New York), pp. 225–242.

A mathematical analysis of the peripheral auditory system mechanics in the goldfish (*Carassius auratus*)

James J. Finneran[a] and Mardi C. Hastings[b]
Department of Mechanical Engineering, The Ohio State University, 206 West 18th Avenue, Columbus, Ohio 43210

(Received 29 September 1998; accepted for publication 30 March 2000)

The dynamic response of the goldfish peripheral auditory system has been analyzed using lumped-parameter mechanical and fluid system models for the swimbladder, Weberian apparatus, and saccule. The swimbladder is treated as a two degree-of-freedom mechanical system consisting of two coupled mass-spring-damper arrangements. The swimbladder is coupled to the Weberian ossicles using a phenomenological analysis of the anterior swimbladder tunica externa which permits both stretching and sliding. Analysis of the saccule features only a single degree of freedom, corresponding to the direction of orientation of the ciliary bundles. Inputs to the saccule consist of the transverse canal fluid motion and the motion of the animal's head (assumed to match the local acoustic particle motion). Mechanical properties required for the system equations were estimated from published literature, direct measurements, and curve fits to experimental data for the motions of the swimbladders. The results indicate that the Weberian apparatus has a significant impact on hearing ability over the entire auditory bandwidth, not just at higher frequencies, and that the saccule functions as a displacement sensor above approximately 300 Hz. © *2000 Acoustical Society of America.* [S0001-4966(00)02207-4]

PACS numbers: 43.80.Lb [WA]

INTRODUCTION

The term ''hearing specialist'' is often used to describe teleost fish, which have some functional adaptation that enhances auditory sensitivity. Goldfish, as well as other members of the series Otophysi, are considered hearing specialists because they possess Weberian ossicles—a series of specialized vertebrae that mechanically couples the swimbladder to the endolymphatic fluid of the ear. Because of their low auditory thresholds [down to approximately 55 dB re: 1 μPa (Jacobs and Tavolga, 1967)] and physical hardiness, goldfish have been popular test subjects and a large amount of behavioral and physiological data exists for them.

The objective of this study was to develop a mathematical model for the mechanics of the goldfish peripheral auditory system. Previous theoretical analyses of the peripheral auditory system in bony fishes have focused individually on the swimbladder or otoliths. This study represents the first time that individual models for the swimbladder, Weberian apparatus, and otoliths have been coupled and analyzed as a complete system.

I. BACKGROUND

The teleost ear contains three semicircular canals and three otolithic organs: the utricle, the lagena, and the saccule. Each otolithic organ consists of a single, dense otolith suspended in a fluid-filled chamber whose wall is lined with a sensory epithelium, or macula, containing mechanoreceptive hair cells. The hair-cell ciliary bundles project into the otolithic chamber and are coupled to the otolith through the otolithic membrane. Relative motion between the otolith and sensory epithelium induces bending of the hair-cell ciliary bundles and ultimately stimulates the auditory nerve. In addition to the ear, many fish possess auxiliary structures which are believed to enhance audition. The most common of these is the swimbladder, a gas-filled chamber (or chambers) located in the abdomen and primarily used to regulate buoyancy. The goldfish swimbladder has two chambers: the anterior swimbladder and the posterior swimbladder. The two chambers are connected by a short duct surrounded by a sphincter (Evans, 1925).

Fish receive sound through at least two pathways: the direct path, which is common to all fish, and the indirect path, which occurs only in fish with a swimbladder. In the direct path, the otoliths, which are about three times as dense as the rest of the fish, will move with an amplitude and phase different from that of the body of the fish, including the sensory epithelium, in the presence of an impinging acoustic wave (Dijkgraaf, 1960). This relative motion results in bending of the hair-cell ciliary bundles. In the indirect path, acoustic pressure acting on the swimbladder will cause the swimbladder walls to vibrate and radiate sound which will stimulate the otolithic organs in a manner directly analogous to the direct path (von Frisch, 1938; van Bergeijk, 1964).

In addition to the direct and indirect paths, a third path, which will be referred to as the ''Weberian path,'' exists in goldfish because of their Weberian ossicles. The Weberian ossicle arrangement in the right side of a goldfish is illustrated in Fig. 1(a). The largest ossicle, the tripus, connects posteriorly to the anterior swimbladder and anteriorly, through a ligament ($L3$), to the second ossicle, the interca-

[a]Current address: Space and Naval Warfare Systems Center, San Diego, 49620 Beluga Rd., San Diego, CA 92152.
[b]Author to whom correspondence should be addressed, Biomedical Engineering Center, The Ohio State University, 1080 Carmack Rd., Columbus, OH 43210-1002.

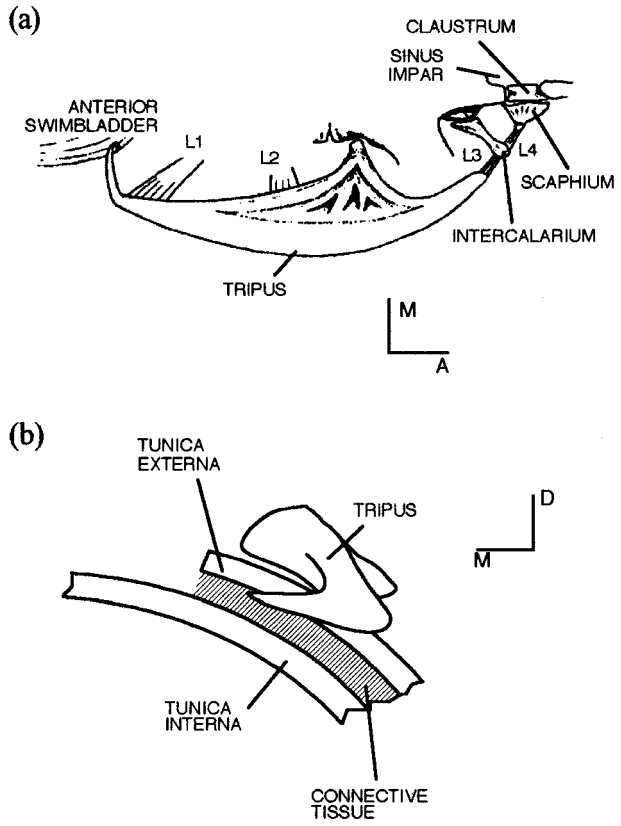


FIG. 1. (a) Weberian ossicle arrangement in the right side of a goldfish (Modified from Krumholz, 1943). $L1-L4$ are ligaments. A—anterior, M—medial. (b) Simplified representation of a transverse slice through the anterior swimbladder showing the relationship between the tunica interna, tunica externa, and the right tripus. D—dorsal, M—medial.

larium, which in turn is connected through a ligament ($L4$) to the third ossicle, the scaphium. A fourth ossicle, the claustrum, is present in some species, including the goldfish. The scaphium and claustrum are imbedded in the cartilaginous walls of the sinus impar, a fluid-filled canal which projects anteriorly and communicates with the transverse canal, which directly opens into the left and right saccular chambers (Chranilov, 1927).

Figure 1(b) shows a transverse slice through the right side of the goldfish anterior swimbladder. The anterior swimbladder has two walls: the inner wall, the tunica interna, is complete; the outer wall, the tunica externa, contains a longitudinal slit and connects to the posterior tip of the tripus. When acoustic pressure acts on the swimbladder, the tunica interna stretches, and the tunica externa not only stretches, but may also slide over the tunica interna. As the edges of the slit in the tunica externa move farther apart, the ossicles rotate forward. When the swimbladder contracts, the sequence is reversed and the ossicles move backwards. Thus, expansion of the swimbladder will cause forward motion of the ossicles and fluid flow into the saccule, while compression will cause fluid flow out of the saccule. The saccular otolith features delicate wing-like projections or flutes which lie in the path of the fluids within the transverse canal, so fluid motion caused by Weberian ossicle motion is likely to cause additional relative motion between the saccular otolith and the sensory epithelium.

Regardless of how sound reaches the ear, it can only be detected if bending of the hair-cell apical ciliary bundles occurs. In the direct path, the acoustic particle motion is the stimulus that brings about relative motion between otolith and macula. Acoustic pressure is the stimulus for both the indirect and Weberian paths.

Early mathematical analyses [e.g., Alexander (1966)] treated the swimbladder as a free spherical air bubble. Experimental measurements of the resonance frequency and damping of intact swimbladders in living fish, however, do not generally agree with those of a free spherical air bubble. At depths less than 200 m, experimental values for the resonance frequency and damping are both higher than predicted (Batzler and Pickwell, 1970; McCartney and Stubbs, 1971; Sand and Hawkins, 1973; Løvik and Hovem, 1979). Accordingly, the spherical gas-bubble model has been modified by several authors to include additional stiffness and damping terms.

Andreeva (1964), McCartney and Stubbs (1971), and Ye and Farmer (1994) treated the swimbladder as infinite, thin-walled, and finite spherical elastic shells, respectively, surrounding an air cavity. Love (1978) considered the swimbladder to be a spherical fluid shell (composed of a viscous, heat-conducting Newtonian fluid) enclosing an air cavity. Feuillade and Nero (1998) combined the elastic solid and fluid shell formulations. Their analysis featured an air cavity surrounded by an elastic shell, which was itself surrounded by a viscous fluid shell. deMunck and Schellart (1987) treated the swimbladder as a prolate spheroid air bubble and derived the scattered acoustic field. The only previous analysis of otolith mechanics was a simple lumped-parameter model proposed by deVries (1950, 1956).

Although single degree-of-freedom (1-DOF) analyses have been shown to work well for a single-chambered swimbladder (Derenburger *et al.*, 1997; Lewis, 1994), goldfish and other Cypriniformes have two-chambered swimbladders. Zhou (1992) and Lewis (1994) demonstrated the inadequacy of a 1-DOF analysis to fit the measured amplitude response of the goldfish swimbladder. In particular, the measured response of the goldfish swimbladder may exhibit twin peaks in the amplitude because of the individual resonances of the two swimbladder chambers.

II. ANALYSIS

The analysis presented here is broken down into three major subsystems for the swimbladder, the Weberian apparatus, and the saccule (which is considered the main acoustic receptor in goldfish). Equations are also required to describe the direct and indirect acoustic fields. Figure 2 is a simplified block diagram which illustrates the relationships among the subsystems. Only the steady-state frequency response is considered here. An attempt was made, where possible, to include the effects of fish size on the system parameters appearing in the equations.

The total ciliary bundle shear consists of individual contributions from the three pathways described previously: the direct path, the indirect path, and the Weberian path. The main goal of the analysis is to estimate not only the total ciliary bundle shear, but also its contributions from each of

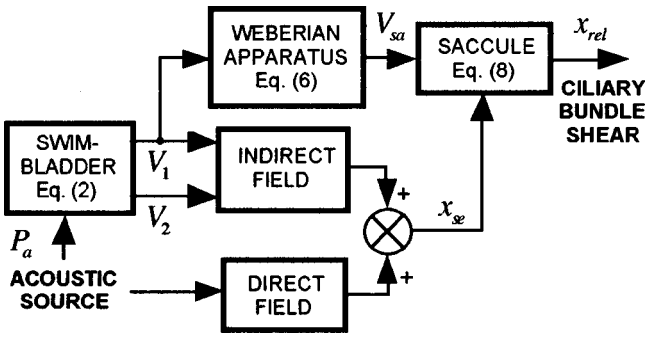


FIG. 2. Simplified block diagram showing the subsystem breakdown used in the analysis.

the three pathways. The response of each swimbladder chamber, the coupling between the anterior swimbladder and the Weberian apparatus, and the effects of various material and geometric properties are also considered.

A. Swimbladder

The goldfish swimbladder chambers are roughly prolate spheroidal in shape. Although Strasberg (1953) and Weston (1967) have shown that a nonspherical air bubble has an increased resonance frequency compared to a spherical bubble of the same volume, the effect is small for prolate or oblate spheroids of nearly spherical shape. For example, an oblate or prolate spheroid with a major axis to minor axis ratio of 2 exhibits an increase in resonance of only about 2% (Weston, 1967). Measurements of goldfish bladder dimensions from the radiographs of 30 individual fish revealed an average major/minor axis ratio of 1.44 (standard deviation 0.12) for the anterior swimbladder and 1.66 (standard deviation 0.33) for the posterior; therefore, the increase in resonance frequency for the goldfish is considered negligible and the bladders are approximated as spherical in shape. Accordingly, each swimbladder chamber is considered to be a spherical air bubble in water, with additional mass, stiffness, and damping added to account for the body tissues surrounding it.

The two bladder chambers are coupled by a stiffness k_{12} and a damping b_{12} to form a 2-DOF system, as shown in Fig. 3(a). The anterior and posterior bladder radii are R_1 and R_2, respectively; the bladder centers are separated by a distance d_{12}. The radial velocities of the anterior and posterior bladder walls are $V_1(\omega)$ and $V_2(\omega)$, respectively. The mass, stiffness, and damping for the anterior swimbladder are m_1, k_1, b_1, respectively, and for the posterior swimbladder m_2, k_2, b_2, respectively. These are acoustical quantities and thus have units of kg/m^2, N/m^3, and N·s/m^3, respectively.

In writing the equation of motion for each bladder, one must include both the incident (direct field) acoustic pressure and the scattered (indirect field) pressure from the other bladder. To assist this, the mass coupling coefficients m_{12} and m_{21} are defined, using

$$m_{mn}=\rho_w\frac{R_n^2}{d_{12}}\exp[-jk(d_{12}-R_n)], \tag{1}$$

where m_{mn} indicates the mass coupling effect of bladder n on bladder m, ρ_w is the density of water, k is the acoustic wave

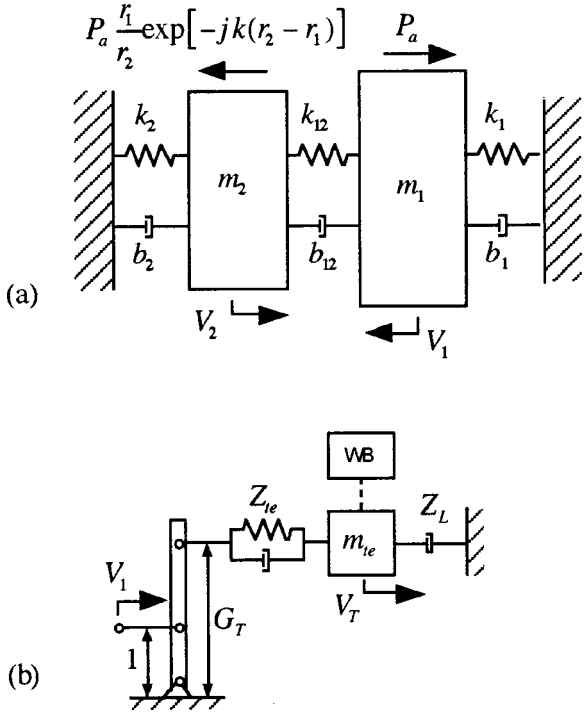


FIG. 3. (a) Two degree-of-freedom mechanical system representation for the anterior and posterior swimbladders. (b) Mechanical system representation of the tunica externa sliding motion. WB—Weberian apparatus subsystem.

number, and $j=\sqrt{-1}$. The resulting set of equations for the system of Fig. 3(a) is

$$\begin{bmatrix} Z_{\mathrm{sb},1}(\omega) & Z_{12}(\omega) \\ Z_{21}(\omega) & Z_{\mathrm{sb},2}(\omega) \end{bmatrix}\begin{bmatrix} V_1(\omega) \\ V_2(\omega) \end{bmatrix}=\begin{bmatrix} -P_a(\omega) \\ -H_{12}(\omega)P_a(\omega) \end{bmatrix}, \tag{2}$$

where

$$\begin{aligned} H_{12}(\omega)&=r_1/r_2\exp[-jk(r_2-r_1)] \\ Z_{\mathrm{sb},m}(\omega)&=(j\omega)m_m+b_m+b_{12}+(k_m+k_{12})/(j\omega) \\ Z_{mn}(\omega)&=(j\omega)m_{mn}+b_{12}+k_{12}/(j\omega), \end{aligned} \tag{3}$$

r_1 and r_2 are the distances from the acoustic source to the anterior and posterior bladders, respectively, $P_a(\omega)$ is the incident acoustic pressure at the anterior swimbladder, and ω is the circular frequency.

Equations (1)–(3) describe the radial motion of the two swimbladder chambers. The coupling between the anterior bladder and the Weberian ossicles is accomplished using a mechanical system model for the tangential motion of the tunica externa. It is assumed that the loading effect of the additional subsystems is negligible.

1. Properties of the swimbladder

Most existing analyses of the swimbladder require knowledge of the complex shear modulus for its surrounding tissues. To the authors' knowledge, the only available shear modulus data are from Lebedeva (1965). Because these data are *post mortem*, from nonotophysans, and do not cover the lower frequencies of interest here, it was decided to estimate the mass, stiffness, and damping of each swimbladder by fitting the model predictions to experimentally determined

TABLE I. Estimated mechanical properties for the swimbladder model.

Parameter	Range	Units
$k_1 \cdot R_1$	$4\times10^5-5\times10^6$	N/m^2
$b_1 \cdot R_1$	30–500	$N\cdot s/m^2$
m_1/R_1	600–5600	kg/m^3
$k_2 \cdot R_2$	$4\times10^5-1\times10^7$	N/m^2
$b_2 \cdot R_2$	25–420	$N\cdot s/m^2$
m_2/R_2	760–9150	kg/m^3
k_{12}	$2\times10^7-3\times10^8$	N/m^3
b_{12}	0–12 600	$N\cdot s/m^3$

frequency response data, rather than calculating these parameters from existing theories. The available data consist of the *in vivo* amplitude and phase response of the anterior and posterior bladders of 23 goldfish, ranging in size from 14 to 74 g and 79 to 136 mm. These data were collected using a noninvasive, noncontact ultrasonic measurement system developed by Rogers and Hastings (1989) and modified by Finneran and Hastings (1996) to allow phase angle measurement. The measurement system and experimental methodology are described by Cox and Rogers (1987). In addition to these experimental data, a limited amount of data were obtained for the *uncoupled* response of the individual swimbladder chambers; that is, the response of one bladder with the second deflated [see Derenburger (1997), for the deflation procedure].

The model predictions were fit to both the uncoupled and coupled data. For the uncoupled system, the fit was accomplished using a modified version of Levy's method (Levy, 1959; Sanathanan and Koerner, 1963). The fit to the coupled data was performed by manually adjusting the various parameters. In each case, the amplitudes were normalized with respect to $\rho_w c_w \sqrt{E(\omega)/\rho_w}$, where c_w is the sound speed in water and $E(\omega)$ is the acoustic energy density. The phase data were normalized with respect to the phase of the acoustic pressure. For the analysis, the incident acoustic field was tailored to match the experimentally measured pressure and particle velocity.

Table I summarizes the results. To properly compare and combine the curve-fit results, the data were normalized. The swimbladder parameters were assumed to vary with bladder radius according to the elastic shell model of Andreeva (1964), so $k_m = f(R^{-1})$, $b = f(R^{-1})$, and $m = f(R)$. The parameters in Table I, with the exception of k_{12} and b_{12}, were therefore normalized with respect to the proper bladder radius. The coupling stiffness and damping were not normalized because their variation with the size of either bladder is unknown. For comparison, again using Andreeva's model, the values for the stiffnesses result in shear moduli within the range $2\times10^4-3\times10^6\ N/m^2$, which is within the same order of magnitude of the range cited by Lebedeva ($10^5-10^6\ N/m^2$).

B. Weberian apparatus

The analysis of the Weberian apparatus consists of lumped-parameter mechanical models for the Weberian ossicles and fluid canals which are coupled to the 2-DOF system model for the swimbladder described previously. The coupling is achieved using a phenomenological analysis of the sliding motion of the tunica externa.

1. Tunica externa

As suggested by Chranilov (1927, 1929), it is assumed that the tunica interna is limited to stretching with swimbladder expansion, but that the tunica externa may not only stretch, but also slide over the tunica interna by virtue of the longitudinal slit it contains. Furthermore, the motion of the largest Weberian ossicle, the tripus, is assumed to follow the motion of the edges of the slit in the tunica externa.

The mean radius, thickness, and mass of the tunica externa are R_{te}, h_{te}, and m_{te}, respectively. The longitudinal slit in the tunica externa is approximated as circular with a radius T_1. The connective tissue within the gap between the tunica interna and tunica externa is treated as an incompressible fluid with dynamic viscosity ξ_L. The tangential motion of the tunica interna is assumed to be zero and a linear velocity profile is assumed in the connective tissue.

The equation of motion for the tunica externa may be manipulated to arrive at

$$(j\omega)m_{te}V_T(\omega) = Z_{te}(\omega)[G_T V_1(\omega) - V_T(\omega)] - Z_L V_T(\omega), \tag{4}$$

where

$$Z_{te}(\omega) = \frac{4\pi R_{te} h_{te} T_1}{(j\omega)J_{te}(\omega)(4R_{te}^2 - T_1^2)}$$
$$Z_L = 2\pi R_{te}^2 \xi_L / h_L \tag{5}$$
$$G_T = 4R_{te}/T_1,$$

where $J_{te}(\omega)$ is the tunica externa compliance and h_L is the thickness of the gap between the tunica externa and tunica interna. The system of Eq. (4) has the mechanical analog shown in Fig. 3(b), where the coupling to the anterior swimbladder occurs through the rigid, massless lever at the left. The tunica externa mechanical impedance $Z_{te}(\omega)$ is shown graphically as a Voigt element; however, the actual form of $Z_{te}(\omega)$ depends on the functional form of $J_{te}(\omega)$. Although the effects of the tripus and other Weberian ossicles are neglected, the attachment point for the Weberian apparatus subsystem is shown.

2. Weberian ossicles and fluid canals

Figure 4 shows the mechanical system model for the Weberian ossicles combined with the tunica externa model of Fig. 3(b). The tripus and intercalarium are represented by the rotational inertias J_{tr} and J_{ic}. The scaphium and claustrum are assumed to move together as a rigid body with rotational inertia J_{sc}. The medial surface of the claustrum forms a part of the wall of the atrium sinus impar; thus, the motion of the scaphium/claustrum body is opposed by the pressure within the sinus impar, $P_{si}(\omega)$. The cross-sectional (wetted) area of the claustrum is A_{cl}. The ossicles are free to rotate about pinned (frictionless) joints with the vertebrae. The Weberian apparatus ligaments $L1-L4$ [see Fig. 1(a)]

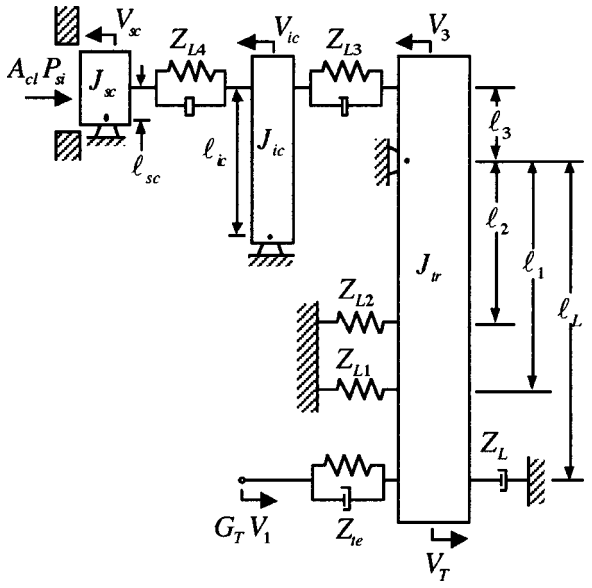


FIG. 4. Mechanical system representation of the Weberian ossicles.

are represented by the mechanical impedances $Z_{L1}(\omega)$, $Z_{L2}(\omega)$, $Z_{L3}(\omega)$, and $Z_{L4}(\omega)$. The velocities indicated in Fig. 4 are defined at the lever-arm locations. The mass of the tunica externa is negligible compared to the inertia of the tripus.

Figure 5(a) shows the fluid component of the Weberian apparatus. It is considered to be a rigid-walled cylindrical pipe system with three geometrically different sections. The fluid within each section is treated as incompressible. Section 1 represents the sinus impar, which contains perilymph. Section 2 represents the sinus endolymphaticus, which projects posteriorly into the sinus impar and contains endolymph. A thin membrane exists between the sinus impar and the posterior projection of the sinus endolymphaticus; thus, fluid motion within the sinus impar may be transmitted to the sinus endolymphaticus. This membrane is neglected in the formulation; continuity of pressure and volume velocity is assumed at the junction. Section 3 represents the transverse canal, which connects the sinus endolymphaticus to the left and right saccular chambers and contains endolymph. The ventro-medial wall of each saccular chamber contains a thin, membraneous region (deBurlet, 1929; von Boutteville, 1935; Alexander, 1966). This feature has also been described in certain catfish by Jenkins (1977), who used the term *release membrane*. The presence of the relatively compliant release membrane increases the overall compliance of the saccular chamber.

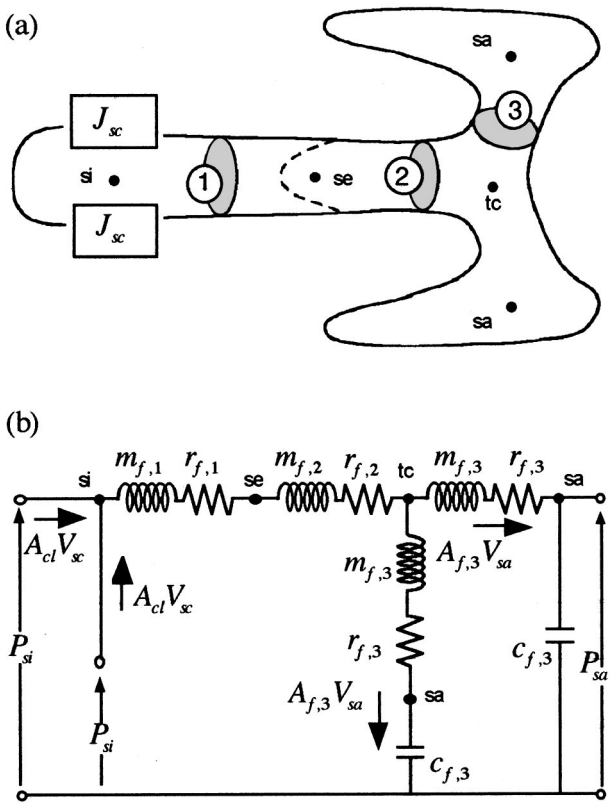


FIG. 5. Weberian apparatus fluid canal system. (a) Simplified representation, and (b) equivalent circuit model.

Figure 5(b) shows an equivalent circuit for the fluid system of Fig. 5(a). The acoustic wavelengths are large relative to the dimensions of the canals; thus, each one is treated as a single lump. The fluid sections have cross-sectional areas $A_{f,1}$, $A_{f,2}$, $A_{f,3}$, lumped fluid inertias $m_{f,1}$, $m_{f,2}$, $m_{f,3}$, and lumped fluid resistances $r_{f,1}$, $r_{f,2}$, $r_{f,3}$. Each saccular chamber has a fluid capacitance $c_{f,3}$. The pressure within each saccular chamber is $P_{\mathrm{sa}}(\omega)$.

When the Weberian ossicles and fluid canals are coupled, the resulting system is described by the set of equations

$$\begin{bmatrix} Z_{\mathrm{tr}}(\omega) & -Z_{L3}(\omega) & 0 & 0 \\ -Z_{L3}(\omega) & Z_{\mathrm{ic}}(\omega) & -Z_{L4}(\omega) & 0 \\ 0 & -Z_{L4}(\omega) & Z_{\mathrm{sc}}(\omega) & A_{\mathrm{cl}} \\ 0 & 0 & A_{\mathrm{cl}} & -1/Z_{\mathrm{fc}}(\omega) \end{bmatrix} \times \begin{bmatrix} V_3(\omega) \\ V_{\mathrm{ic}}(\omega) \\ V_{\mathrm{sc}}(\omega) \\ P_{\mathrm{si}}(\omega) \end{bmatrix} = \begin{bmatrix} G_L G_T Z_{\mathrm{te}}(\omega) V_1(\omega) \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tag{6}$$

where $G_1 = l_1/l_3$, $G_2 = l_2/l_3$, $G_L = l_L/l_3$, and

$$\begin{aligned} Z_{\mathrm{tr}}(\omega) &= (j\omega)\frac{J_{\mathrm{tr}}}{l_3^2} + G_1^2 Z_{L1}(\omega) + G_2^2 Z_{L2}(\omega) + G_L^2 Z_{\mathrm{te}}(\omega) \\ &\quad + G_L^2 Z_L + Z_{L3}(\omega) \\ Z_{\mathrm{ic}}(\omega) &= (j\omega)\frac{J_{\mathrm{ic}}}{l_i^2} + Z_{L3}(\omega) + Z_{L4}(\omega) \\ Z_{\mathrm{sc}}(\omega) &= (j\omega)\frac{J_{\mathrm{sc}}}{l_{\mathrm{sc}}^2} + Z_{L4}(\omega) \\ Z_{\mathrm{fc}}(\omega) &= (j\omega)(2m_{f,1} + 2m_{f,2} + m_{f,3}) \\ &\quad + (2r_{f,1} + 2r_{f,2} + r_{f,3}) + \frac{1}{(j\omega)c_{f,3}}. \end{aligned} \tag{7}$$

The velocity of fluid entering the saccule, $V_{\mathrm{sa}}(\omega)$, is found using $A_{f,3}V_{\mathrm{sa}}(\omega) = A_{\mathrm{cl}}V_{\mathrm{sc}}(\omega)$.

3. Properties of the Weberian apparatus

The tunica externa thickness and gap thickness were based on measurements from dissected animals and from data presented by Alexander (1961). The viscosity of the tissue within the gap between the tunica externa and tunica interna was estimated to lie within the range 0.01–1.0 $N \cdot s/m^2$. The ratio T_1/R_1 was based on measurements from dissected animals.

The tunica externa compliance was estimated by curve-fitting *post mortem* creep test data from isolated samples of the tunica externa and tunica interna from several species of Cypriniformes (Alexander, 1961). A power function with the form $A_{te}(t/t_0)^n$, where A_{te} and n are constants, was fit to the data using a least-squares technique. The time t was normalized by the constant $t_0 = 1$ s in order to keep A_{te} in units of m^2/N. The least squares solution yielded $A_{te} = 3 \times 10^{-6}\ m^2/N$ and $n = 0.26$ ($R^2 = 0.893$). Since Alexander's data were *post mortem*, one would expect the *in vivo* compliance to be lower than the curve-fit result.

The Weberian ossicle rotational inertias were estimated by modeling each ossicle as being composed of a number of simple geometric solids. The dimensions of the ossicles and the lever-arm lengths in Fig. 4 were estimated from direct examination of goldfish and from sketches and data presented by Chranilov (1929, 1927), Watson (1939), Alexander (1962), Weber (1820), and Krumholz (1943) for a number of different Cypriniformes. The lever-arm lengths and ossicle dimensions were assumed to be proportional to the fish length. The density of the ossicles, ρ_{wb}, equals that of fish bone, which is 1570–2040 kg/m^3 for roach, carp, and dace (Alexander, 1959).

The fluid canal dimensions were determined from dissections and sketches presented by Chardon and Vandewalle (1991), Popper (1971a), and von Frisch (1938). The density and viscosity of perilymph (section 1) were based on pigeon data from Money *et al.* (1971) and equal 1008 kg/m^3 and $0.76 \times 10^{-3}\ N \cdot s/m^2$, respectively. The density and viscosity of endolymph (sections 2 and 3) were from ten Kate and Kuiper (1970), for the pike, and equal 1010 kg/m^3 and $1.20 \times 10^{-3}\ N \cdot s/m^2$, respectively.

The fluid resistance and inertance were calculated using expressions for incompressible fluid in rigid, circular pipes. Separate expressions were used for low frequencies, where $\omega < 32\xi_f \rho_f^{-1} d_f^{-2}$ (ρ_f is the fluid density, d_f is the pipe diameter, and ξ_f is the fluid dynamic viscosity) and for high frequencies, where $\omega > 7200\xi_f \rho_f^{-1} d_f^{-2}$. Linear interpolation was used at intermediate frequencies.

The compliance of the saccular capsule was estimated by modeling it as a fluid-filled, spherical cavity surrounded by an infinite elastic shell. The release membrane was treated as a uniformly loaded thin circular membrane of radius R_{rm} and thickness t_{rm} with clamped edges. Assuming that R_{rm} is approximately 1.0 mm, t_{rm}/R_{rm} is 0.24, and the saccular chamber wall has properties similar to collagen, then $c_{f,3}$ is approximately $3 \times 10^{-17}\ m^3/Pa$. If there is no release membrane, the saccular compliance is controlled by the elasticity of the saccular chamber itself, estimated to be $8 \times 10^{-18}\ m^3/Pa$.

C. Saccule

In the goldfish saccule, the sensory epithelium lies roughly in a vertical plane with its longitudinal axis oriented at approximately 30° with respect to the rostral–caudal axis of the fish (Fay and Olsho, 1979). The goldfish saccule possesses a vertical hair-cell orientation pattern (Popper and Fay, 1993); therefore, the hair-cell ciliary bundles are assumed to be oriented 120° and −60° with respect to the rostral–caudal axis. For this study, only translation of the saccular otolith, or sagitta, within a vertical plane, in the direction of the ciliary bundle orientation was considered. Although the sagitta likely moves in a complex manner including rotation and translation in three dimensions, given the lack of data on its structure and material properties and those of the surrounding structures, a three-dimensional analysis including rotation was beyond the scope of this work. For this reason the analysis of the sagitta motion was limited to one dimension. It seemed appropriate, in considering only a single direction of sagitta motion, to perform the analysis in the vertical direction, the orientation direction of the saccular hair-cell ciliary bundles.

Furukawa and Ishii (1967a, 1967b) reported that in goldfish the ventral fluting of the saccular otolith is attached to the release membrane. For this study, the release membrane is included in the calculation of the saccular chamber compliance; however, because of a lack of conclusive evidence, no direct connection between the otolith and the saccular chamber or the release membrane is assumed.

The saccule is assumed to contain N_{hc} identical hair cells which are distributed uniformly across the macula. The mechanical impedance of a single hair-cell ciliary bundle to shear along its axis is $Z_{hc}(\omega)$. The hair-cell ciliary bundles are coupled to the otolith through the otolithic membrane, which is assumed to be homogeneous and isotropic. For the analysis, the otolithic membrane is divided into elemental regions which couple each individual ciliary bundle to the otolith. The mechanical impedance of each of these identical regions is $Z_{om}(\omega)$.

Figure 6(a) shows the mechanical model for the saccule. The sagitta has a density ρ_{sg}, volume v_{sg}, and mass m_{sg}. It is coupled to the sensory epithelium through the otolithic membrane and hair-cell ciliary bundle mechanical impedances $N_{hc}Z_{om}(\omega)$ and $N_{hc}Z_{hc}(\omega)$, which represent the cumulative effects of the N_{hc} ciliary bundles and elemental regions of the otolithic membrane. The saccular chamber is filled with endolymph having a density ρ_e and a dynamic viscosity ξ_e. Fluid displaced from the transverse canal enters the saccular capsule with a velocity $V_{sa}(\omega)$ relative to that of the fish's body. The main goal is to find the ciliary bundle shear, defined as the relative displacement $x_{rel}(\omega) = x_{hc}(\omega) - x_{se}(\omega)$, where $x_{hc}(\omega)$ is the displacement of the hair-cell ciliary bundle tip and $x_{se}(\omega)$ is the displacement of the sensory epithelium.

Since the acoustic impedance of a fish is roughly identical to that of the surrounding water, the sensory epithelium displacement will closely follow that of the surrounding water, which, within an acoustic field, is a harmonic function of time. The mechanical system model in Fig. 6(a) thus represents a body in an accelerated fluid. Therefore, in addition to

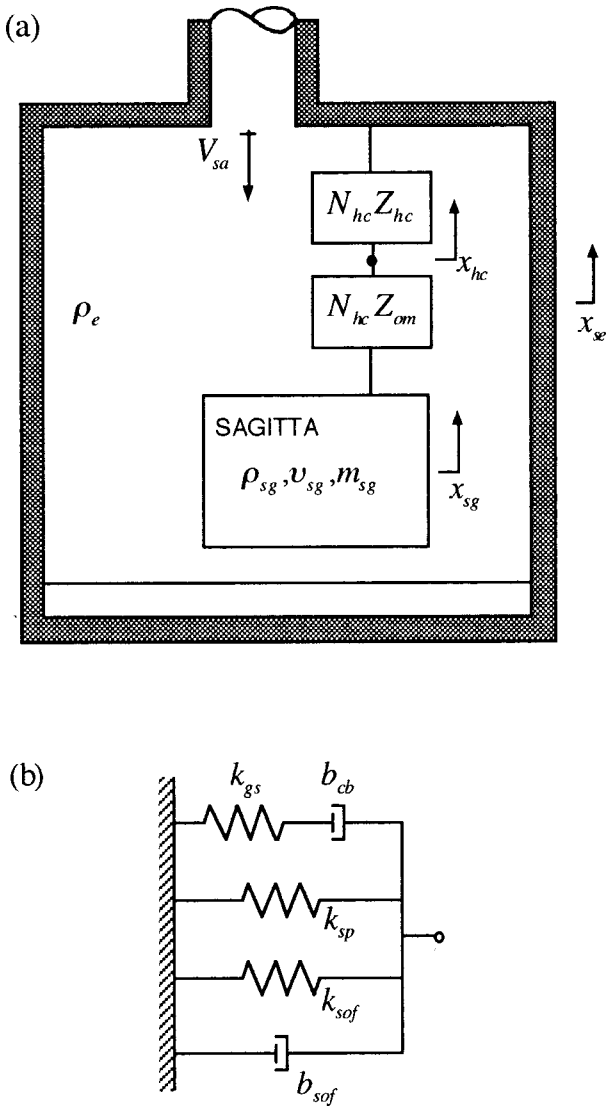


FIG. 6. (a) One-dimensional mechanical system representation of the saccule. (b) Mechanical system model for the single hair-cell ciliary bundle.

viscous drag and the reaction of the otolithic membrane, the sagitta also experiences a ''lift'' force proportional to the acceleration of the surrounding fluid and a drag force proportional to the *relative* acceleration between the otolith and the surrounding fluid and the ''apparent'' volume of the sagitta, $\upsilon'_{\rm sg}$, which is the volume of the fluid displaced by the moving otolith (Prandtl, 1952). Writing the equation of motion for the otolith and solving for $x_{\rm rel}(\omega)$ yields

$$x_{\rm rel}(\omega) = -\frac{Z_m(\omega)}{Z_{\rm sg}(\omega)} x_{\rm se}(\omega) - \frac{Z_f(\omega)}{(j\omega) Z_{\rm sg}(\omega)} V_{\rm sa}(\omega), \qquad (8)$$

where

$$Z_{\rm sg}(\omega) = (j\omega) m_e \left[1 + \frac{Z_{\rm hc}(\omega)}{Z_{\rm om}(\omega)}\right] + N_{\rm hc} Z_{\rm hc}(\omega) + Z_{\rm vis}(\omega)\left[1 + \frac{Z_{\rm hc}(\omega)}{Z_{\rm om}(\omega)}\right]$$

$$Z_m(\omega) = (j\omega) m_{\rm sg}(1 - \rho_e/\rho_{\rm sg}) \qquad (9)$$

$$Z_f(\omega) = (j\omega)\rho_e(\upsilon_{\rm sg} + \upsilon'_{\rm sg}) + Z_{\rm vis}(\omega),$$

and the effective mass m_e, defined as the mass of the sagitta plus any entrained fluid, is $m_e = m_{\rm sg} + \rho_e \upsilon'_{\rm sg}$. The velocity $V_{\rm sa}(\omega)$ is known from the Weberian apparatus model. The sensory epithelium displacement $x_{\rm se}(\omega)$ is assumed to match the acoustic particle displacement in the direction of the ciliary bundles.

1. Properties of the saccule

The sagitta density and mass are based on data from deVries (1956) and Furukawa and Ishii (1967a), respectively. The apparent volume is assumed to equal the otolith volume. The relationship between the fluid velocity and viscous drag on the sagitta, $Z_{\rm vis}(\omega)$, was estimated using expressions presented by Stokes (1851) and Freeman and Weiss (1985) for a circular cylinder immersed in a uniform flow (the effects of the saccular capsule wall are ignored). The sagitta radius and length are from direct measurements and from Adams (1940); the resulting values were then increased to reflect the increased surface area of the sagitta compared to a cylinder. The viscosity ξ_e was also increased to account for the jelly-like structure of the otolithic membrane.

For a single 51-mm goldfish, Platt (1977) reported 7494 individual hair cells in the saccular macula. For the lagenar macula, he reported 3651 large-bundled hair cells in a 44-mm goldfish and 4852 in a 51-mm goldfish. The number of hair cells therefore appears to increase with the size (i.e., the age) of a fish. This conclusion is also supported by Lombarte and Popper (1994), and Popper and Hoxter (1984). Using the hair-cell densities for the saccular macula of five goldfish (Platt, 1977), the average density is estimated as 336 hair cells HC for an area of $100\times100\,\mu$m, or 336 HC/$10^4\,\mu$m^2. Using the average density and the total number of hair cells above, the macula area for a 51-mm fish is estimated as $22.3\times10^4\,\mu$m^2. The manner in which the macula area scales with the size of the fish is unknown; however, since the saccular macula is rather elongated, it is assumed that the macula area scales with the length of the fish. The proportionality constant relating $N_{\rm hc}$ and $L_{\rm gf}$ is therefore 147 HC/mm.

The model for the individual hair-cell ciliary bundles is a modification of those presented by Howard and Hudspeth (1987) and Benser *et al.* (1993). Howard and Hudspeth's model was considered problematic because it predicts an instantaneous response to a step input. Corey and Hudspeth (1983) have shown that the time lag between an applied force and the resulting bundle deflection is on the order of 20–25 μs for a ciliary bundle from the bullfrog saccule; therefore, Howard and Hudspeth's model was modified to include the subotolithic stiffness $k_{\rm sof}$, as suggested by Benser *et al.* (1993), as well as a damping, presumably also caused by the presence of subotolithic filaments, in parallel with the subotolithic stiffness. This damping, $b_{\rm sof}$, also includes any viscous damping from fluid drag on the bundle as it moves within the endolymph.

Figure 6(b) shows the resulting mechanical system model. The ciliary bundle itself is represented by a standard linear solid consisting of a damping $b_{\rm cb}$ and two stiffnesses, one attributed to the stereociliary pivot ($k_{\rm sp}$) and one arising

TABLE II. Parameter values for the ciliary bundle model.

Parameter	Range	Units	Reference/Notes
k_{cb}	0.55–2.15	mN/m	Benser *et al.* (1993)
k_{gs}/k_{cb}	0.225	···	Howard and Hudspeth (1987)
k_{sp}/k_{cb}	0.375	···	Howard and Hudspeth (1987)
k_{sof}/k_{cb}	0.4	···	Benser *et al.* (1993)
b_{cb}	6	μNs/m	Howard and Hudspeth (1987)
b_{sof}	0.011–0.043	μNs/m	Corey and Hudspeth (1983) Benser *et al.* (1993)

from the gating spring (k_{gs}). In parallel with the ciliary bundle elements exist a stiffness and damping caused by the presence of subotolithic filaments that connect the otolithic membrane to the sensory epithelium. Nonlinearities present in the ciliary bundle response are neglected. The damping b_{cb} was given by Howard and Hudspeth (1987) as approximately 6 μN·s/m for saccular hair cells in the bullfrog. Benser *et al.* (1993) indicated that the ciliary bundle elements account for 50%–70% of the total stiffness in Fig. 6(b), again for a bullfrog. Howard and Hudspeth (1987) stated that the gating spring stiffness is roughly 60% of the stereociliary pivot stiffness. Using this information and measured values for the total ciliary bundle stiffness k_{cb} (Benser *et al.*, 1993), the magnitudes of the individual stiffnesses in Fig. 6(b) were estimated. The subotolithic filament damping b_{sof} was estimated from the total ciliary bundle stiffness along with a time constant of 20 μs (Corey and Hudspeth, 1983). Table II summarizes these estimates.

The otolithic membrane was assumed to be perfectly rigid and to transmit the motion of the otolith evenly to all ciliary bundles. The otolith is therefore fixed relative to the ciliary bundle tip; that is, the otolith motion is identical to the ciliary bundle tip motion. This is essentially the same treatment as in Kachar *et al.* (1990). Benser *et al.* (1993) experimentally demonstrated that the otolithic membrane in the bullfrog saccule does not move as a rigid body; when excited at a point location the response amplitude declined at successive points away from the stimulus. However, if the excitation of the otolithic membrane is assumed to be uniformly distributed, this effect is negligible. Also, Corey and Hudspeth (1983) have shown that the phase lag produced by the otolithic membrane is negligible at the frequencies of interest here; thus, the rigid body assumption may be justified. For the rigid body case, $Z_{om}(\omega)\rightarrow\infty$ and $x_{sg}(\omega)\approx x_{hc}(\omega)$.

III. RESULTS AND DISCUSSION

Since the analytical predictions strongly depend on the choice of parameters, experimental data for the motions of the swimbladders and tripus of a representative fish were used to guide the choice of parameters in order to closely match the results with the experimental data. The parameters were then varied from these baseline values to show the relative impact of each on the overall system response.

A. Swimbladder and Weberian apparatus

Figures 7–9 compare experimental data for a goldfish (53.7 g; 126 mm) to the analytical predictions for a fish of the same size. Table III lists values of the most significant system parameters. The swimbladder dimensions were based on *in vivo* radiographs of intact fish. The other parameters were adjusted to get the best visual fit to the experimental data.

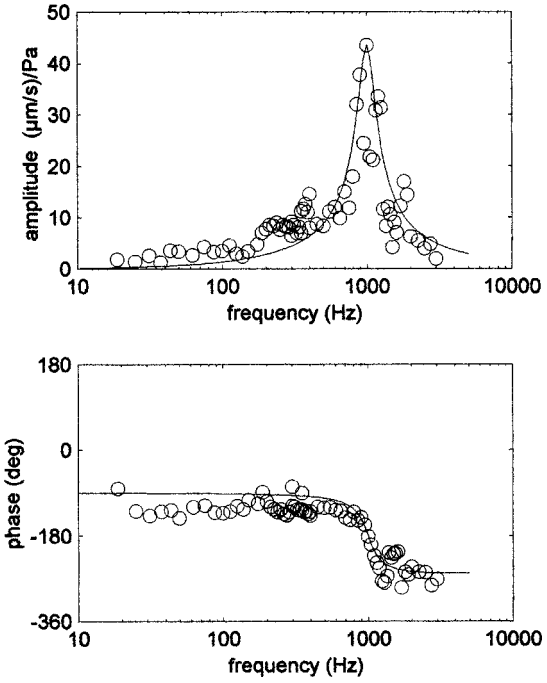


FIG. 7. Comparison of the analytical predictions for the anterior swimbladder radial velocity with experimental measurements. Solid line—analysis; symbols—experiment. The swimbladder system parameters are listed in Table III.

Figures 7 and 8 show the results for the anterior and posterior swimbladders, respectively. The bladder reso-

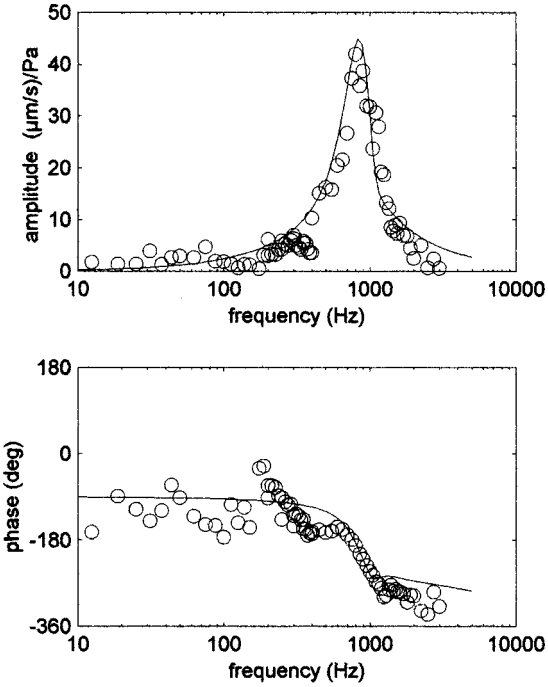


FIG. 8. Comparison of the analytical predictions for the posterior swimbladder radial velocity with experimental measurements. Solid line—analysis; symbols—experiment. The swimbladder system parameters are listed in Table III.

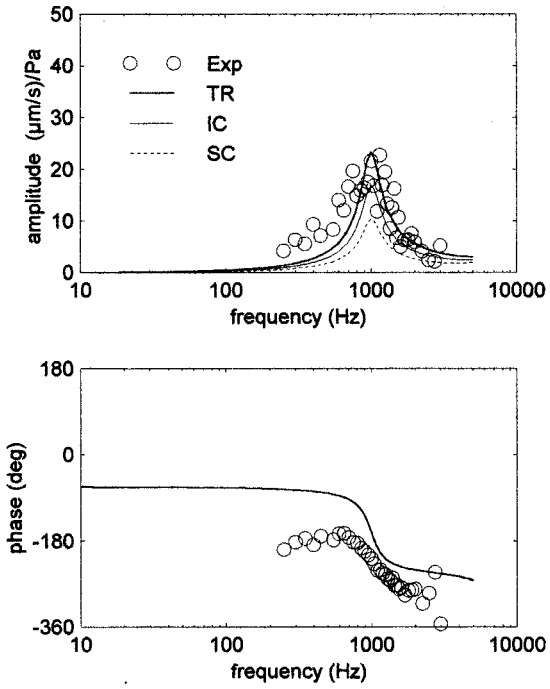


FIG. 9. Comparison of the analytical predictions for the Weberian ossicles and experimental measurements for the tripus. Solid line—analysis; symbols—experiment. The system parameters are listed in Table III.

nances are visible near 1000 Hz. Below the resonance frequencies, the phase angles are near $-90°$; above resonance, the phase angles approach $-270°$. The analysis results for both bladder chambers agree reasonably well with the experimental data, especially near the resonance frequency. The quality of the agreement and the fact that the system parameters are within the bounds considered physically reasonable verifies the form of the swimbladder system equations.

Figure 9 compares experimental data for the anterior portion of the tripus to the predictions for the motion of the tripus, intercalarium, and scaphium. The analytical and experimental amplitudes are reasonably close; however, the experimental curve is more heavily damped. This trait is also exhibited in the phase, which also shows a consistent lag between the experimental and calculated values. The Weberian ossicles are in phase with each other and with the anterior swimbladder. Using the sign conventions of Fig. 4, these phase relationships confirm the expected operation of the Weberian apparatus: positive pressure causes inward radial motion of the anterior swimbladder, forward rotation of the ossicles, and fluid flow into the sinus impar.

TABLE III. Baseline parameters for the swimbladder model.

Parameter	Value	Units
Fish mass	53.71	g
R_1	8.14	mm
R_2	6.77	mm
$k_1 \cdot R_1$	3.76×10^6	N/m^2
$b_1 \cdot R_1$	156.4	N·s/m^2
m_1/R_1	1733	kg/m^3
$k_2 \cdot R_2$	1.68×10^6	N/m^2
$b_2 \cdot R_2$	170.4	N·s/m^2
m_2/R_2	1790	kg/m^3
k_{12}	7.62×10^7	N/m^3
b_{12}	7280	N·s/m^3
A_{te}	6×10^{-7}	m^2/N

Alexander (1961) proposed that the viscoelastic properties of the tunica externa produce a high-pass filter effect that prevents low-frequency pressure changes from causing displacement of the Weberian ossicles. To examine this, the amplitude ratios between the Weberian ossicles and the anterior swimbladder were computed. Each ratio exhibits a high-pass effect; however, the low-frequency roll-off is only 5 dB/decade, which seems too gradual to prevent a change of depth from resulting in a significant input to the Weberian ossicles. If a change in depth results in a pressure input with frequency content up to only 0.05 Hz, the Weberian ossicle input will be roughly -20 dB relative to the same stimulus applied at 500 Hz. The pressure changes accompanying a change in depth are several orders of magnitude larger than those of an acoustic wave: a 1-m change in depth results in a pressure change of roughly 9800 Pa, which is almost 200 dB $re{:}1\ \mu$Pa. Although movement to a lower depth would increase the swimbladder stiffness (as a result of the larger hydrostatic pressure) and thus lower the bladder response at low frequencies, a depth change of 1 m would only result in a stiffness increase of 10%, which is insufficient to offset the massive pressure change. It thus appears that the high-pass filters within the Weberian apparatus model would not be able to prevent relatively large inputs from passing to the Weberian ossicles. It is likely that nonlinearities neglected in this analysis prevent overstimulation of the Weberian apparatus at low frequencies; for example, as the motions input to the ossicles increase, the ligament stiffnesses may increase accordingly. It may also be possible for the goldfish to transfer gas from the posterior to the anterior swimbladder (or the reverse) in order to keep the anterior swimbladder volume relatively constant within a range of depths. Alexander (1966) stated that proprioceptors in the bladder walls of Cyprinidae allow detection of swimbladder volume changes, and Evans (1925) suggested that the swimbladder nerve and sphincter muscle arrangement appears to act as a mechanism for maintaining uniform pressure in the anterior swimbladder.

Changes to the mass, stiffness, or damping of a swimbladder affect its response in a fashion typical of a mechanical second-order system, but have little effect on the second bladder, with the following exception: If the bladder resonance frequencies are sufficiently far apart, then a second peak, corresponding to the resonance of the other bladder, may appear. Figure 10 shows an example of multiple resonances appearing in the experimentally measured response of the posterior swimbladder of a goldfish (20.7 g; 88 mm). Also shown in Fig. 10 is the theoretical curve for a fish of the same size, with swimbladder parameters again adjusted to give the best fit. Figure 10 indicates that the analysis is capable of predicting multiple resonances in the swimbladder response.

The coupling parameters k_{12} and b_{12} act similar to the parameters for the individual bladders, but affect both bladders. Larger values for the coupling coefficients also increase

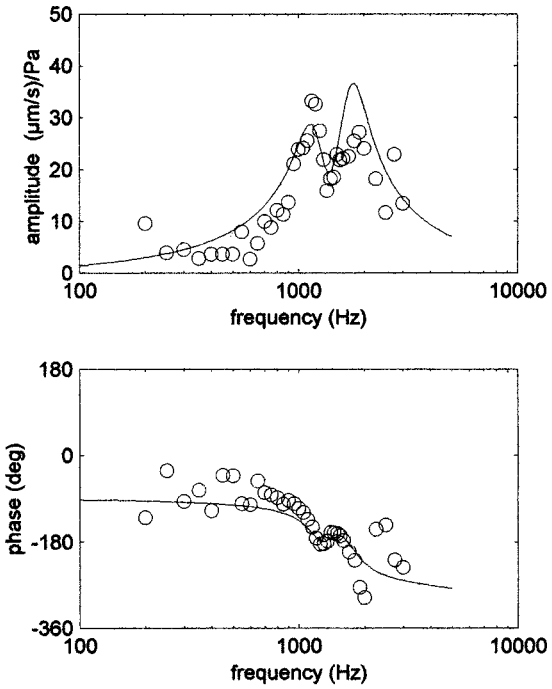


FIG. 10. Example of experimental and analytical results showing multiple resonances in the response of the posterior swimbladder.

the sensitivity of one bladder to changes in the properties of the other bladder.

The most significant parameters for the Weberian apparatus are the tunica externa compliance and thickness, the stiffness of ligament $L1$, the transverse canal diameter, and the saccular chamber compliance. The stiffness of the tunica externa is represented using the ratio $K_{te}=h_{te}/A_{te}$, rather than considering its compliance and thickness individually. Figure 11(a) shows the effect of K_{te} on the velocity of the tripus, $V_3(\omega)$. Increasing the stiffness causes an increase in the tripus velocity amplitude; decreasing the stiffness reduces the tripus amplitude. The stiffness of the tunica externa determines the degree to which the tunica externa slides rather than stretches; a stiff tunica externa will not stretch and must slide over the tunica interna. Since the motion of the tripus is directly related to the sliding motion of the tunica externa, a stiff tunica externa results in large tripus motion. The phase of the tripus is not affected by the tunica externa stiffness.

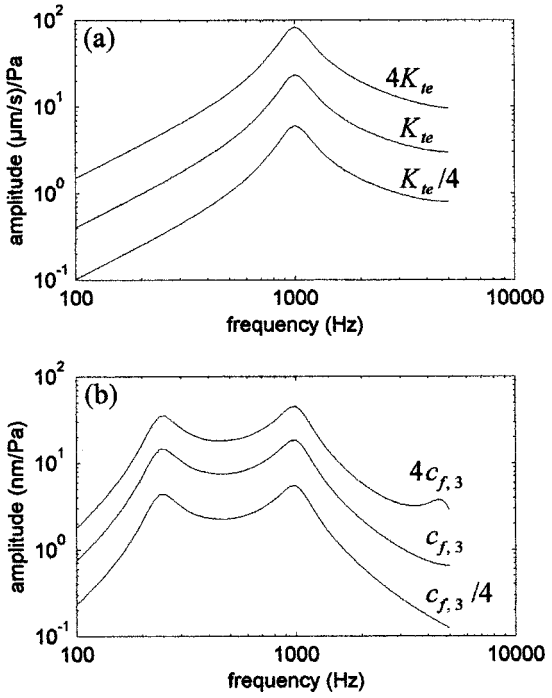


FIG. 11. Effects of varying Weberian apparatus parameters. (a) Effect of tunica externa stiffness K_{te}. (b) Effect of saccule compliance $c_{f,3}$.

Analogs to Fig. 11(a) may be generated by varying the stiffness of ligament $L1$, the transverse canal diameter, or the saccular chamber compliance. Although the magnitudes of the changes will be different, the general shape of the curves is preserved. Increasing the saccular compliance causes an upward shift in the curves of Fig. 11(a) (the same effect as changes to K_{te}); increasing the canal diameter and ligament stiffness cause a downward shift in Fig. 11(a) (opposite effect of K_{te}). At small diameters and high saccular compliances, an additional resonance appears at the extreme upper frequencies.

Figure 11(b) shows the effects of the saccular chamber compliance, $c_{f,3}$, on the relative displacement between the sagitta and the sensory epithelium. Varying the saccular chamber compliance causes a shift in the normalized displacement amplitude curves. A second resonance also appears at the upper frequency range when the compliance is high. Below the second resonance, the phase angle does significantly change with changes in the saccular chamber compliance. Again, analogous plots may be generated by varying the tunica externa compliance and thickness, the stiffness of ligament $L1$, or the transverse canal diameter.

B. Saccule

The total relative displacement $x_{rel}(\omega)$ may be interpreted as the sum of the relative displacements between the sagitta and sensory epithelium due to each of the three hearing mechanisms: the direct path, the indirect path, and the Weberian path. These individual components may be directly calculated and used to assess the relative contribution of each as a function of frequency, as shown in Fig. 12. The Weberian component clearly dominates throughout most of the frequency range. The direct path has a strong influence only at low frequencies, below 50 Hz. The indirect path component amplitude is larger than the direct path at frequencies near the swimbladder resonance, but still well below that of the Weberian component. The phase angles show the direct and Weberian components in phase at low frequencies, and the Weberian and indirect components 180° out of phase throughout the frequency range. It should be pointed out that these results apply only to the case of a monopole source 1 m in front of the fish. At larger source distances, the influence of the direct path diminishes, while the effects of the Weberian and indirect paths do not. This is because the direct path relies on the acoustic particle velocity as the input to the saccule; at larger distances from the source, the particle velocity decreases, which decreases the direct path component. Thus, for large source distances the Weberian path dominates throughout the entire frequency range. This con-

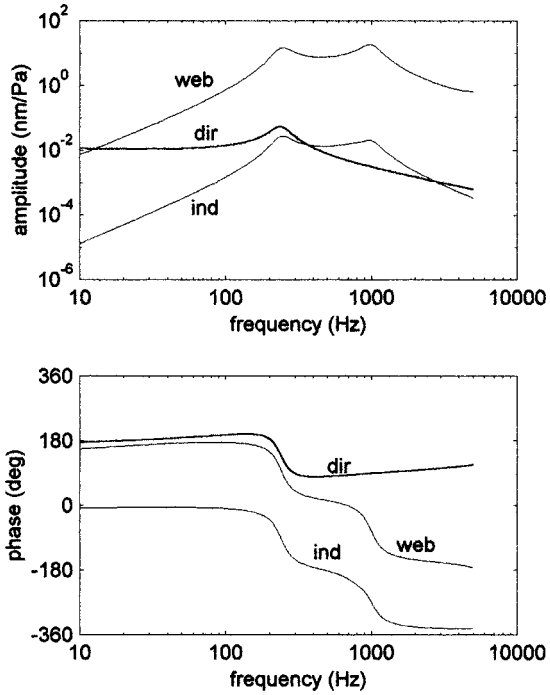


FIG. 12. Relative displacement between the sagitta and sensory epithelium for a monopole source 1 m in front of the fish. dir—direct path; ind—indirect path; web—Weberian path.

dition would also be present in any experimental procedures where the fish is excited using a loudspeaker in air. Although the pressure waves in the air may couple to the water in a small tank at low frequencies by exciting the boundaries of the tank, this would not produce particle velocities as large as those accompanying an acoustic wave underwater at low frequencies close to a source (Parvulescu, 1964, 1967). If the source is moved from a position in front of the fish to the sides or rear, the phase angles show different relationships. In particular, the direct component phase is shifted by 180° when the source is located behind the fish. With the source behind the fish, the direct and indirect path components are in phase at low frequencies and 180° out of phase with the Weberian path component. These phase relationships may contribute to directional hearing capability (Schuijf, 1976).

Above 50 Hz, where the Weberian apparatus component dominates, the total relative displacement between the sagitta and sensory epithelium may be inferred from the Weberian apparatus component of Fig. 12. From approximately 300 to 900 Hz, the total relative displacement is roughly in phase with the acoustic pressure. Using the sign conventions of Fig. 6, this translates to dorsal sagitta (and hair-cell ciliary bundle tip) motion in response to positive acoustic pressure. The phase lag between $x_{\mathrm{rel}}(\omega)$ and the acoustic pressure is approximately 0°–20° between 600 and 800 Hz. These results are in general agreement with those of Furukawa and Ishii (1967a), who measured microphonic potentials in goldfish and reported that the hair-cell ciliary bundles were deflected dorsally in response to positive acoustic pressure and ventrally in response to negative pressure. Furukawa and Ishii also reported a time delay of 0.1–0.2 ms (corresponding to a phase lag of 20°–60°) between the pressure stimulus and the onset of the microphonic potential.

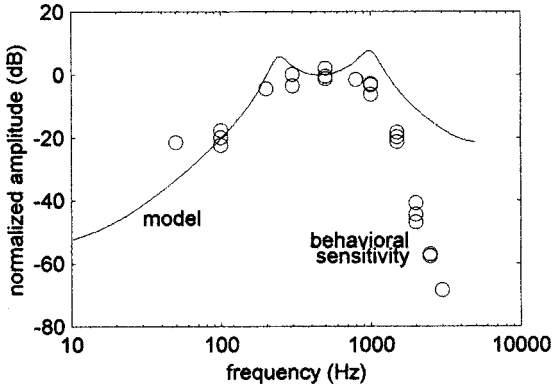


FIG. 13. Model predictions compared to goldfish behavioral sensitivity. Both data series are normalized to 0 dB at 500 Hz. Solid line: Total relative displacement between the sagitta and sensory epithelium for a monopole source 1 m in front of the fish (sum of the direct, indirect, and Weberian components from Fig. 12). Symbols: Goldfish auditory sensitivity, defined as the reciprocal of the auditory threshold. These data are pooled from Jacobs and Tavolga (1967) and Popper (1971b).

Based on Fig. 12, one would expect the removal of the swimbladder or the Weberian ossicles to result in a severe loss of hearing sensitivity, up to approximately 40 dB. This was observed by Poggendorf (1952), who found a loss of 25–40 dB in hearing sensitivity following removal of the tripus in the catfish. Fay and Popper (1974) also reported a decline of 20–35 dB in sound-pressure sensitivity after removal of the swimbladder in goldfish.

Figure 13 compares the total relative displacement, $x_{\mathrm{rel}}(\omega)$, which is the sum of the individual components shown in Fig. 12, to the goldfish auditory sensitivity, defined as the reciprocal of the auditory threshold obtained from behavioral conditioning experiments. The sensitivity data are pooled from Jacobs and Tavolga (1967) and Popper (1971b) and are, along with the model predictions, normalized to 0 dB at 500 Hz. The auditory sensitivity and the model response agree reasonably well, except at frequencies above 1–2 kHz. It is most likely that dynamics present in the real system but not appearing in the model are responsible for the steep slope of the sensitivity curve at high frequencies. At lower frequencies, the model and sensitivity curves match well and both show a relatively flat region from 200–1000 Hz. The behavioral data appear more heavily damped than the model predictions as well; however, this may be in some part caused by scatter in the behavioral data: trial-to-trial differences in the measured thresholds would tend to lower the amplitudes of any sharp peaks. Small perturbations in the swimbladder, Weberian apparatus, or saccule parameters would have the same effect in the model.

Equation (8) may be rewritten to express the relative displacement between the sagitta and the sensory epithelium as the sum

$$x_{\mathrm{rel}}(\omega)=H_{\mathrm{se}}(\omega)x_{\mathrm{se}}(\omega)+H_{\mathrm{sa}}(\omega)\frac{V_{\mathrm{sa}}(\omega)}{j\omega}, \tag{10}$$

where $H_{\mathrm{se}}(\omega)$ is the transfer function between the sensory epithelium displacement and the relative displacement, and $H_{\mathrm{sa}}(\omega)$ is the transfer function between the transverse canal fluid displacement and the relative displacement. These

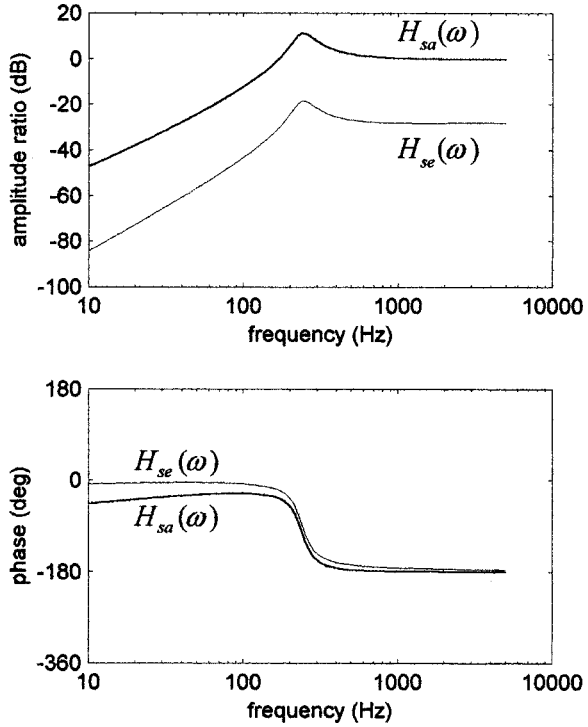


FIG. 14. Transfer functions relating the relative motion between sagitta and sensory epithelium to the sensory epithelium displacement (H_{se}) and the displacement of fluid entering the saccule from the transverse canal (H_{sa}).

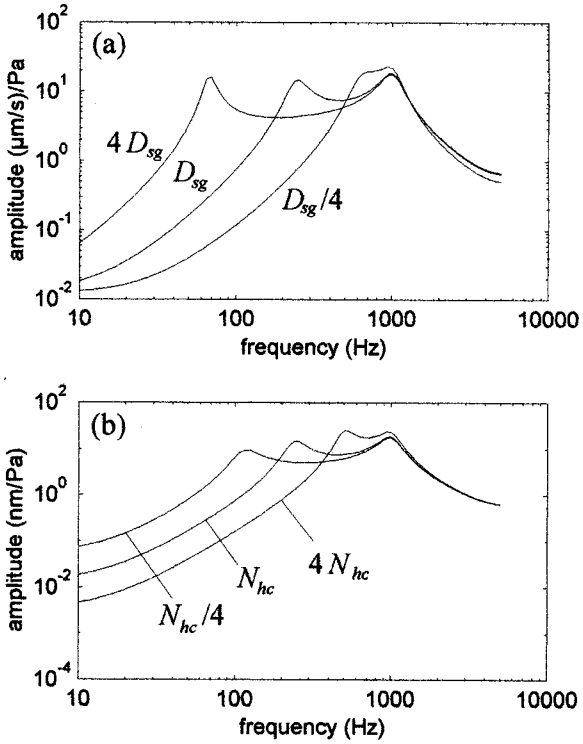


FIG. 15. Effects of varying saccule parameters. (a) Effect of changing sagitta diameter D_{sg}. (b) Effect of varying the number of hair cells N_{hc}.

transfer functions are displayed in Fig. 14 as functions of frequency. At low frequencies, the slope of $H_{sa}(\omega)$ is +20 dB/decade; the slope of $H_{se}(\omega)$ is +40 dB/decade. At higher frequencies, above 300 Hz, both transfer functions have a flat amplitude ratio, indicating that the saccule operates as a displacement sensor above 300 Hz. At low frequencies the saccule responds to the *velocity* of fluid entering the saccular chamber, and to the *acceleration* of the sensory epithelium. At low frequencies, the phase angle of $H_{sa}(\omega)$ does not approach 0°, as might be expected, but rather −90°. This is a consequence of Stokes' Law, which dictates that at sufficiently low frequencies the force on an object moving in a viscous fluid is proportional to the relative velocity between the fluid and the object (Stokes, 1851).

The relative velocity between the sagitta and the sensory epithelium has two distinct regions. At high frequencies, near the swimbladder resonance and above, the swimbladder and Weberian apparatus control the motion. At lower frequencies, below the swimbladder resonance, the response is primarily controlled by the saccule system parameters, including the number and stiffness of the ciliary bundles and the size and shape of the otolith. Changing these parameters greatly affects the low-frequency response of the system, but shows little or no effect at or above the swimbladder resonance.

Figure 15(a) shows the effect of varying the sagitta effective diameter on the relative velocity between the sagitta and sensory epithelium. Increasing the diameter shifts the curve to the left and broadens the range of frequencies where the relative displacement is reasonably flat and in phase with the acoustic pressure. Though not shown, the effects of changing the sagitta length are similar: increasing the length shifts the curves to the left; decreasing the length shifts the curves to the right. Figure 15(b) shows the effect of varying the number of hair cells, N_{hc}, on the relative displacement between the sagitta and sensory epithelium. Increasing N_{hc} shifts the curve to the right; decreasing N_{hc} shifts the curve left. The effects of changing the ciliary bundle stiffness k_{cb} are similar to those of N_{hc}.

IV. CONCLUSIONS

The auditory system analysis presented here predicts the correct amplitude and phase relationships between the swimbladders and shows the coupling observed between the anterior swimbladder and tripus. The resonance frequencies for the individual ossicles are well above the auditory range; thus, the Weberian apparatus acts essentially as a spring at audible frequencies and couples the swimbladder motion to the saccule. The results also show a high-pass filter effect caused by the tunica externa compliance; however, the low frequency roll-off of 5 dB/decade is insufficient to prevent a change in depth from transmitting a large amplitude input through the Weberian apparatus to the saccule. Other mechanisms may exist to prevent large, low-frequency changes in swimbladder volume from affecting the Weberian apparatus performance and thus protect the inner ear.

The results for the relative displacement between the sagitta and the sensory epithelium show the saccule response to be analogous to that of a displacement sensor above approximately 300 Hz. At low frequencies the response is controlled by the properties of the saccule itself; at higher frequencies the swimbladder controls the response. The performance predictions for the saccule generally agree with the data presented by Furukawa and Ishii (1967a) which

show that the relative motion between the sagitta and the sensory epithelium lags the acoustic pressure by 20°–60° between 600 and 800 Hz. The results predict a sagitta amplitude on the order of 10 nm/Pa.

If the relative displacement is broken down into the direct, indirect, and Weberian paths, the Weberian path component dominates above 50 Hz and is roughly 40 dB larger than the others at frequencies above 100 Hz. This agrees with data showing a severe loss of hearing sensitivity after removal of the swimbladder or tripus (Poggendorf, 1952; Fay and Popper, 1974). The phase relationships between the direct component and the others change as the source is moved from the front to the back of the fish. Phase relationships such as this may influence directional hearing (Schuijf, 1976).

Because of the uncertainties present in the system parameters, the results are most useful in predicting general trends, rather than exact responses. The swimbladder mass, stiffness, and damping, tunica externa compliance, and hair-cell ciliary bundle number and stiffness are among the most significant parameters.

Analysis of the peripheral auditory system suffers mainly from a lack of reliable material property data and the limited availability of morphological data. This is most obvious in the saccule, where the manner of coupling between the otolith, otolithic membrane, ciliary bundles, and the saccular chamber is largely unknown. The first step to improve the system analysis would be to use medical imaging techniques to obtain detailed information on the structure of the auditory system, and the sizes of and spatial relationships between the auditory organs. Direct experimental evaluation of the properties of the tunica externa and the Weberian ossicle ligaments may also improve the reliability of analytical predictions.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research Grant No. N00014-94-1-0337 and the Frederick V. Hunt Postdoctoral Research Fellowship in Acoustics.

Adams, L. A. (**1940**). ''Some characteristic otoliths of American Ostariophysi,'' J. Morphol. **66**, 497–527.

Alexander, R. McN. (**1959**). ''The physical properties of the isolated swimbladder in cyprinidae,'' J. Exp. Biol. **36**, 341–346.

Alexander, R. McN. (**1961**). ''Visco-elastic properties of the tunica externa of the swimbladder in cyprinidae,'' J. Exp. Biol. **38**, 747–757.

Alexander, R. McN. (**1962**). ''The structure of the Weberian apparatus in the Cyprini,'' Proc. Zool. Soc. London **139**, 451–473.

Alexander, R. McN. (**1966**). ''Physical aspects of swimbladder function,'' Biol. Rev. Cambridge Philos. Soc. **41**, 141–176.

Andreeva, I. B. (**1964**). ''Scattering of sound by air bladders of fish in deep sound-scattering ocean layers,'' Sov. Phys. Acoust. **10**, 17–20.

Batzler, W. E., and Pickwell, G. V. (**1970**). ''Resonant acoustic scattering from gas-bladder fishes,'' Proceeding International Symposium on Biological Sound Scattering in the Ocean, pp. 168–179 (unpublished).

Benser, M. E., Issa, N. P., and Hudspeth, A. J. (**1993**). ''Hair-bundle stiffness dominates the elastic reactance to otolithic-membrane shear,'' Hear. Res. **68**, 243–252.

Chardon, M., and Vandewalle, P. (**1991**). ''Acoustico-lateralis system,'' in *Cyprinid Fishes*, edited by I. J. Winfield and J. S. Nelson (Chapman and Hall, New York), pp. 332–352.

Chranilov, N. S. (**1927**). ''Beiträge zur kenntnis des Weber'schen apparatus der ostariophysi. I. Vergleichend-anatomische. Üebersich der knochenelemente das Weber'schen apparatus bei Cypriniformes,'' Zool. Jahrb. Abt. Anat. **49**, 501–597.

Chranilov, N. S. (**1929**). ''Beiträge zur kenntnis des Weber'schen apparatus der ostariophysi. II. der Weber'schen apparatus bei Siluridae,'' Zool. Jahrb. Abt. Anat. **51**, 323–462.

Corey, D. P., and Hudspeth, A. J. (**1983**). ''Kinetics of the receptor current in bullfrog saccular hair cells,'' J. Neurosci. **3**, 962–976.

Cox (a.k.a. Hastings), M., and Rogers, P. H. (**1987**). ''Automated noninvasive motion measurement of auditory organs in fish using ultrasound,'' J. Vibr., Acoustics, Stress, and Rel. in Des. **109**, 55–59.

de Munck, J. C., and Schellart, N. A. M. (**1987**). ''A model for the nearfield acoustics of the fish swimbladder and its relevance for directional hearing,'' J. Acoust. Soc. Am. **81**, 556–560.

deBurlet, H. M. (**1929**). ''Zur vergleichende anatomie und physiologie des perilymphatishen raumes,'' Acta Oto-Laryngol. **13**, 153–187.

Derenburger, C. (**1997**). ''Experimental study of the low-frequency response of the peripheral auditory organs in the goldfish and oscar,'' M. S. thesis, The Ohio State University, Columbus, Ohio.

Derenburger, C. J., Hastings, M. C., and Finneran, J. J. (**1997**). ''A comparison of experimentally determined swimbladder resonances in two species of fish,'' Proceedings ASME Noise Control and Acoustics Division, **NCA-24**, edited by T. M. Farabee (The American Society of Mechanical Engineers, New York), pp. 299–306.

deVries, H. L. (**1950**). ''The mechanics of the labyrinth otoliths,'' Acta Oto-Laryngol. **38**, 262–273.

deVries, H. L. (**1956**). ''Physical aspects of the sense organs,'' Progr. Biophys. Biophysical Chem. **6**, 207–264.

Dijkgraaf, S. (**1960**). ''Hearing in bony fishes,'' Proc. R. Soc. London, Ser. B **152**, 51–54.

Evans, H. M. (**1925**). ''A contribution to the anatomy and physiology of the air-bladder and Weberian ossicles in Cyprinidae,'' Proc. R. Soc. London, Ser. B **97**, 545–576.

Fay, R. R., and Olsho, L. W. (**1979**). ''Discharge patterns of lagenar and saccular neurons of the goldfish eighth nerve: Displacement sensitivity and directional characteristics,'' Comp. Biochem. Physiol. **62A**, 377–386.

Fay, R. R., and Popper, A. N. (**1974**). ''Acoustic stimulation of the ear of the goldfish (*Carassius auratus*),'' J. Exp. Biol. **61**, 243–260.

Feuillade, C., and Nero, R. W. (**1998**). ''A viscous-elastic swimbladder model for describing enhanced-frequency resonance scattering from fish,'' J. Acoust. Soc. Am. **103**0, 3245–3255.

Finneran, J. J., and Hastings, M. C. (**1996**). ''A noninvasive ultrasonic technique for measuring the dynamic properties of biological tissues,'' J. Acoust. Soc. Am. **100**, 2617(A).

Freeman, D. M., and Weiss, T. F. (**1985**). ''On the role of fluid inertia and viscosity in stereociliary tuft motion: Analysis of isolated bodies of regular geometry,'' in *Lecture Notes in Biomathematics*, edited by S. Levin (Springer, New York), pp. 147–154.

Furukawa, T., and Ishii, Y. (**1967a**). ''Effects of static bending of sensory hairs on sound reception in the goldfish,'' Jpn. J. Physiol. **17**, 572–588.

Furukawa, T., and Ishii, Y. (**1967b**). ''Neurophysiological studies on hearing in goldfish,'' J. Neurophysiol. **30**, 1377–1403.

Howard, J., and Hudspeth, A. J. (**1987**). ''Mechanical relaxation of the hair bundle mediates adaptation in mechanoelectrical transduction by the bullfrog's saccular hair cell,'' Proc. Natl. Acad. Sci. USA **84**, 3064–3068.

Jacobs, D. W., and Tavolga, W. N. (**1967**). ''Acoustic intensity limens in the goldfish,'' Anim. Behav. **15**, 324–335.

Jenkins, D. B. (**1977**). ''Light microscopic study of the saccule and lagena certain catfishes,'' Am. J. Anat. **150**, 605–630.

Kachar, B., Parakkal, M., and Fex, J. (**1990**). ''Structural basis for mechanical transduction in the frog vestibular sensory apparatus: I. The otolithic membrane,'' Hear. Res. **45**, 179–190.

Krumholz, L. A. (**1943**). ''A comparative study of the Weberian ossicles in North American Ostariophysine fishes,'' Copeia **1943(1)**, 33–40.

Lebedeva, L. P. (**1965**). ''Measurement of the dynamic complex shear modulus of animal tissues,'' Sov. Phys. Acoust. **11**, 163–165.

Levy, E. C. (**1959**). ''Complex curve fitting,'' IRE Trans. Auto. Cont. **AC-4**, 37–44.

Lewis, T. (**1994**). ''Detection of scattered ambient noise by fish,'' Ph.D. thesis, Georgia Institute of Technology, Atlanta, Georgia.

Lombarte, A., and Popper, A. N. (**1994**). ''Quantitative analyses of postembryonic hair cell addition in the otolithic endorgans of the inner ear of the European hake, *Merluccius merluccius*,'' J. Comp. Neurol. **345**, 419–428.

Love, R. H. (**1978**). ''Resonant acoustic scattering by swimbladder-bearing fish,'' J. Acoust. Soc. Am. **64**, 571–580.
Løvik, A., and Hovem, J. M. (**1979**). ''An experimental investigation of swimbladder resonance in fishes,'' J. Acoust. Soc. Am. **66**, 850–854.
McCartney, B. S., and Stubbs, A. R. (**1971**). ''Measurements of the acoustic target strengths of fish in dorsal aspect, including swimbladder resonance,'' J. Sound Vib. **15**, 397–420.
Money, K. E., Bonen, L., Beatty, J. D., Kuehn, L. A., Sokoloff, M., and Weaver, R. S. (**1971**). ''Physical properties of fluids and structures of the vestibular apparatus of the pigeon,'' Am. J. Phys. **220**, 140–147.
Parvulescu, A. (**1964**). ''The acoustics of small tanks,'' in *Marine Bioacoustics*, edited by W. N. Tavolga (Pergamon, Oxford), Vol. I, pp. 7–13.
Parvulescu, A. (**1967**). ''Problems of propagation and processing,'' in *Marine Bioacoustics*, edited by W. N. Tavolga (Pergamon, Oxford), Vol. II, pp. 87–100.
Platt, C. (**1977**). ''Hair cell distribution and orientation in goldfish otolith organs,'' J. Comp. Neurol. **172**, 283–298.
Poggendorf, D. (**1952**). ''Die absoluten hörschwellen des zwergwelses (*Amiurus nebulosus*) und bieträge zur physik des Weberschen apparates der Ostariophysen,'' Z. Vergl. Physiol. **34**, 222–257.
Popper, A. N. (**1971a**). ''The morphology of the Weberian ossicles of two species of the genus *Astyanax* (Ostariophysi: Characidae),'' J. Morphol. **133**, 179–188.
Popper, A. N. (**1971b**). ''The effect of size on the auditory capacities of the goldfish,'' J. Aud. Res. **11**, 239–247.
Popper, A. N., and Fay, R. R. (**1993**). ''Sound detection and processing by fish: Critical review and major research questions,'' Brain Behav. Evol. **41**, 14–38.
Popper, A. N., and Hoxter, B. (**1984**). ''Growth of a fish ear. I. Quantitative analysis of sensory hair cell and ganglion cell proliferation,'' Hear. Res. **15**, 133–142.
Prandtl, L. (**1952**). *Essentials of Fluid Dynamics* (Hafner, New York).
Rogers, P. H., and Hastings, M. C. (**1989**). ''Noninvasive vibration measurement system and method for measuring amplitude of vibration of tissue in an object being investigated,'' U.S. Patent No. 4,819,643.
Sanathanan, C. K., and Koerner, J. (**1963**). ''Transfer function synthesis as a ratio of two complex polynomials,'' IEEE Trans. Autom. Control **AC-3**, 56–58.
Sand, O., and Hawkins, A. D. (**1973**). ''Acoustic properties of the cod swimbladder,'' J. Exp. Biol. **58**, 797–820.
Schuijf, A. (**1976**). ''The phase model of directional hearing in fish,'' in *Sound Reception in Fish*, edited by A. Schuijf and A. D. Hawkins (Elsevier, Amsterdam), pp. 63–86.
Stokes, G. G. (**1851**). ''On the effect of the internal friction of fluids on the motion of pendulums,'' Trans. Cambridge Philos. Soc. **9**, 6–106.
Strasberg, M. (**1953**). ''The pulsation frequency of nonspherical gas bubbles in liquids,'' J. Acoust. Soc. Am. **25**, 536–537.
ten Kate, J. H., and Kuiper, J. W. (**1970**). ''The viscosity of the pike's endolymph,'' J. Exp. Biol. **53**, 495–500.
van Bergeijk, W. A. (**1964**). ''Directional and nondirectional hearing in fish,'' in *Marine Bioacoustics*, edited by W. N. Tavolga (Pergamon, Oxford), pp. 281–303.
von Boutteville, K. F. (**1935**). ''Untersuchungen über den gehörsinn bei characiniden und gymnotiden und den bau ihres labyrinthes,'' Z. Vergl. Physiol. **22**, 162–191.
von Frisch, K. (**1938**). ''Über die bedeutung des sacculus und der lagena für den gehörsinn der fische,'' Z. Vergl. Physiol. **25**, 703–747.
Watson, J. M. (**1939**). ''The development of the Weberian ossicles and anterior vertebrae in the goldfish,'' Proc. R. Soc. London, Ser. B **127**, 452–472.
Weber, E. H. (**1820**). *De aure et auditu hominis et animalium. Pars I. De aure animalium aquatilium* (Gerhard Fleischer, Leipzig).
Weston, D. E. (**1967**). ''Sound propagation in the presence of bladder fish,'' in *Underwater Acoustics*, edited by V. M. Albers (Plenum, New York), Vol. 2, pp. 55–88.
Ye, Z., and Farmer, D. M. (**1994**). ''Acoustic scattering from swim-bladder fish at low frequencies,'' J. Acoust. Soc. Am. **96**, 951–956.
Zhou, W. (**1992**). ''Two-peak resonance of swimbladders in fish (acoustic measurements and analytical model),'' M.S. thesis, Georgia Institute of Technology.

Masking in three pinnipeds: Underwater, low-frequency critical ratios

Brandon L. Southall, Ronald J. Schusterman, and David Kastak
Long Marine Laboratory, University of California at Santa Cruz, 100 Shaffer Road, Santa Cruz, California 95060

(Received 22 December 1999; accepted for publication 8 June 2000)

Behavioral techniques were used to determine underwater masked hearing thresholds for a northern elephant seal (*Mirounga angustirostris*), a harbor seal (*Phoca vitulina*), and a California sea lion (*Zalophus californianus*). Octave-band white noise maskers were centered at five test frequencies ranging from 200 to 2500 Hz; a slightly wider noise band was used for testing at 100 Hz. Critical ratios were calculated at one masking noise level for each test frequency. Above 200 Hz, critical ratios increased with frequency. This pattern is similar to that observed in most animals tested, and indicates that these pinnipeds lack specializations for detecting low-frequency tonal sounds in noise. However, the individual pinnipeds in this study, particularly the northern elephant seal, detected signals at relatively low signal-to-noise ratios. These results provide a means of estimating zones of auditory masking for pinnipeds exposed to anthropogenic noise sources. © *2000 Acoustical Society of America.* [S0001-4966(00)02509-1]

PACS numbers: 43.80.Lb [WA]

I. INTRODUCTION

Many pinnipeds (seals, sea lions, and walruses) use low-frequency, broadband acoustic signals under water for intraspecific communication (Schusterman, 1978; Watkins and Wartzok, 1985). In addition to detection of conspecific signals, passive listening to other natural sounds may also be biologically significant for pinnipeds in localization of predators or prey, or in spatial orientation and navigation (see Schusterman *et al.*, 2000). Since they spend considerable time in coastal environments where noise levels from natural sources such as wind, waves, and biologics are frequently high (Urick, 1983), hearing in the presence of interfering noise is very important for pinnipeds. Additional sources of noise, however, have resulted from the industrialization of marine environments, greatly increasing overall noise levels, particularly at low frequencies, in many areas (see Ross, 1976). The inundation of new noise sources, such as large vessels, petroleum exploration and recovery efforts, and low-frequency military sonar systems, is likely to adversely impact some pinniped species either by displacing individuals from important foraging or reproductive areas or by interfering with hearing (Richardson *et al.*, 1995; Kastak and Schusterman, 1998; Kastak *et al.*, 1999).

One way noise can compromise hearing is by masking a signal. The magnitude of interference depends upon signal and noise frequency and is greatest when they are similar. The lowest signal-to-noise ratio at which a subject can detect a tonal signal over broadband masking noise is called the critical ratio (Fletcher, 1940). Critical ratios (CRs) are calculated as the ratio of the sound pressure level of a just-audible tonal signal (dB *re*: 1 μPa) to that of the masking noise sound pressure level at the center frequency of the masking band, calculated from the measured masker spectral density level (dB *re*: 1 μPa^2/Hz). While the CR method of investigating auditory masking appears to be of little use in predicting critical bandwidth and auditory filter shape, it does provide simple measures of signal-to-noise ratios required for detection (Scharf, 1970). In nearly all mammals, CRs increase gradually with frequency, except at very low frequencies (see Fay, 1988). In contrast, some mammals have markedly lower CRs at specific frequencies. For instance, the horseshoe bat (*Rhinolophus ferrumequinum*) has dramatically lower CRs near 83 kHz, the constant frequency component of its echolocation signal (Long, 1977). Obtaining CRs within a range of frequencies prevalent in species-typical vocalizations is one of several methods used to determine if specific frequencies are processed particularly efficiently. Anatomical analyses of some mammals with such processing specializations have revealed relatively shallow stiffness gradients in certain regions of the basilar membrane (see Echteler *et al.*, 1994).

Despite the fact that the majority of pinniped vocalizations, as well as the natural background noise over which they must be detected, contain considerable energy at low frequencies, most studies of pinniped masked hearing have been conducted at or above 2000 Hz. Individuals of four species have been tested: the harp seal, *Phoca groenlandica* [Terhune and Ronald, 1971 ($n=1$)], ringed seal, *Phoca hispida* [Terhune and Ronald, 1975 ($n=2$)], harbor seal, *Phoca vitulina* [Renouf, 1980 ($n=2$); Turnbull and Terhune, 1990 ($n=1$); Terhune, 1991 ($n=1$)], and northern fur seal, *Callorhinus ursinus* [Moore and Schusterman, 1987 ($n=2$)]. Of these, only Terhune (1991) investigated masking below 2000 Hz. In general, pinniped CRs appear similar in form across frequencies to those of unspecialized mammals in that they increase gradually with frequency. Northern fur seal, harbor seal, and harp seal CRs are similar in magnitude to those of humans and some odontocete cetaceans, but are somewhat lower than those of the ringed seal (Terhune and Ronald, 1975) and some terrestrial mammals (see Fay, 1988).

The general lack of knowledge about low-frequency pinniped hearing and the effects of noise has prompted research

in response to concerns regarding anthropogenic noise pollution in marine environments. There has been recent emphasis on obtaining low-frequency absolute hearing thresholds (e.g., Kastak and Schusterman, 1998). However, the relevance of unmasked thresholds to problems involving animals in environments dominated by low-frequency noise is questionable. In such contexts, knowledge of masked hearing capabilities is needed to understand noise impacts. While absolute hearing thresholds are basic in predicting when masking will occur, masked hearing thresholds are necessary for determining the degree of masking. Thus both measures are required to adequately assess signal detection in noise.

This study investigated auditory masking at six frequencies between 100 and 2500 Hz in a northern elephant seal (*Mirounga angustirostris*), a California sea lion (*Zalophus californianus*), and a harbor seal (*Phoca vitulina*). Combined with the absolute thresholds measured using similar techniques for the same individuals by Kastak and Schusterman (1998), the masking data of this study will provide some basis for evaluating low-frequency anthropogenic noise impacts on free-ranging pinnipeds.

II. METHODS

Subjects

The subjects were a 4-year-old female northern elephant seal (Burnyce), a 9-year-old male harbor seal (Sprouts), and a 12-year-old female California sea lion (Rio). They were housed at Long Marine Laboratory in Santa Cruz, California in free-flow saltwater pools and adjacent haul-out space. Subjects received 20%–50% of their daily allotment of fish during experimental sessions.

Each of the subjects had extensive experience performing behavioral audiometric tasks; no deficits or abnormalities were apparent in their absolute, low-frequency audiograms (see Kastak and Schusterman, 1998).

Apparatus

The testing enclosure was a circular 7.5-m diameter, 2.5-m deep, concrete pool. A testing apparatus was placed in a fixed position on the pool edge for each experimental session. The apparatus comprised a 1-in PVC frame on which a chin station, a movable opaque door, and a square, plastic response paddle were mounted. The chin station was located 1.4 m from the surface of the water, 1.2 m from the side of the pool, and 0.4 m from the response paddle.

Test stimuli and masking noise

Test stimuli were 500-ms pure tones (40-ms rise/fall time) generated with Stanford Research Systems (SRS) Arbitrary Waveform Composer software and triggered from a SRS DS345 function generator. Signals were fed to a Hewlett-Packard 350C stepwise attenuator and then to a Realistic MPA-20 power amplifier. The output of this amplifier was connected to either a J-9 or a J-11 Naval Undersea Warfare Center (NUWC) underwater transducer. The transducer was suspended into the test pool by a PVC harness and could be moved along a steel support. It was placed in the same position, approximately 5 m from the testing apparatus, for all sessions at each test frequency.

The masking stimuli consisted of octave-band white noise centered at each test frequency, although a slightly wider band was used at 100 Hz. Noise pressure spectral density levels (dB *re*: 1 μPa^2/Hz) were 20 dB above directly measured or interpolated absolute hearing thresholds (i.e., 20 dB sensation level) determined by Kastak and Schusterman (1998). Since CRs obtained using bands of noise have been shown to be independent of masker level for most of the dynamic range in other mammals (see Fay, 1988), masked thresholds were obtained at a single masker level.

Gaussian white noise was generated with CoolEdit® software (Syntrillium) and identical 6-s intervals were recorded onto audio tapes. For each trial, one interval of masking noise was played back on a Technics RS-686 tape recorder and filtered with a Krohn-Hite 3550 bandpass filter. The output of this filter was connected to the input of the amplifier that drove the transducer. Thus signals and masking noise were mixed before being amplified and projected from the same source.

Signal and masking noise levels were calibrated at the chin station of the test apparatus before and after each experimental session. A NUWC H-56 hydrophone was mounted on the test apparatus so that its acoustic center was in approximately the same position as the center of the animal's head during testing. The hydrophone output was monitored on either a Hitachi V202 or a Kikusui cos 5041 oscilloscope, as well as on a real-time PC-based spectrum analyzer (SpectraPlus®, Pioneer Hill). Using the spectrum analyzer, narrow-band analysis (2.7-Hz analysis bandwidth) was performed and the noise pressure spectral density level was measured at the masking noise band center frequency. However, in order to more accurately represent the average noise levels in a slightly wider band surrounding each test frequency, masking noise spectral density levels were calculated from measured 1/3-octave levels. Calculated levels from the 1/3-octave measurement were generally within 1 dB of directly measured values at the center frequency of the noise band.

Careful calibration procedures and strict criteria on variability in signal and noise fields were employed in this study (see Moore and Schusterman, 1987). The acoustic response of the testing enclosure to low-frequency pure tones had previously been mapped (Kastak and Schusterman, 1998). Additional mapping was conducted to measure: (1) the distribution of noise energy within masker bands, (2) variation in received signal and noise levels at the chin station and many other positions around it, and (3) how these levels fluctuated over time. Large variations in received masker levels within noise bands (up to $\pm$20 dB) were obtained when white noise was projected. To compensate for this, masking noise was digitally filtered prior to recording intervals onto analog tape. This adjustment resulted in received noise levels that were relatively flat (within $\pm$3 dB) across the entire band. Masking noise was calibrated before and after each experimental session to ensure that this criterion was achieved when noise was projected into the testing enclosure. Large fluctuations in

received signal and noise levels (up to ±10 dB) were also initially measured within a few cm from the chin station. Transducer positions were selected for each test frequency that minimized signal and noise variability at and around the chin station. To ensure maximum uniformity in signal and noise fields, signal and noise levels were calibrated at 26 positions within a 20×20×20 cm region centered at the chin station and encompassing all reasonable positions of the subject's head during testing. Values at each position were maintained within ±3 dB of central calibration position measurements. Additionally, temporal variation in received levels was measured; transducer positions were selected so that these fluctuations were within ±3 dB at all calibration positions.

Procedure

Before each trial, a trainer signaled the subject to swim to the test apparatus and place its muzzle in the chin station. A trial began when the apparatus door was raised. Noise was presented for the duration of each 6-s trial and was terminated with the closing of the apparatus door; no noise was presented during inter-trial intervals. Noise alone was presented on 50% of trials while a signal and noise were presented on the remaining trials. A go/no-go procedure was used in which the sequence of noise and signal plus noise trials was pseudorandom (Moore and Schusterman, 1987). A subject indicated that it detected a signal by pressing the response paddle with its nose (go response) and indicated that it did not detect a signal by remaining stationed (no-go response). Correct responses were rewarded with fish; incorrect responses were not reinforced.

Each experimental session consisted of an 8–10 trial warm-up phase in which signal levels were constant and clearly audible over masking noise. This was followed by a threshold phase of 40–60 trials in which signal levels varied. Finally, a cool-down phase of six to eight trials was run in which signal levels were again constant and clearly audible. Warm-up and cool-down phases were used to evaluate subject motivation and stimulus control over responding.

Two different psychophysical procedures were used in the threshold phase. Initially a staircase method (Cornsweet, 1962) was used in which signal levels were attenuated 4 dB following each signal detection until the subject failed to detect a signal (miss). Adjustments were then made in 2-dB steps (increased following a miss, decreased following signal detection). Between 2 and 10 staircase sessions were conducted at each test frequency to provide a preliminary estimate of masked hearing thresholds. Subsequently, a method of constant stimuli procedure (Stebbins, 1970) was used in which five to seven discrete signal levels, separated by 4 dB and bracketing the estimated masked hearing threshold, were randomly presented. Masked hearing thresholds were determined to be the signal level corresponding to 50% correct detection using probit analysis (Finney, 1971). Between three and six method of constant stimuli sessions were conducted until 95% confidence limits of calculated thresholds fell within ±3 dB. False alarm rates, defined as the percentage of noise trials on which go responses occurred, were measured in method of constant stimuli sessions to obtain estimates of response bias.

TABLE I. Underwater masked hearing thresholds and critical ratios obtained from test sessions using the psychophysical method of constant stimuli. For each pinniped species listed, a single individual was tested.

Species	Frequency (Hz)	Masker Level (dB *re*: 1 μPa^2/Hz)	Masked threshold (dB *re*: 1 μPa)	Critical ratio (dB)
N. elephant seal	100	110	124	14
Harbor seal	100	115	131	16
N. elephant seal	200	93	103	10
Harbor seal	200	104	117	13
Calif. sea lion	200	112	130	18
N. elephant seal	500	98	111	13
Harbor seal	500	98	113	15
Calif. sea lion	500	107	127	20
N. elephant seal	800	95	109	14
Harbor seal	800	100	115	15
Calif. sea lion	800	94	122	18
N. elephant seal	1200	96	111	15
Harbor seal	1200	94	114	20
Calif. sea lion	1200	97	119	22
N. elephant seal	2500	95	112	17
Harbor seal	2500	97	114	17
Calif. sea lion	2500	86	108	22

Test frequencies were selected based on the ability to generate adequate test signals and masking noise that did not exceed our *a priori* maximum allowable variance in the area surrounding the chin station. Underwater masked hearing thresholds were obtained for all subjects at 200, 500, 800, 1200, and 2500 Hz. The elephant seal and harbor seal were also tested at 100 Hz where a relatively wider band of noise (50–200 Hz) was used to ensure coverage of the entire critical bandwidth. It was not possible to test the sea lion at 100 Hz because of this subject's higher absolute threshold at this frequency and limitations of the equipment. Critical ratios were calculated as the ratio (in dB) between the masked hearing threshold and the masking noise sound-pressure level at the center frequency of the masking band, calculated from the measured masker spectral density level.

III. RESULTS

Standard deviations for masked hearing thresholds were <1 dB and false alarm rates were <20% for all subjects at all frequencies. Table I shows the principal results of our study which include masked hearing thresholds and CRs for each subject representing three species of pinniped. The CRs of the pinnipeds in this study are graphically depicted in Fig. 1 so they may be compared with those of selected marine and terrestrial mammals. As Table I and Fig. 1 show, all subjects of this study displayed the general mammalian trend of increasing CRs with frequency, except at very low frequencies. Calculated CRs were slightly lower for the harbor seal than the California sea lion and were lower still in the elephant seal, whose values are similar to the lowest reported mammalian values (see Fay, 1988). These inter-individual differences were small, but consistent at all test frequencies.

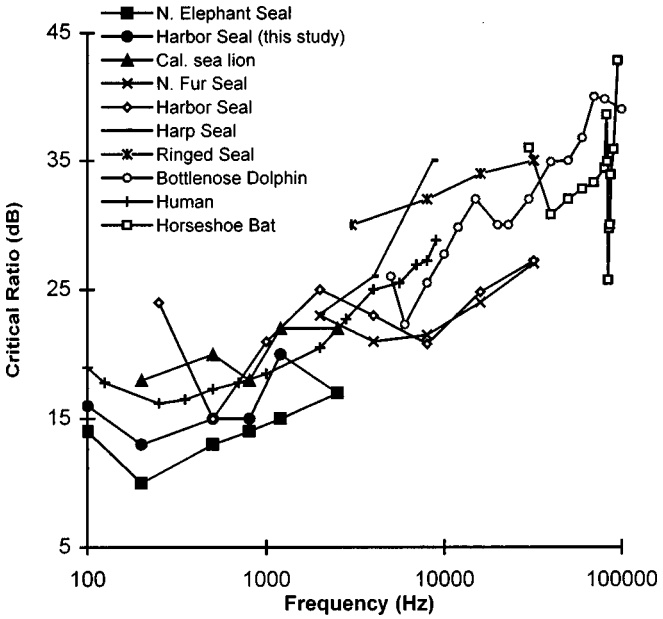


FIG. 1. Critical ratios for a northern elephant seal, a harbor seal, and a California sea lion shown with data from: northern fur seal (Moore and Schusterman, 1987), harbor seal (combined data from Turnbull and Terhune, 1990; Terhune, 1991), harp seal (Terhune and Ronald, 1971), ringed seal (Terhune and Ronald, 1975) bottlenose dolphin (Johnson, 1968), human (Hawkins and Stevens, 1950), and horseshoe bat (Long, 1977).

IV. DISCUSSION

Across a range of low frequencies, the CRs of the pinnipeds in this study are similar in form to those of most mammals tested. The apparent lack of specialization for enhanced detection of specific tonal signals is consistent with the fact that most pinniped vocalizations and other biologically significant sounds, such as incidental noise associated with schooling prey, are broadband and/or frequency modulated. This has probably limited the selective advantage of specialization for enhanced detection of specific frequency sounds and favored similar detection efficiency across frequencies.

The pinniped CRs in this study are somewhat lower in average magnitude than most other animals tested at similar frequencies (see Fay, 1988). Generally similar results have been obtained in other marine mammals with the exception of the ringed seal (see Richardson *et al.*, 1995). Low CRs in marine mammals might be a signal processing adaptation, facilitating detection in naturally noisy marine environments. Signal production strategies also thought to serve this purpose, such as signal redundancy, have been noted in pinnipeds (see Schusterman, 1978). With the caveat that auditory masking studies have been conducted with very few species and individuals, most of the available data indicate pinnipeds hear signals relatively well in noise. However, more masking studies, as well as additional research on pinniped cochlear mechanics and auditory neurophysiology with respect to frequency resolution, are needed to determine if this is in fact characteristic of pinnipeds. Additionally, more research on masking for individuals both in air and under water is needed to corroborate Turnbull and Terhune's (1990) conclusion that CRs should not differ between the two media. Future studies should contain adequate controls of spatiotemporal variability in signal and noise fields.

Unfortunately, the numerous methodological differences among studies of auditory masking in pinnipeds and other mammals and the generally small sample sizes limit comparisons between species. One potentially important variable in CR measurements that has received little attention is the duty cycle of masking noise presentation. In some mammalian masking studies, including ours, noise exposure was limited to the duration of each trial. However, in other studies noise was presented continuously during the entire experimental session, while still others fail to specify if masking noise was gated or continuous. In human subjects, continuous monaural exposure to moderate level (40–70 dB SPL) bands of noise changed a subject's perceived loudness of dichotically presented tonal signals in the same ear relative to a comparison signal in the ''rested'' ear by 2.3 to 9.9 dB in approximately 1 min. (Carterette, 1956). This effect is known as loudness adaptation and the change in perceived loudness of a signal following noise exposure is likely related to central neural processes (Gelfand, 1981). We are unaware of research on the potential effects of loudness adaptation on CR measurements obtained when masking noise was presented continuously and binaurally in humans or other animals. If adaptation occurs but affects both the masker and signal similarly, there should be no difference in the CRs obtained. However, since this is unknown, it is possible that loudness adaptation has some confounding effect on CR measures when masking noise is not gated for each trial. In order to avoid this possibility, intermittent masker presentation was used in this study.

Recent data on low-frequency absolute hearing (Kastak and Schusterman, 1998), temporary threshold shift (Kastak *et al.*, 1999), and auditory masking (this study) in pinnipeds are particularly relevant in considering noise impacts on free-ranging pinnipeds. This is because these studies involve frequencies produced by most anthropogenic noise sources and contained in most biologically significant signals, particularly conspecific vocalizations. The results of this study provide measures for estimating zones of auditory masking caused by specific anthropogenic noise sources such as large ships, as in the following, very simplified example. The 200-Hz component of a male harbor seal underwater call should be detectable by another harbor seal at a distance of approximately 160 m based on the calculated CR of 12.9 dB and assuming the following: the 200-Hz component of the call has a spectrum level of 105 dB (*re*: 1 μPa @ 1 m), $10 \log R$ spreading occurs because both animals are assumed to be in shallow water, and sea state four conditions [the 200-Hz spectrum level of resulting ambient noise would be approximately 70 dB (*re*: 1 μPa)]. If a supertanker 10 km away and in much deeper water were generating noise in which the 200-Hz spectrum level was 160 dB (*re*: 1 μPa @ 1 m), the audible distance between the two seals would be reduced to approximately 8.1 m, assuming ship noise propagation loss based on $20 \log R$ spreading for the first 5 km and $10 \log R$ for the final 5 km. An assumption made in this example is that CRs at a particular frequency do not differ for different masker levels. In other masking studies using

wideband noise, measured masking effects generally increased directly with increasing effective masker level across most of the dynamic range (see Fay, 1988). However, studies using tonal maskers have shown that more remote masking occurs with higher masker levels, particularly at frequencies above the masker, in both human (Ehmer, 1959) and bottlenose dolphin subjects (Johnson, 1971). If this is true for pinnipeds as well, the above example must be further qualified. While the CR at a particular frequency will likely remain the same in different noise levels, the degree of remote masking may differ, changing the perception of other frequency components of the salient signal.

While the majority of available data on masking indicate low CRs in pinnipeds and other marine mammals, this trend is contradicted by results obtained in a single study with one ringed seal (Terhune and Ronald, 1975). However, the finding that some pinnipeds hear signals well in noise should not be construed as an indication that anthropogenic noise is in some way innocuous to pinnipeds. Indeed, such sensory processing capabilities would in fact underscore the importance of effective hearing in pinnipeds and reiterate concerns regarding anthropogenic noise. Further, the masking effects of loud noise sources may still be very great even if individuals of a species are assumed to have low CR values. Regulatory agencies estimating zones of auditory masking for wild pinnipeds should opt for conservative estimates based on the upper limits of the range of CR values until additional data are available.

ACKNOWLEDGMENTS

This work was supported by Grant No. N000149510936 from the Office of Naval Research to R.J.S. and funding provided by the Friends of Long Marine Laboratory, American Cetacean Society and the Myers Oceanographic Trust to B.L.S. Colleen Reichmuth was integral in conducting experimental sessions and reviewing earlier drafts of this manuscript. The authors acknowledge the useful comments of two anonymous reviewers. The assistance of student volunteers of the Pinniped Research in Cognition and Sensory Systems group at Long Marine Laboratory is also appreciated.

Carterrette, E. C. (**1956**). ''Loudness adaptation for bands of noise,'' J. Acoust. Soc. Am. **28**, 865–871.

Cornsweet, T. N. (**1962**). ''The staircase method in psychophysics,'' Am. J. Psychol. **75**, 485–491.

Echteler, S. M., Fay, R. R., and Popper, A. N. (**1994**). ''Structure of the mammalian cochlea,'' in *Comparative Hearing: Mammals*, edited by R. R. Fay and A. N. Popper (Springer-Verlag, New York), pp. 134–171.

Ehmer, R. H. (**1959**). ''Masking patterns of tones,'' J. Acoust. Soc. Am. **31**, 1115–1120.

Fay, R. R. (**1988**). *Hearing in Vertebrates: A Psychophysics Databook* (Hill-Fay, Winnetka).

Finney, D. J. (**1971**). *Probit Analysis*, 3rd ed. (Cambridge University Press, Cambridge, England).

Fletcher, H. (**1940**). ''Auditory patterns,'' Rev. Mod. Phys. **12**, 47–65.

Gelfand, S. A. (**1981**). *Hearing: An Introduction to Psychological and Physiological Acoustics* (Marcel Dekker, New York).

Hawkins, J. H., and Stevens, S. S. (**1950**). ''The masking of pure tones and of speech by white noise,'' J. Acoust. Soc. Am. **22**, 6–13.

Johnson, C. S. (**1968**). ''Masked tonal thresholds in the bottle-nosed porpoise,'' J. Acoust. Soc. Am. **44**, 965–967.

Johnson, C. S. (**1971**). ''Auditory masking of one pure tone by another in the bottlenosed porpoise,'' J. Acoust. Soc. Am. **49**, 1317–1318.

Kastak, D., and Schusterman, R. J. (**1998**). ''Low frequency amphibious hearing in pinnipeds: Methods, measurements, noise, and ecology,'' J. Acoust. Soc. Am. **103**, 2216–2228.

Kastak, D., Schusterman, R. J., Southall, B. L., and Reichmuth, C. J. (**1999**). ''Underwater temporary threshold shift induced by octave-band noise in three species of pinniped,'' J. Acoust. Soc. Am. **106**, 1142–1148.

Long, G. R. (**1977**). ''Masked auditory thresholds from the bat, *Rhinolophus ferrumequinum*,'' J. Comp. Physiol. **116**, 247–255.

Moore, P. W. B., and Schusterman, R. J. (**1987**). ''Audiometric assessment of northern fur seals (*Callorhinus ursinus*),'' Marine Mammal Sci. **3**, 31–53.

Renouf, D. (**1980**). ''Masked hearing thresholds of harbor seals (*Phoca vitulina*) in air,'' J. Aud Res. **20**, 263–269.

Richardson, J. W., Greene, Jr., C. R., Malme, C. I., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, San Diego).

Ross, D. (**1976**). *Mechanics of Underwater Noise* (Pergamon, New York).

Scharf, B. (**1970**). ''Critical bands,'' in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, San Diego), pp. 159–202.

Schusterman, R. J. (**1978**). ''Vocal communication in pinnipeds,'' in *Behavior of Captive Wild Animals*, edited by H. Markowitz and V. J. Stevens (Nelson-Hall, Chicago), pp. 247–308.

Schusterman, R. J., Kastak, D., Levenson, D. L., Reichmuth, C. R., and Southall, B. L. (**2000**). ''Why pinnipeds don't echolocate,'' J. Acoust. Soc. Am. **107**, 2256–2264.

Stebbins, W. C. (**1970**). ''Principles of animal psychophysics,'' in *Animal Psychophysics: The Design and Conduct of Sensory Experiments*, edited by W. C. Stebbins (Appleton-Century-Crofts, New York), pp. 1–19.

Terhune, J. M. (**1991**). ''Masked and unmasked pure tone thresholds of a harbor seal listening in air,'' Can. J. Zool. **69**, 2059–2066.

Terhune, J., and Ronald, K. (**1971**). ''The harp seal *Pagophilus groenlandicus* (Erxleben, 1777). X. The air audiogram,'' Can. J. Zool. **49**, 285–290.

Terhune, J., and Ronald, K. (**1975**). ''Masked hearing thresholds of ringed seals,'' J. Acoust. Soc. Am. **58**, 515–516.

Turnbull, S. D., and Terhune, J. (**1990**). ''White noise and pure tone masking of pure tone thresholds of a harbor seal listening in air and under water,'' Can. J. Zool. **68**, 2090–2097.

Urick, R. J. (**1983**). *Principles of Underwater Sound*, 3rd ed. (McGraw-Hill, New York).

Watkins, W. A., and Wartzok, D. (**1985**). ''Sensory biophysics of marine mammals,'' Marine Mammal Sci. **1**, 219–260.

A software model to estimate zones of impact on marine mammals around anthropogenic noise

Christine Erbe[a] and David M. Farmer
Institute of Ocean Sciences, 9860 West Saanich Road, Sidney, British Columbia V8L 4B2, Canada

(Received 27 July 1999; accepted for publication 16 June 2000)

Anthropogenic noise impacts marine mammals in a variety of ways. In order to estimate over which ranges this happens, we first need to understand the propagation of noise through the ocean away from the noise source, and, second, understand the relationship between received noise levels and impact thresholds. A software package combining both aspects is presented. (1) A sound propagation model based on ray theory was developed to calculate received noise levels as a function of range, depth, and frequency. (2) Current knowledge of noise impact thresholds for marine mammals was gathered and included in software routines predicting zones of impact on marine mammals around industrial underwater noise sources. As input parameters, this software package requires the source level and spectrum of the noise of interest; physical oceanography data about the local ocean environment such as bathymetry, bottom and surface loss data, and sound speed profiles; and bioacoustical information about the target species in the form of an audiogram, critical auditory bandwidths, spectra of typical animal vocalizations, reported sound levels of disturbance, and criteria for hearing damage. As output, the software produces data files and plots of the zones of audibility, masking, disturbance, and potential hearing damage around a noise source. © *2000 Acoustical Society of America.* [S0001-4966(00)04509-4]

PACS numbers: 43.80.Nd, 43.30.Cq [WA]

I. INTRODUCTION

The effects of anthropogenic noise on marine mammals are currently an active field of research in many countries. The problem is urgent, because the populations of several marine mammal species are classified as ''at risk.'' Noise is known to affect marine mammals in a variety of ways and under certain circumstances can be damaging.

A. Masking

Water conducts sound very well but is essentially opaque for light. Chemical transfer in the ocean is greatly limited by low levels of turbulence. Therefore, audition generally is the most developed sense in marine mammals compared to vision and chemoreception. Marine mammals use acoustics for a variety of purposes, which can differ from species to species. The masking of acoustic signals can have detrimental effects—not only for the individual, but ultimately for the entire population. Underwater noise has the potential for interfering with odontocete echolocation signals, impeding the animals' ability to navigate or find food. Noise can mask communication signals that might play a role in social cohesion, group activities, mating, warning, or individual identification. Noise can further interfere with environmental sounds animals might listen to. For example, it has been postulated that animals recognize the sound of surf, which guides them away from shallow water. Noise also affects the detection of sounds of predators and sounds of prey. Basic signal detection abilities of marine mammals in noise have been measured in a few projects.[1–10] Masking will be ''biologically significant'' if the animals' biological fitness is reduced (decreased rate of reproduction). The extent to which masking can affect biological fitness is not yet understood.

B. Behavioral disturbance

Noise has the potential of disrupting ''normal'' animal behavior. The literature on observed behavioral changes due to human presence, aircraft, ships, boats, offshore construction, geophysical exploration, sonars, ocean acoustic research, etc. is vast (Richardson *et al.*[11] for review). Reported animal reactions include a cessation of feeding, resting, socializing, and an onset of alertness or avoidance. Often, the distance or even the received noise level at which the change in behavior was observed, was reported. For many marine mammals, disturbance occurred at broadband received noise levels of about 120 dB *re* 1 μPa for continuous (in opposition to transient) noise.

If noise ''only'' causes temporary alterations in swimming path, heart rate, or breathing, the behavioral disturbance may not be biologically significant. If, on the other hand, noise scares animals away from critical habitat for an extended period or impedes foraging, mating, or nursing, the effect will most likely be of biological significance. In general, however, long-term effects on individuals and particularly on a population remain unknown.

C. Hearing impairment

Similar to masking, hearing loss affects an animal's ability to navigate, communicate, and detect predators and prey. Any type of noise at some level has the ability to damage hearing by causing decreased auditory sensitivity. The extent of ear damage determines whether the resulting threshold

[a]Electronic mail: erbec@dfo-mpo.gc.ca

shift is temporary (TTS: temporary threshold shift) or permanent (PTS). Threshold shift depends on a variety of factors such as the spectral characteristics of the noise (frequency and amplitude), the amount of energy per time for impulsive noise, the hearing sensitivity (audiogram) of the subject, the duration of noise exposure and the duty cycle, or recovery time in between exposures. There are two studies on TTS in marine mammals in the literature. A bottlenose dolphin was exposed to octave band noise between 5 and 10 kHz for 30–50 min.[12] The level was 96 dB above the normal center-frequency threshold at 7.5 kHz. Immediately afterward, a TTS of 12–18 dB at the center frequency was measured. A harbor seal, two California sea lions, and one northern elephant seal were exposed to octave band noise 60–75 dB above the normal center-frequency thresholds.[13] After 20 min, threshold shifts of, on average, 4.8 dB were measured at frequencies between 100 and 2000 Hz. Hearing recovered to normal within 24 h. There are no data on permanent hearing loss due to noise exposure in marine mammals. Repeated exposure to TTS indicating noise is thought to cause a PTS.

D. Nonhearing physiological effects

There are no concrete data on nonhearing physiological effects in marine mammals; only humans and some terrestrial animals have been studied. Effects such as concussion, embolism, resonance of hollow organs, hemorrhage, and rupture of organs and tissues have been reported. It is difficult to extrapolate form one noise source and one organism to another.

E. Prey effects

Marine mammals are indirectly affected by noise in the case that noise reduces the availability of prey. For example, noise-induced effects in fish include swim bladder resonance, blast injury of fish, eggs and larvae, a decrease in reproductivity, slower growth, and possibly habitat avoidance.

In order to assess the impact of a variety of anthropogenic noise on marine mammals and in order to design protective measures, an efficient and rational tool is required. We have developed a software package to estimate zones of impact around underwater noise sources affecting marine mammals. The package consists of two parts. Part 1 is a sound propagation model calculating received noise levels and spectra as a function of depth and range for a particular ocean environment. Part 2 is an impact assessment model estimating the range over which damaging effects such as acoustic masking, behavioral disturbance, and hearing damage might occur.

II. DEVELOPMENT OF THE SOFTWARE PACKAGE

A. Sound propagation modeling

The sound propagation software forms the first part of the software package developed for this project. The goal was to design a sound propagation model for broadband sounds such as industrial noise and animal vocalizations between a few Hz and about 20 kHz. Because of the high frequencies involved and for computational speed, we decided to use ray theory. As a starting point, we based our sound propagation model on Bowlin's RAY code.[14] The main modifications we made were a search for eigenrays, an inclusion of surface loss, and frequency-dependent absorption by ocean water. As input parameters, the ray model reads in a bathymetry file (depth versus range), a sound speed profile (which can change with range), and a source location (specified in range and depth). The ray code then traces rays in two dimensions over depth and range. Rays arriving at the same receiver location are added incoherently. With knowledge of the source level in dB *re* 1 μPa @1 m and the source spectrum, the results of the sound propagation model are matrices of received sound pressure levels and spectra as a function of range and depth. For data reduction, rather than using a long vector of spectrum density levels in 1 Hz bands, 12th octave band levels are used for sound propagation. This approach is also reasonable in relation to critical bandwidths in some marine mammals.[15] Table I lists the center frequencies of the 12th octave bands used. The width of each 12th octave band can be calculated as

TABLE I. Center frequencies (Hz) of adjacent 12th octave bands.

100	106	112	119	126	133
141	150	159	168	178	189
200	212	224	238	252	267
283	300	317	336	356	378
400	424	449	476	504	534
566	599	635	673	713	755
800	848	898	951	1008	1068
1131	1199	1270	1345	1425	1510
1600	1695	1796	1903	2016	2136
2263	2397	2540	2691	2851	3020
3200	3390	3592	3805	4032	4271
4525	4795	5080	5382	5702	6041
6400	6781	7184	7611	8063	8543
9051	9589	10 159	10 763	11 404	12 082
12 800	13 561	14 368	15 222	16 127	17 086
18 102	19 178	20 319			

$$\Delta f = (2^{1/24} - 2^{-1/24}) \cdot f_{\mathrm{ctr}}. \tag{1}$$

Certain aspects of our sound propagation model are described in more detail in the following paragraphs.

1. Attenuation in seawater

To include frequency-dependent absorption, we used the summarizing formula given by Jensen *et al.*,[16] Eq. (1.34), based on individual studies at differing frequencies:[17–19]

$$\alpha = 3.3\times 10^{-3} + \frac{0.11 f^2}{1+f^2} + \frac{44 f^2}{4100+f^2} + 3.0\times 10^{-4} f^2. \tag{2}$$

The absorption coefficient α has the unit dB/km.

2. Surface loss

The sound propagation code allows the inclusion of surface loss for two environments: an ice-covered arctic environment and a windy open ocean surface. We applied a two-layer fluid–solid model with parameters used by Laible and Rajan[20] to calculate reflection coefficients under ice. A com-

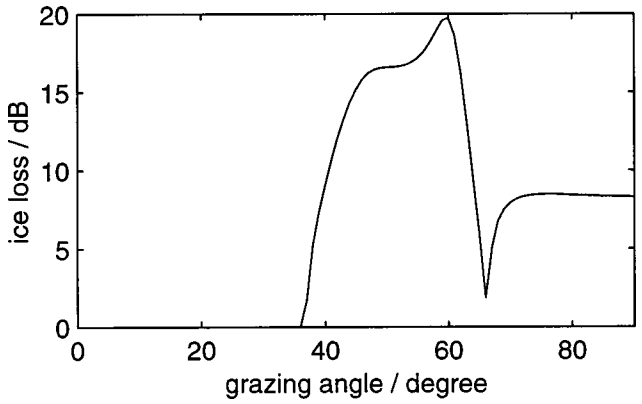


FIG. 1. Reflection loss versus grazing angle at the ice-covered surface.

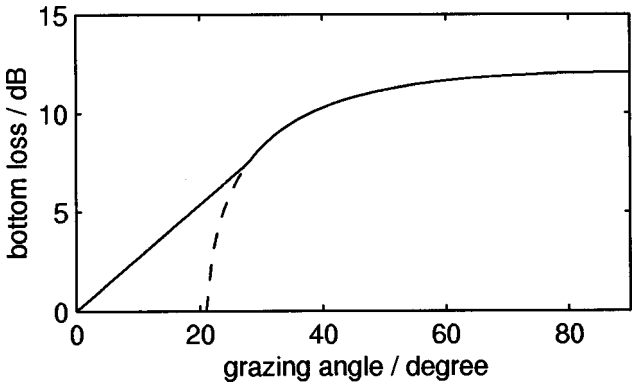


FIG. 2. Bottom loss versus grazing angle for a sand–silt–clay mixture.

pressional (P) plane wave in the fluid encountering a solid half-space gives rise to a ray reflected back into the fluid, a compressional (P) wave in the solid, and a shear (S) wave in the solid. With Z, Z_p, and Z_s denoting the acoustic impedances of the P wave in the fluid, the P wave in the solid and the S wave in the solid, respectively, and θ_s denoting the angle of the S wave toward the horizontal, the reflection coefficient R as the ratio of incident to reflected pressure is calculated according to Brekhovskikh,[21] Eq. (4.25):

$$R=\frac{Z_p\cos^2 2\theta_s+Z_s\sin^2 2\theta_s-Z}{Z_p\cos^2 2\theta_s+Z_s\sin^2 2\theta_s+Z}. \tag{3}$$

Figure 1 shows the inverse of the reflection loss in dB. The critical angle for the excitation of the S wave in the solid is 37°; the critical angle for the P wave in the solid is 66° (measured to the horizontal). For grazing angles less than either, total reflection without loss occurs, according to our model, without attenuation.

We further accounted for scattering loss from a randomly rough ice cover. In general, a ray incident on a rough surface gives rise to a number of scattered rays leaving at various angles. We treated scattering as a coherent loss of energy from the specularly reflected ray. According to Brekhovskikh and Lysanov,[22] Eq. (9.8.13), the mean reflection coefficient R_m from a randomly rough boundary is equal to the reflection coefficient from a planar boundary R times an exponential amplitude factor:

$$R_m=R\cdot e^{-P^2/2}. \tag{4}$$

The Rayleigh parameter P is defined as

$$P=2k\sigma\sin\theta=4\pi\frac{f}{c}\sigma\sin\theta. \tag{5}$$

Here, k is the acoustic wave number, f the corresponding frequency, c the speed of sound in water, and σ the root-mean-square roughness. The equation holds for wavelengths greater than σ. We chose a rms roughness of 1 cm for first-year ice.[20] We neglected low-frequency loss into elastic modes of the ice.

In the open ocean environment without ice cover, surface scattering is due to wind and waves causing local deviations from a planar surface and introducing scattering air bubbles into the upper ocean. Surface scattering due to a rough sea surface is included in our ray model similarly to the ice scattering algorithm with $R=1$ at the water–air interface (except for near-normal incidence) due to the large difference in acoustic impedance.

3. Bottom loss

We applied a simple two-fluid-layer model to include bottom loss. Using Clay and Medwin,[23] Eq. (2.9.5) for the reflection coefficient, and density, porosity, and sound speed data for sediments from Hamilton,[24] we computed the following relationship between bottom loss and grazing angle for the case of a sand–silt–clay bottom, Fig. 2. The critical angle was 21°. For incident angles less than critical, there was no bottom loss but total reflection according to the model (dashed lines). Absorption by the sediment was manually included by tapering the bottom loss for angles less than critical (solid line) according to Urick,[17] Chap. 11. We neglected scattering by the bottom that would have an effect for high frequencies. We further neglected sound propagation through sediment layers that would play a role for low frequencies and could add to the energy received in the water column at long ranges.

B. Calculation of the zone of audibility

The subroutine calculating the zone of audibility requires the following input data. First, it imports the received sound band levels of the manmade noise of interest as a function of range and depth as provided by the sound propagation algorithm. Second, it needs an audiogram (pure tone detection thresholds at a number of frequencies) of the marine mammal target species. Third, knowledge of the species' critical bandwidths (the widths of the auditory filters[25,26]) is required. Fourth, the software asks for a typical ambient noise spectrum for the location of interest. The output of this subroutine is a binary matrix whose columns represent range and whose rows represent depth. The matrix only contains the values 1 or 0, with 1 indicating that at least part of the noise spectrum is audible to the animal, and 0 indicating that the noise is entirely inaudible. The subroutine also creates a two-dimensional plot showing the extent of the zone of audibility with range and depth.

The zone of audibility of a sound source is limited by two factors: (1) by the critical band levels falling below the animal's audiogram, and (2) by the critical band levels fall-

ing below ambient noise levels. It is often assumed that a sound source stops being audible as soon as its critical band levels become equal or less than those of the ambient noise. Masking, however, is obviously more complex than that. Humans[27] and some species of marine mammals (Richardson *et al.*,[11] Chap. 8.5.3, for a review) have been shown to detect signals a few dB below ambient noise levels. Masking depends on the particular frequency and time structure of both signal and noise and on directional hearing abilities of the target animal. For simplicity, this subroutine limits the range of audibility by the critical band levels of the sound source dropping below ambient levels.

The subroutine carries out the following steps. Critical band levels are computed for both the manmade noise and the ambient noise. In each band, the maximum of the sound pressure level of the animal audiogram and the ambient noise band level is taken to yield a combined audibility-limiting spectrum. At each range and depth, this limiting spectrum is subtracted from the band levels of the received sound spectrum. If, in any critical band, the difference is greater than 0, the sound is considered audible, otherwise not.

C. Calculation of the zone of masking

The maximum range of masking is determined as the distance from a noise source at which the noise just prevents the detection of a faint signal, which would be just audible in the absence of the noise. The subroutine calculating the zone of masking needs the following input data: (1) the received noise spectra as a function of depth and range calculated by the sound propagation software, (2) a spectrum or spectra of typical animal vocalizations or other sounds that are being masked, and (3) the critical bandwidths in which masking will be determined.

Critical band levels for the signal to be masked are calculated. The band levels are taken such that they represent the quietest recognizable signal in the absence of masking noise. In most cases, this level will not be known. The user has to compare the signal level in critical bands to the audiogram and decide at what level the signal will just be recognizable. As default, the signal spectrum is taken such that the level in one critical band just surpasses the audiogram. Erbe[10] distinguished between signal detection and signal recognition. The signal was a beluga vocalization containing a base frequency and harmonic and nonharmonic higher components. In the absence of noise, the animal stopped recognizing the call as soon as the lower frequencies dropped below audibility. The higher frequencies, however, would have been audible to much lower sound pressure levels.

The algorithm for calculating masking in this subroutine is based on Fletcher's principle[28] stating that a signal will be masked by noise with equal or higher energy in the same critical band. This argument was, in fact, corroborated by Erbe,[10] measuring masked hearing thresholds for complex signals and noises with a beluga whale. Thus, comparing critical band levels, if all of the spectral components of the received noise lie above those of the signal, the signal will be completely masked. The software outputs a binary matrix of the zone of masking as a function of depth and range and plots the data.

Alternatively, instead of using Fletcher's principle, the software can be linked to automatic signal detection routines.[10,15] These were designed to model masked hearing thresholds in beluga whales as measured with a captive, trained beluga.[9] An artificial neural network trained with back-propagation achieved the best results. However, the neural network was designed to work with 2-D sound spectrograms containing both frequency and temporal information rather than sound spectra that only contain frequency information. In order to make full use of the neural network tool, the sound propagation part of this software package should be extended to include time. For those animal species where Fletcher's critical band theory seems to work, a critical band comparison might still be the preferable algorithm for the masking subroutine for reasons of computational speed.

D. Calculation of the zone of disturbance

The subroutine predicting the extent of the zone of disturbance requires knowledge of sound pressure levels that have reportedly led to observed behavioral changes in the target species. In the literature, this information might be available as broadband received sound pressure levels or as more narrowband (often third octave band) levels.[11] Often, in the case of continuous sounds, a broadband sound pressure level of 120 dB *re* 1 μPa is assumed. The software subroutine takes the received sound spectra provided by the sound propagation routine and calculates band levels as a function of depth and range for the frequencies at which behavioral reaction thresholds have been reported. If the received sound levels are greater than or equal to the reported threshold levels, disturbance is predicted to occur, otherwise not. The output of this subroutine is a binary matrix indicating over which range and depth disturbance exists, and a plot of the results.

E. Calculation of the zone of potential hearing damage

The subroutine estimating the range over which hearing damage might occur takes the received sound spectra provided by the sound propagation model as input. It further requires the animal audiogram. Octave band levels of the received sound spectra are calculated at a series of frequencies. If these are more than 96 dB above the audiogram at center frequency, a TTS of 12–18 dB can occur after 30 min, according to the study by Au *et al.*[12] If they are more than 60 dB above the audiogram, a TTS of 4.8 dB is possible, according to the study by Kastak *et al.*[13] The subroutine returns two binary matrices and plots the data.

III. CONCLUSION

A software package has been put together that estimates zones of impact on marine mammals around sources of anthropogenic noise. The first subroutine is a sound propagation model for broadband noise. It yields received noise band levels as a function of depth and range in two dimensions. Its output serves as input to the following subroutines estimating the depth and range over which the noise is audible to a

target species, over which it interferes with marine mammal communication, over which it disturbs the animals, and over which it could cause hearing damage. As new biological information becomes available, e.g., data on organ damage in marine mammals, the software package can readily be updated and further subroutines can be included. This software package can be applied to a variety of ocean environments, to any marine mammal species (given that required biological input data such as an audiogram are known), and to various types of noise. In the case of baleen whales, who are more sensitive at lower frequencies, the ray propagation model should be replaced by a model more appropriate at low frequencies, e.g., a parabolic equation model. Anthropogenic noise sources in marine mammal habitat include offshore hydrocarbon exploration (drill ships, oil rigs, tankers, trenchers, pipeline lay barges), seismic exploration, mineral mining, ocean dredging, fishing vessels, cargo vessels, ocean acoustic research, military activities, ocean linears, ferries, pleasure boats, private boats, and the whale watching fleet. The example of icebreaker noise affecting beluga whales in the Beaufort Sea is discussed in a companion paper.[29]

ACKNOWLEDGMENTS

We wish to thank the Canadian Coast Guard, Central and Arctic Region, in particular Patrice St-Pierre, for continued support and funding of this project. Thanks to John Richardson for his very helpful and elaborate comments on this manuscript.

[1] C. S. Johnson, ''Masked tonal thresholds in the bottlenosed porpoise,'' J. Acoust. Soc. Am. **44**, 965–967 (1968).
[2] C. S. Johnson, ''Auditory masking of one pure tone by another in the bottlenose porpoise,'' J. Acoust. Soc. Am. **49**, 1317–1318 (1971).
[3] V. I. Burdin, V. I. Markov, A. M. Reznik, V. M. Skornyakov, and A. G. Chupakov, ''Ability of *Tursiops truncatus* Ponticus Barabasch to distinguish a useful signal against a noise background,'' in *Morphology and Ecology of Marine Mammals*, edited by K. K. Chapskii and V. E. Sokolov (Wiley, New York, 1973), pp. 162–168.
[4] C. S. Johnson, M. W. McManus, and D. Skaar, ''Masked tonal hearing thresholds in the beluga whale,'' J. Acoust. Soc. Am. **85**, 2651–2654 (1989).
[5] W. W. L. Au and P. W. B. Moore, ''Critical ratio and critical bandwidth for the Atlantic bottlenose dolphin,'' J. Acoust. Soc. Am. **88**, 1635–1638 (1990).
[6] J. A. Thomas, J. L. Pawloski, and W. W. L. Au, ''Masked hearing abilities in a false killer whale (*Pseudorca crassidens*),'' in *Sensory Abilities of Cetaceans—Laboratory and Field Evidence*, edited by J. A. Thomas and R. A. Kastelein (Plenum, New York, 1990), pp. 395–404.
[7] A. Supin and V. Popov, ''Frequency-selectivity of the auditory system in the bottlenose dolphin, *Tursiops truncatus*,'' in Ref. 6, pp. 385–393.
[8] A. Y. Supin, V. V. Popov, and V. O. Klishin, ''ABR frequency tuning curves in dolphins,'' J. Comp. Physiol. A **173**, 649—656 (1993).
[9] C. Erbe and D. M. Farmer, ''Masked hearing thresholds of a beluga whale (*Delphinapterus leucas*) in icebreaker noise,'' Deep-Sea Res. II **45**, 1373–1388 (1998).
[10] C. Erbe, ''Detection of whale calls in noise: Performance comparison between a beluga whale, human listeners and a neural network,'' J. Acoust. Soc. Am. **108**, 297–303 (2000).
[11] W. J. Richardson, C. R. Greene, Jr., C. I. Malme, and D. H. Thomson, *Marine Mammals and Noise* (Academic, San Diego, CA, 1995).
[12] W. W. L. Au, P. E. Nachtigall, and J. L. Pawloski, ''Temporal threshold shift in hearing induced by an octave band of continuous noise in the bottlenose dolphin,'' J. Acoust. Soc. Am. **106**, 2251 (1999).
[13] D. Kastak, R. J. Schusterman, B. L. Southall, and C. J. Reichmuth, ''Underwater temporary threshold shift induced by octave-band noise in three species of pinniped,'' J. Acoust. Soc. Am. **106**, 1142–1148 (1999).
[14] J. Bowlin, J. Spiesberger, T. Duda, and L. Freitag, ''Ocean acoustical ray-tracing software RAY,'' Tech. Rep. WHOI-93-10, Woods Hole Oceanographic Institution, Woods Hole, MA, 1992.
[15] C. Erbe, A. R. King, M. Yedlin, and D. M. Farmer, ''Computer models for masked hearing experiments with beluga whales (*Delphinapterus leucas*),'' J. Acoust. Soc. Am. **105**, 2967–2978 (1999).
[16] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, Woodbury, NY, 1994).
[17] R. J. Urick, *Sound Propagation in the Sea* (Peninsula, Los Altos, CA, 1982).
[18] W. H. Thorp, ''Analytic description of the low-frequency attenuation coefficient,'' J. Acoust. Soc. Am. **42**, 270 (1967).
[19] F. M. Fisher and V. P. Simmons, ''Sound absorption in sea water,'' J. Acoust. Soc. Am. **62**, 558–564 (1977).
[20] H. A. Laible and S. D. Rajan, ''Temporal evolution of under ice reflectivity,'' J. Acoust. Soc. Am. **99**, 851–865 (1996).
[21] L. M. Brekhovskikh, *Waves in Layered Media* (Academic, New York, 1960).
[22] L. Brekhovskikh and Yu. Lysanov, *Fundamentals of Ocean Acoustics* (Springer-Verlag, Berlin, 1982).
[23] C. S. Clay and H. Medwin, *Acoustical Oceanography: Principles and Applications* (Wiley, New York, 1977).
[24] E. L. Hamilton, ''Geoacoustic modeling of the sea floor,'' J. Acoust. Soc. Am. **68**, 1313–1340 (1980).
[25] J. O. Pickles, *An Introduction to the Physiology of Hearing* (Academic, San Diego, CA, 1988).
[26] B. C. J. Moore, *An Introduction to the Psychology of Hearing* (Academic, San Diego, CA, 1997).
[27] B. Scharf, ''Critical bands,'' in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York, 1970), pp. 157–202.
[28] H. Fletcher, ''Auditory patterns,'' Rev. Mod. Phys. **12**, 47–65 (1940).
[29] C. Erbe and D. M. Farmer, ''Zones of impact around icebreakers affecting beluga whales in the Beaufort Sea,'' J. Acoust. Soc. Am. **108**, 1332–1340 (2000).

Zones of impact around icebreakers affecting beluga whales in the Beaufort Sea

Christine Erbe[a] and David M. Farmer
Institute of Ocean Sciences, 9860 West Saanich Road, Sidney, British Columbia V8L 4B2, Canada

(Received 27 July 1999; accepted for publication 16 June 2000)

A software model estimating zones of impact on marine mammals around man-made noise [C. Erbe and D. M. Farmer, J. Acoust. Soc. Am. **108**, 1327–1331 (2000)] is applied to the case of icebreakers affecting beluga whales in the Beaufort Sea. Two types of noise emitted by the Canadian Coast Guard icebreaker *Henry Larsen* are analyzed: bubbler system noise and propeller cavitation noise. Effects on beluga whales are modeled both in a deep-water environment and a near-shore environment. The model estimates that the *Henry Larsen* is audible to beluga whales over ranges of 35–78 km, depending on location. The zone of behavioral disturbance is only slightly smaller. Masking of beluga communication signals is predicted within 14–71-km range. Temporary hearing damage can occur if a beluga stays within 1–4 km of the *Henry Larsen* for at least 20 min. Bubbler noise impacts over the short ranges quoted; propeller cavitation noise accounts for all the long-range effects. Serious problems can arise in heavily industrialized areas where animals are exposed to ongoing noise and where anthropogenic noise from a variety of sources adds up. [S0001-4966(00)04409-X]

PACS numbers: 43.80.Nd [WA]

I. INTRODUCTION

Over the past few decades, public concern for the well-being of marine mammals has rapidly increased. Due to human impact, many species are listed on endangered species lists around the world. Some populations no longer exist, i.e., are extirpated. While hunting used to be the major factor reducing marine mammal population numbers many years ago, habitat degradation is now the major issue. This includes chemical water contamination (sewage), waste disposal (plastics, glass, metal, nets), over-harvesting of prey (fisheries), and anthropogenic noise. Man-made underwater noise has many sources: offshore hydrocarbon exploration (drill ships, oil rigs, tankers, trenchers, pipeline lay barges), seismic exploration, mineral mining, ocean dredging, fishing vessels, ocean acoustic research, military activities, and ship traffic ranging from large cargo vessels, ocean liners, passenger vessels, and ferries to small private boats, pleasure boats, and whale watching boats.

There are many different ways in which anthropogenic noise affects marine mammals, some of which likely play an important role in a population's survival. Noise can disturb animal behavior. In cases where noise causes avoidance reactions and animals temporarily leave the area of loud noise exposure, this is unlikely to be biologically significant. However, if important behavior such as mating, nursing, or feeding is disrupted, or if animals are scared away from critical habitat over long periods of time, the impact can affect the long-term survival of the population. Noise further has the potential to interfere with the animal's communication signals, echolocation signals in the case of odontocetes, environmental sounds (e.g., surf) animals might listen to for orientation, the sound of prey, and the sound of predators. In extreme cases, loud continuous noise or sudden blasts of noise can cause physiological damage to the ear or other organs and tissues.

In many countries, efforts are being made to reduce the risk of noise damage to marine mammals. There is a need for efficient, i.e., fast and rational, tools to estimate over which ranges noise affects animals in which way. Various mitigation methods can then be applied, see Ref. 1 for a detailed review. As examples, quieter vessels could substitute for noisier ones, loud equipment be altered or replaced, shipping routes changed and construction sites moved away from critical marine mammal habitat. Some operations such as seismic exploration or ocean acoustic research could be timed to take place in seasons of less marine mammal abundance. Operational procedures can be modified if marine mammals are sighted in the vicinity, e.g., ramping-up a sound source, reducing source levels, changing spectral characteristics, selecting duty cycles, or temporarily shutting down if animals are within an ''unsafe radius.''

For all of these mitigation procedures knowledge of the range over which noise affects marine mammals and in which way is crucial. Erbe and Farmer[2] developed a software package that combines a sound propagation model and impact threshold models. Given the source spectrum of a noise, the package estimates zones of audibility, behavioral disturbance, masking, and potential hearing damage for a marine mammal target species and any particular ocean environment. In the current article, this software package is applied to the case of icebreaker noise affecting beluga whales in the Beaufort Sea.

II. INPUT PARAMETERS OF THE SOFTWARE PACKAGE

A. Noise emitted by an icebreaker

Noise emitted by the Canadian Coast Guard icebreaker CCGS *Henry Larsen* was recorded while on route through

[a]Electronic mail: erbec@dfo-mpo.gc.ca

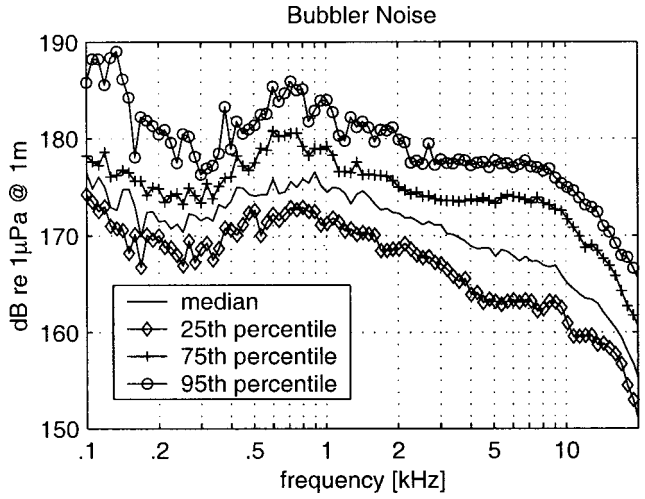


FIG. 1. Bubbler noise statistics. Plotted are 12th octave band levels of source spectra at 1-m range. Source levels are 189 (25th percentile), 192 (median), 195 (75th percentile) and 201 dB *re* 1 μPa at 1 m (95th percentile).

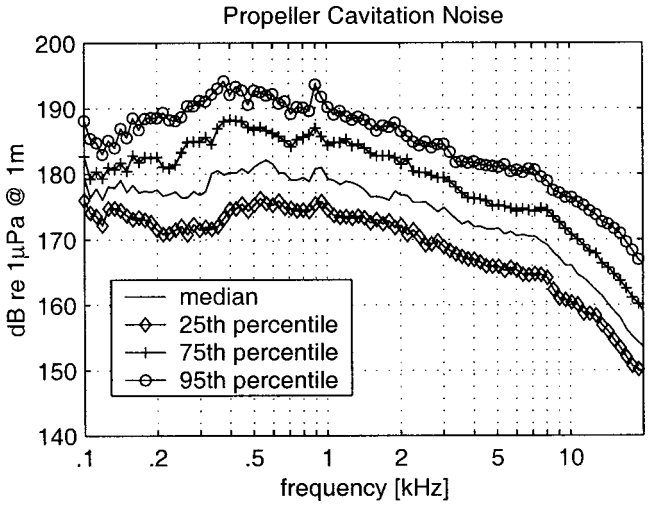


FIG. 2. Propeller cavitation noise statistics. Plotted are 12th octave band levels of source spectra at 1-m range. Source levels are 192 (25th percentile), 197 (median), 201 (75th percentile) and 205 dB *re* 1 μPa at 1 m (95th percentile).

the Beaufort Sea during August 1991. Two main types of noise were identified: bubbler system noise and propeller cavitation noise.[3] Some icebreakers are equipped with a so-called bubbler system blowing high-pressure air into the water in order to push floating ice away from the ship. The air bubbles introduced into the sea surface during the process make the ocean noisy. We call this type of noise ''bubbler noise.'' It was temporarily continuous and had a fairly white sound spectrum with most of its energy below 5 kHz. Spectrograms of single noise clips were shown earlier.[3] Figure 1 shows the statistical source spectrum levels in 12th octave bands between 100 Hz and 20 kHz, based on 36 sound files of 10-s length each. We calculated a median source level of 192 dB *re* 1 μPa at 1 m in this frequency range.

Propeller cavitation noise is caused by every propeller-driven ship. When the *Henry Larsen* tried to break an ice-ridge after building up momentum, but failed and was stopped by the ice with the propeller still turning at full speed, the noise was strongest. In this case, we refer to the noise as ''ramming noise.'' The frequency spectrum of propeller cavitation noise was broadband with energy up to 20 kHz (the maximum sampled frequency in our recording). The noise was not continuous with time, but consisted of sharp pulses occurring about 11 times per second (which was equal to the rotation frequency of the propeller times the number of blades). Figure 2 shows 12th octave band levels of propeller cavitation noise. Eighty-six sound files of 10-s length were used for the statistics. We calculated a median source level of 197 dB *re* 1 μPa at 1 m between 100 Hz and 22 kHz. For the subsequent analysis of ramming noise, we chose the loudest cavitation noise, i.e., the 95th percentile, with a corresponding source level of 205 dB *re* 1 μPa at 1 m.

B. Sound propagation modeling

The Beaufort Sea lies north of Alaska, the Yukon, and the western part of the Northwest Territories. A map of the study location is shown in Fig. 3. The chart shows the coastline and bathymetry lines for the continental shelf, continental slope, and abyssal plain of the Beaufort Sea. Pingo-like features exist on the continental shelf, locally reducing the water depth by 30%. A narrow shipping corridor through the pingo area is marked on nautical charts identifying individual pingos. The effects of ship noise on beluga whales were studied along two transects. Transect 1 assumed a ship in the shipping corridor and beluga whales anywhere between the shipping corridor and Beluga Bay. Transect 2 modeled offshore shipping and beluga whales. The sound propagation model was based on ray theory.[2] We accounted for frequency-dependent absorption by ocean water, surface scattering off the sea ice, and energy loss into the bottom sediment. We assumed a mixture of sand, silt, and clay as the sediment,[4] using its geoacoustic properties from Hamilton.[5] Mean temperature and salinity data with depth for the season of early autumn were obtained from the Levitus database of the International Research Institute for Climate Prediction and used to calculate sound speed profiles. During the fall, the chance of encountering a 50% ice-covered sea surface near the coast is 50%[6] with increasing probabilities further north over the continental slope and abyssal plain of the Beaufort Sea. We modeled first year surface ice with a rms roughness of 1 cm.[7] The sound propagation model took the noise source spectra as plotted in Figs. 1 and 2, and computed received noise spectra on a two-dimensional (depth versus range) grid of receiver locations.

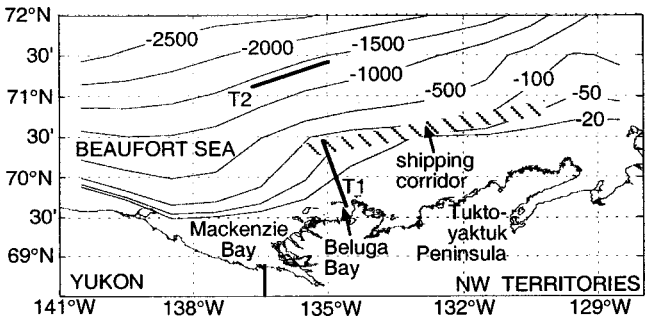


FIG. 3. Study site in the Beaufort Sea depicting transects T1 and T2.

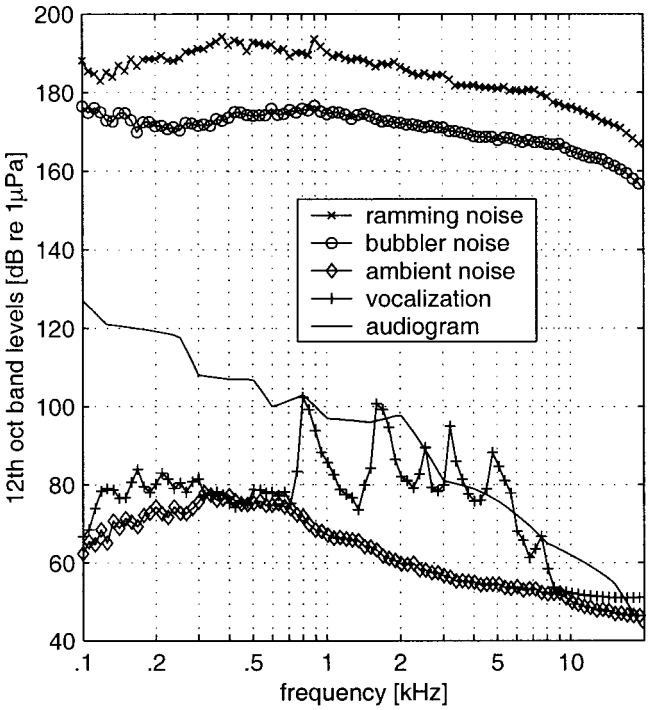


FIG. 4. The 12th octave band levels for ramming noise and median bubbler noise at the source (at 1 m), median ambient noise, a beluga vocalization at minimum recognizable level, and the beluga audiogram.

C. Calculation of the zone of audibility

This subroutine required knowledge of the widths of the critical bands of the animal's auditory filter. Erbe *et al.*[8] estimated critical bandwidths for beluga whales from critical ratio data.[9] It was shown that critical bands were about $\frac{1}{12}$ of an octave wide in the frequency range of interest. This subroutine also needed an audiogram (pure tone detection thresholds) of the target species. For beluga whales, seven audiograms have been published.[3,10–12] We calculated the means and interpolated at the center frequencies of the 12th octave bands, listed in Table I.[2] The resulting audiogram is shown in Fig. 4. Ambient arctic noise in the presence of first-year ice was recorded during an earlier study.[13,14] Figure 4 shows median 12th octave band levels based on ten recordings of 2-s length each, obtained over 5 days. In our case, ambient noise was mainly due to naturally occurring thermal and pressure icecracking, and wind and currents shuffling ice floes. The audibility model considered a noise source audible as long as the energy in at least one 12th octave band exceeded the audiogram and ambient noise.

D. Calculation of the zone of masking

This subroutine required spectra of the signals to be masked. Figure 4 shows the 12th octave band levels of a typical beluga vocalization at its minimum recognizable sound pressure level as determined by hearing experiments with a trained beluga whale.[3] In other words, in the absence of noise, this was the quietest level at which the animal recognized the vocalization. In the presence of noise, the same animal successfully recognized the call, if the energy in the 12th octave bands of the two lower frequency peaks was above the corresponding band levels of the noise.[15] This corroborated Fletcher's principle[16] stating that a signal will be masked by noise with equal or higher energy in the same critical band. In this subroutine, as long as both 12th octave band levels of the two lower-frequency peaks of the call were above the noise, the signal was assumed recognizable, otherwise it was masked.

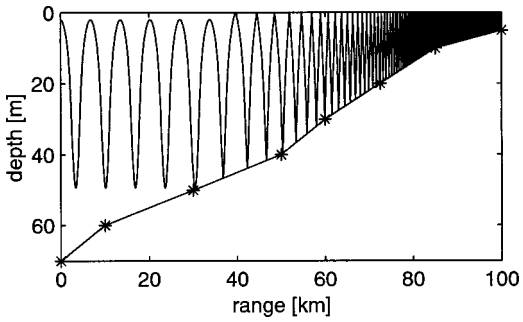


FIG. 5. Ray path up the continental slope along T1.

E. Calculation of the zone of disturbance

To estimate over what range disturbance can occur, data from field experiments are needed. LGL and Greeneridge observed groups of migrating beluga whales in the Beaufort Sea during playbacks of icebreaker noise. Six out of 17 groups altered their path when received levels were around 81 dB *re* 1 μPa (3rd octave band level at 5 kHz) during the strongest phases of propeller cavitation noise. We modified the subroutine to calculate this 3rd octave band level for the noises as a function of range and depth and we used 81 dB as the disturbance threshold.

F. Calculation of the zone of potential hearing damage

Two data sets were used to estimate over what ranges a temporary threshold shift (TTS) in hearing might occur. Au *et al.*[17] measured a TTS of 12–18 dB at 7.5 kHz after exposing a bottlenose dolphin for 30–50 min to an octave band of noise 96 dB above the audiogram. We calculated (overlapping) octave band levels at all frequencies listed in Table I,[2] and applied the 96-dB criterion to all frequencies. Kastak *et al.*[18] measured a TTS of on average 4.8 dB with three pinniped species after 20 min in octave band noise 60–75 dB above the audiogram.

III. RESULTS

A. Sound propagation

For transect T1, Fig. 5 shows a ray leaving a sound source in the shipping corridor at range 0, at 2-m depth, at a near-zero angle of -0.005 degrees. The ray starts out as a surface-and-bottom refracted ray in the mixed layer sound channel. As soon as it hits the upward sloping bottom, the ray follows a bottom-and-surface reflected path. Rays leaving the source at greater angles will turn into a reflected ray sooner. With each surface and bottom reflection, energy is lost. Therefore, only very little sound energy climbs up the

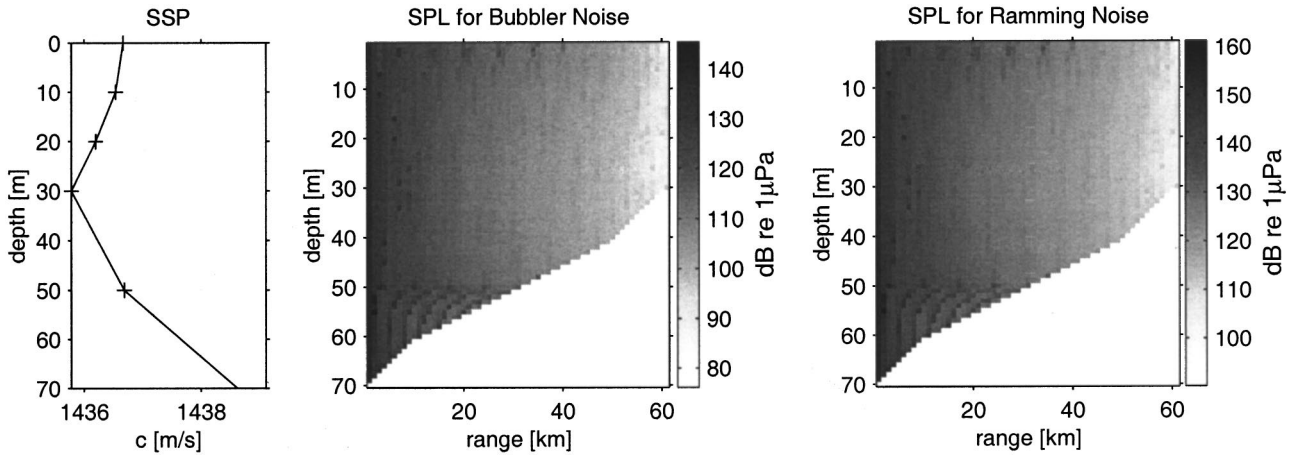


FIG. 6. Received sound pressure levels of median bubbler noise and ramming noise (95th percentile cavitation noise) along T1.

continental slope and is audible in the shallow coastal water of Beluga Bay. Acoustic energy traveling through the sea floor was neglected.

Figure 6 shows the sound speed profile (SSP) used for T1 and the received sound pressure levels (SPL) as a function of range and depth. The gray scale is in dB *re* 1 μPa. The source (the icebreaker) is located in the upper left corner. Received levels are fairly independent of depth in this shallow water and decrease with range. Beyond 60-km range, the water becomes shallower than 30 m and the SSP becomes downward refracting. Energy is lost very fast; basically no sound reaches water shallower than 30 m. We used a mean levitus SSP (early autumn season) from the deep end of T1 for the entire transect. Depending on the freshwater outflow from the Mackenzie River west of Beluga Bay, the SSP near shore might be highly variable and thus change received sound levels. Our sound propagation model allows for range-dependent SSPs, however, we did not have access to enough SSP data to do this.

Along transect T2 over the abyssal plain of the Beaufort Sea, ray propagation is typically arctic (upward refracting, Fig. 7). A shallow sound channel exists in the mixed surface layer. Figure 8 plots received sound pressure levels as a function of range and depth with the icebreaker again located in the upper left corner. Received levels decrease with range and depth. However, multiple convergence zones appear, where rays converge leading to high received sound levels.

B. Zones of audibility

Table I summarizes the impact ranges for the medians and percentiles of bubbler noise and propeller cavitation noise. A plot of the zone of audibility for median bubbler noise along T1 is shown in Fig. 9(a). The noise was audible to beluga ears at all depths out to ranges of 32 km (35 km on average over all depths). Figure 10 shows received bubbler spectra (in 12th octave bands) at a constant depth of 20 m at various ranges. At short ranges, the entire spectrum was audible. With increasing range, the low-frequency end of the bubbler spectrum became inaudible first, due to the insensitivity of the beluga ear to these low frequencies. At ranges greater than 10 km, the high-frequency end of the bubbler spectrum became inaudible, due to the increased absorption of sound energy by seawater at high frequencies. At long ranges, it was only the mid-frequencies between 3 and 10 kHz that were audible.

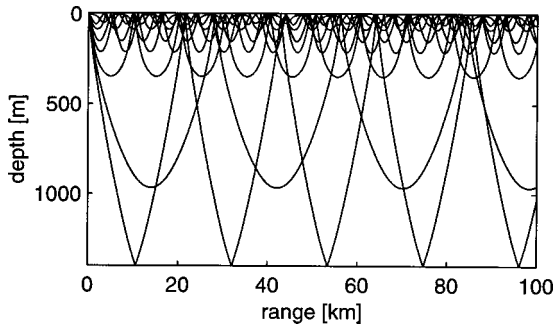


FIG. 7. Ray paths along T2 over the abyssal plain of the Beaufort Sea.

Along T1, ramming noise was audible at all depths to ranges of 50–54 km [Fig. 11(a)]. At short ranges, all frequencies were audible. Beyond 2-km range, frequencies below 500 Hz became inaudible. Beyond 40-km range, the high-frequency end of the spectrum became inaudible. At the longest ranges, it was only the mid-frequencies between 3 and 5 kHz that were audible.

Along T2, bubbler noise was audible to all depths down to 1400 m for ranges of 19 km [Fig. 12(a)]. In less deep water, particularly along the convergence zones, bubbler noise was audible to a maximum range of 53 km. At short ranges, all frequencies were audible. Beyond 15 km, only the band between 3 and 10 kHz was audible. In the case of ramming noise, audibility extended to 33 km at depth and 78 km in the upper 200 m [Fig. 13(a)]. Similar to bubbler noise, at short ranges, the entire spectrum was audible. Beyond 20 km, only spectral energy between 500 Hz and 10 kHz was audible to belugas.

In general, the range of audibility is limited by the ship noise levels dropping either below the animal's audiogram or below the ambient noise. In our case, ambient noise was only audible to beluga whales above 16 kHz. Therefore, only at higher frequencies was the ship audibility limited by the ambient noise. LGL and Greenridge[19] measured 3rd octave levels of ambient noise up to 6.3 kHz in the Beaufort Sea at the same time of the year that our ambient noise was measured. Integrating our measurements into 3rd octave bands, noise levels agreed between 150 Hz and 1 kHz. For higher frequencies, our ambient noise dropped off faster, creating a

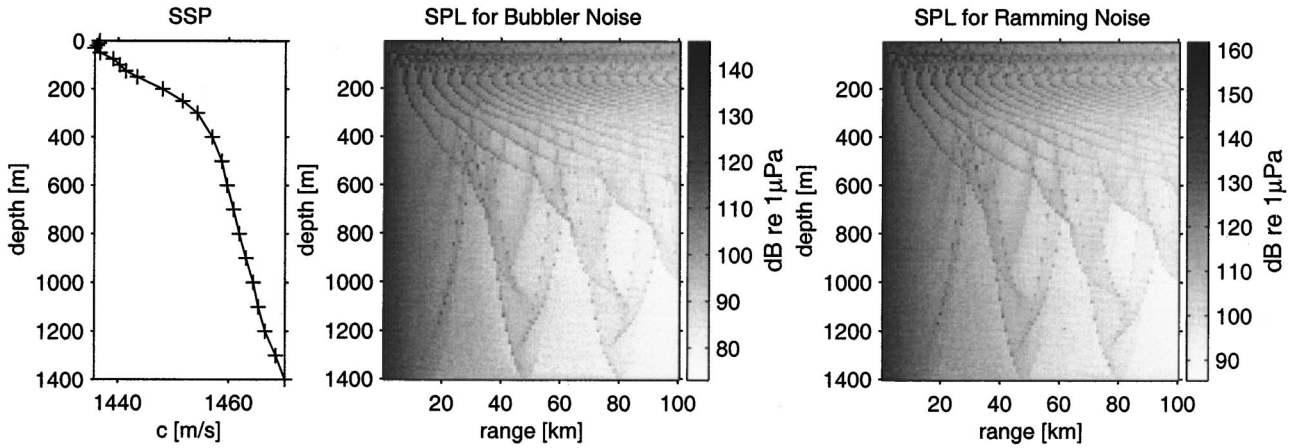


FIG. 8. Received sound pressure levels of median bubbler noise and ramming noise (95th percentile cavitation noise) along T2.

difference of 13 dB at 6 kHz. Greene[20] measured ambient noise up to 1.6 kHz in the same area, obtaining similar band levels to ours. Cosens and Dueck[21] measured spectrum levels between 20 and 5000 Hz. Integrating into 12th octave bands, their levels agreed with ours up to 1.5 kHz. Above, our levels dropped faster, creating a difference of 15 dB at 5 kHz. Ambient arctic noise is highly variable, depending on wind and sea state, on how broken up the ice is, and whether or not measurements were taken in a wind-shaded are behind ice ridges, or floes. Raised high-frequency ambient noise levels measured by LGL and Greeneridge[19] and Cosens and Dueck[21] would be audible to beluga whales above 5 kHz. Only in the case that there is enough ship noise energy above 5 kHz that propagates to long ranges can such ambient noise levels decrease the range of audibility of the ship noise.

C. Zones of masking

The following ranges are maximum ranges of masking, because the model was based on masking a very faint call that was barely recognizable in the absence of the icebreaker. Figure 9(b) indicates that masking of the beluga vocalization by bubbler noise along T1 over the continental slope occurred at all depths out to a range of about 14 km. Ramming noise masked over about 40 km range [Fig. 11(b)]. Over the abyssal plain, Figs. 12(b) and 13(b), the extent of the zone of masking was 6 km at depth and 29 km near the surface in the case of median bubbler noise, compared to 18 km at depth and 71 km near the surface in the case of ramming noise. In our case, ambient arctic noise did not add to the masking of icebreaker noise because it was considerably below audibility at the frequencies of the call. This would also be the case for the higher ambient noise levels measured by LGL and Greeneridge[19] and Cosens and Dueck.[21]

D. Zones of disturbance

As shown in Fig. 9(c) for bubbler noise along T1, the zone of behavioral disturbance extended to ranges of about 32 km and was thus almost as large as the zone of audibility (35 km). For ramming noise, disturbance went out to 46 km [Fig. 11(c)]. Over the abyssal plain [Fig. 12(c)] the range of disturbance was 19 km at depth and 44 km in shallow water

TABLE I. Impact ranges around bubbler and propeller cavitation noise (medians, 25th, 75th, and 95th percentiles). In cases where impact ranges do not vary much with depth, a mean range is given. If only one distance is listed (e.g., bubbler noise, T1, 95th percentile, audibility: 43 km), the distance refers to range and the impact occurred at all depths down to the seafloor. If the zone of impact does not extend to all depths (e.g., bubbler noise, T1, 75th percentile, TTS 12–18 dB), the maximum range ''R'' (in this case 80 m) and the maximum depth ''D'' (in this case 55 m) are given. In cases where impact ranges vary a lot with depth (e.g., bubbler noise, T2, 95th percentile, audibility), the range of impact is given both in deep ''d'' water (here 29 km) and in shallow ''s'' water (here 61 km).

	T1				T2			
	95%	75%	50%	25%	95%	75%	50%	25%
				Bubbler noise				
Audibility	43 km	39 km	35 km	29 km	29 km d, 61 km s	23 km d, 56 km s	19 km d, 53 km s	16 km d, 44 km s
Masking	30 km	19 km	14 km	10 km	12 km d, 47 km s	9 km d, 42 km s	6 km d, 29 km s	3 km d, 20 km s
Disturbance	40 km	37 km	32 km	24 km	26 km d, 53 km s	23 km d, 50 km s	19 km d, 44 km s	14 km d, 36 km s
TTS 12–18 dB	120 m	80 mR, 55 mD	40 mR, 25 mD	20 mR, 13 mD	160 mR, 100 mD	100 mR, 60 mD	40 mR, 30 mD	20 mR, 10 mD
TTS 4.8 dB	3–4 km	2–3 km	1–2 km	1 km	2 km	1–2 km	1kmR, 750 mD	1 kmR, 600 mD
				Propeller cavitation noise				
Audibility	52 km	47 km	41 km	33 km	33 km d, 78 km s	28 km d, 68 km	23 km d, 54 km s	17 km d, 48 km s
Masking	40 km	31 km	21 km	14 km	18 km d, 71 km s	14 km d, 52 km s	8 km d, 37 km s	4 km d, 26 km s
Disturbance	46 km	39 km	35 km	29 km	30 km d, 62 km s	25 km d, 55 km s	21 km d, 48 km s	16 km d, 42 km s
TTS 12–18 dB	120 m	60 mR, 45 mD	20 mR, 15 mD	none	120 mR, 100 mD	80 mR, 50 mD	40 mR, 20 mD	20 mR, 10 mD
TTS 4.8 dB	3–4 km	2 km	1–2 km	1 km	2 km	1 km	1 km R, 500 mD	900 mR, 300 mD

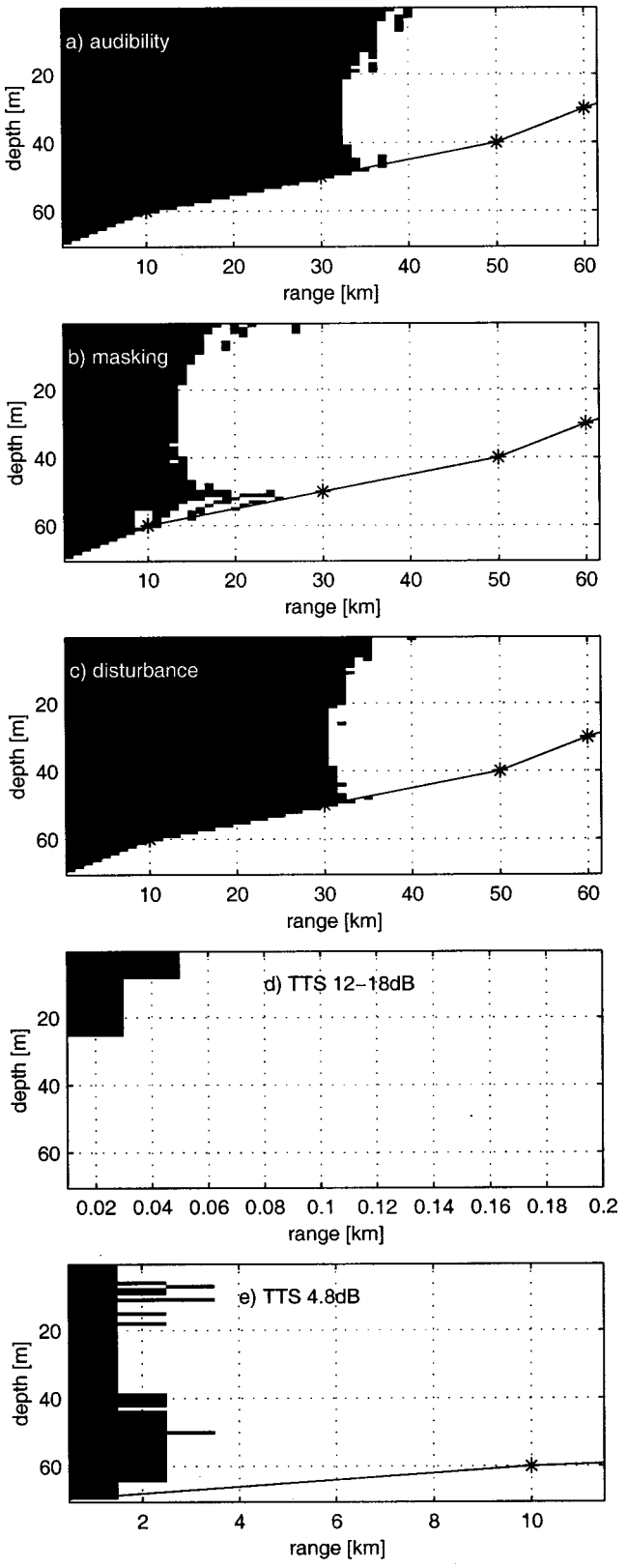


FIG. 9. Zones of impact around median bubbler noise for T1. Note the different range scale in (d) and (e).

for bubbler noise. For ramming noise [Fig. 13(c)] disturbance occurred to 30 km at depth and 62 km near the surface. The predicted zone of disturbance was only slightly smaller than the predicted zone of audibility in all four cases.

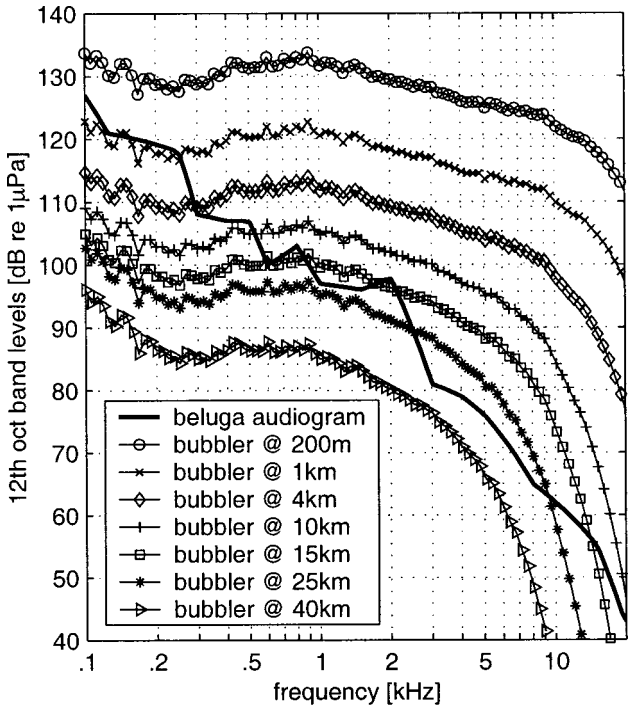


FIG. 10. Received 12th octave band levels of median bubbler noise at various ranges, taken at a depth of 20 m.

E. Zones of potential hearing damage

For median bubbler noise along T1 [Fig. 9(d)] we modeled a TTS of 12–18 dB if the animal stayed within 40-m range and 25-m depth of the icebreaker for over 30 min. It is unlikely that an animal would stay that close for that long. A smaller TTS of 4.8 dB is more likely, if the animal stays within 1–2 km for 20 min [Fig. 9(e)]. Ramming noise [Figs. 11(d) and (e)] had a larger range of impact. We predicted a TTS of 12–18 dB at all depths out to 120-m range, and a TTS of 4.8 dB over 3–4 km range. Over the abyssal plain, in the case of bubbler noise [Figs. 12(d) and (e)], a large TTS was possible to depths of 30 m and ranges of 40 m. A small TTS was possible over 750-m depth and 1-km range. In the case of ramming noise [Figs. 13(d) and (e)], a TTS of 12–18 dB could occur if an animal spent over 30 min within 120-m range and 100-m depth of the icebreaker. A TTS of 4.8 dB could occur if an animal spent over 20 min within 2-km range at all depths down to 1400 m.

From an examination of received spectra, we concluded that in all cases, hearing damage would occur only at the highest frequencies looked at, i.e., mostly between 10 and 20 kHz. The smaller TTS of 4.8 dB occurred more broadband, i.e., to lower frequencies, than the higher TTS of 12–18 dB. Also, the higher the percentile of the statistical noise spectrum (i.e., the louder the source), and the closer the animal was to the ship, the more did the hearing damage extend to lower frequencies. For example, the 25th percentile of bubbler noise along T1 was predicted to cause a TTS of 4.8 dB between 1 and 20 kHz. The 95th percentile, however, was predicted to cause a TTS of 4.8 dB between 300 Hz and 20 kHz. Ramming noise (95th percentile propeller cavitation noise) along T2 caused a TTS of 12–18 dB between 5 and 20 kHz and a TTS of 4.8 dB between 170 Hz and 20 kHz.

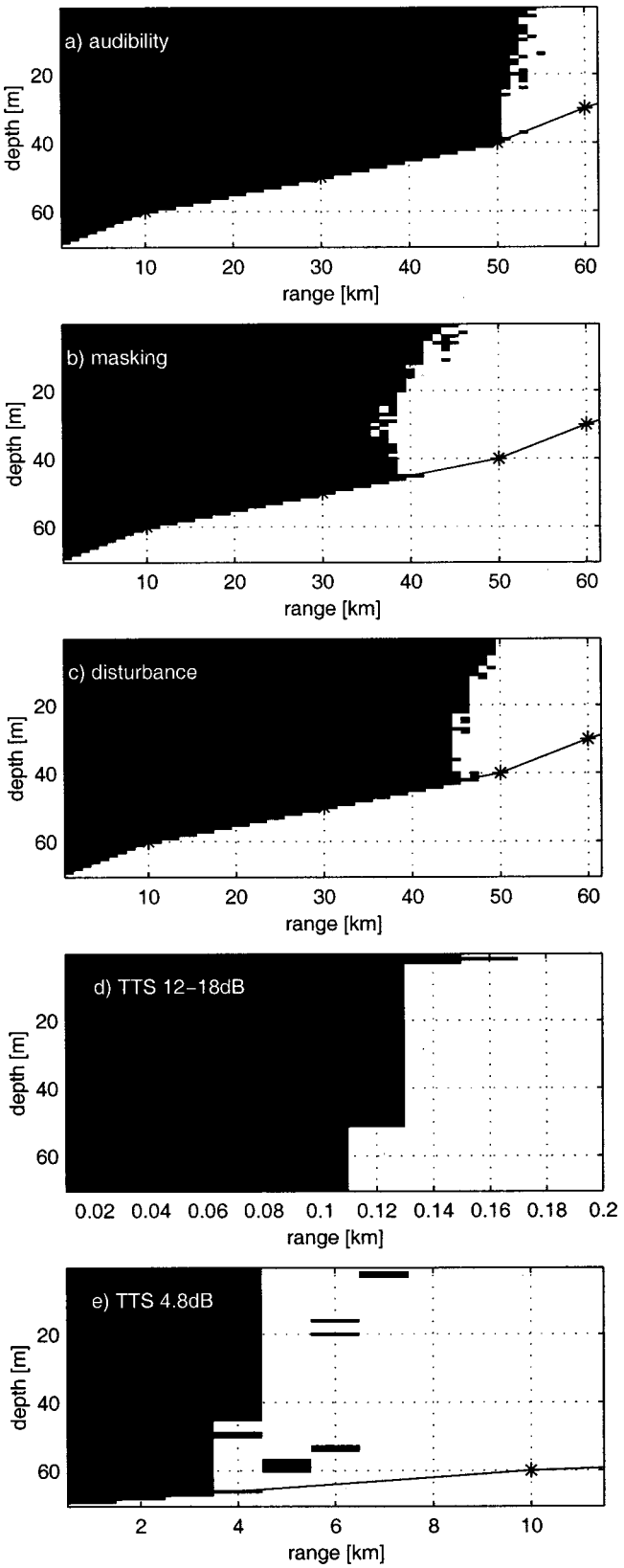


FIG. 11. Zones of impact around ramming noise for T1. Note the different range scale in (d) and (e).

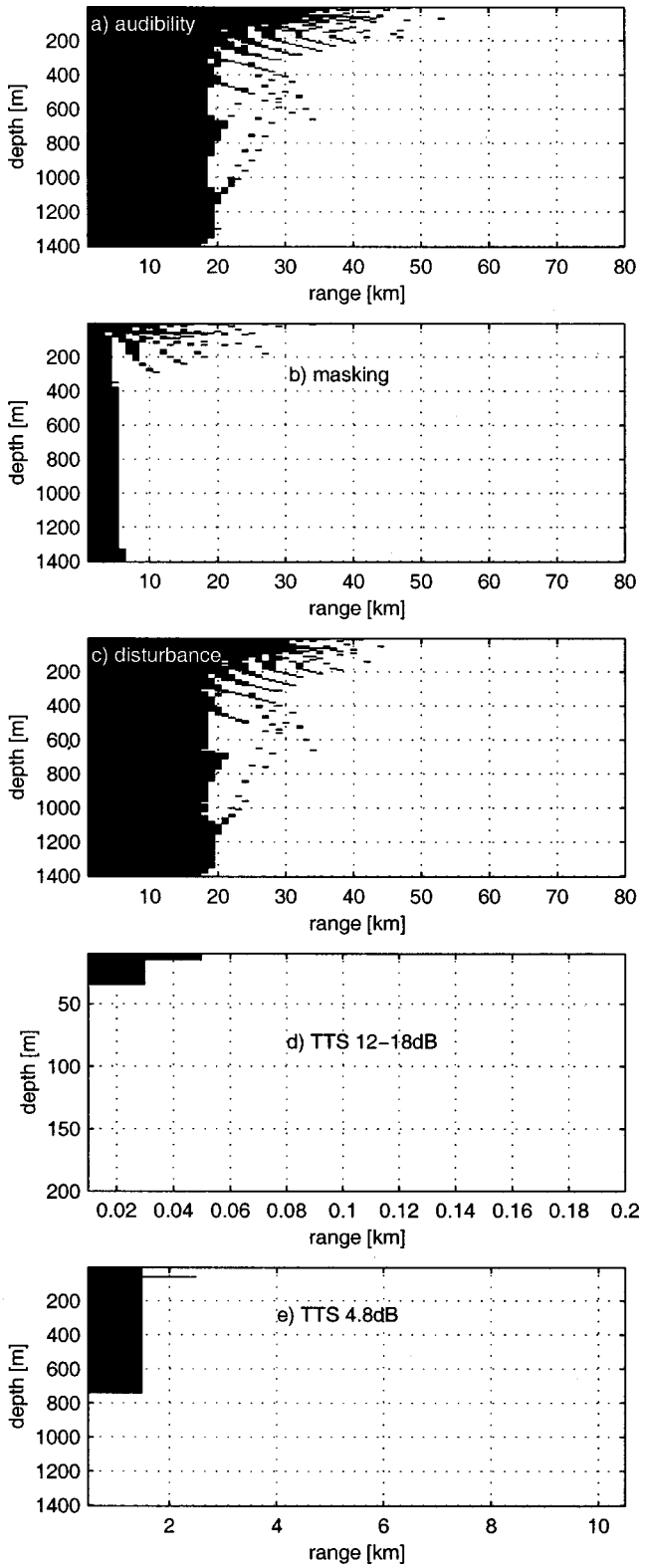


FIG. 12. Zones of impact around median bubbler noise for T2. Note the different range scale in (d) and (e), and the different depth scale in (d).

IV. DISCUSSION

A software package has been developed which combines a sound propagation model for broadband sound and impact threshold models for noise effects on marine mammals.[2] In this article, the tool was applied to estimate the ranges over which icebreakers may affect beluga whales in the Beaufort Sea. Bubbler system noise and propeller cavitation noise recorded from the Canadian Coast Guard icebreaker *Henry Larsen*[3] were analyzed for two ship locations: over the abyssal plain in the Beaufort Sea (T2) and in the shipping corridor over the continental shelf near Beluga Bay (T1). For both

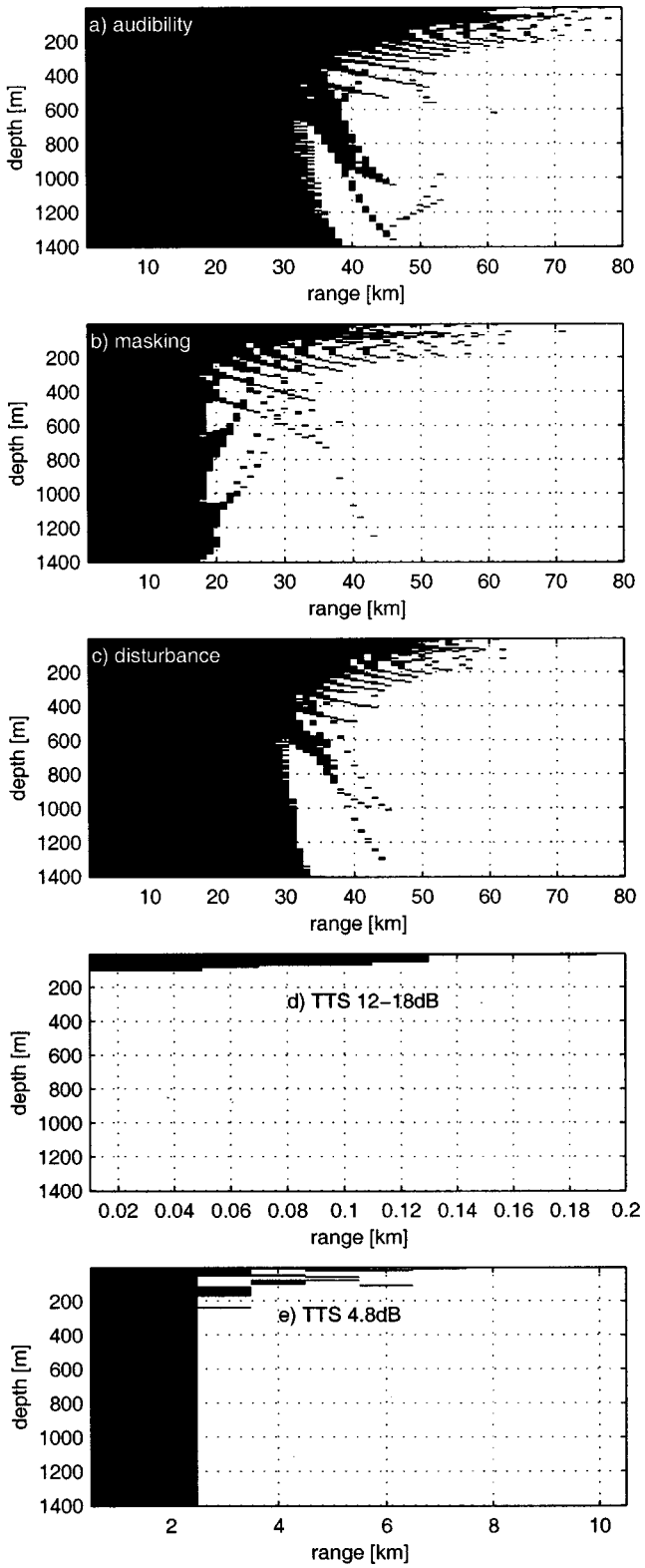


FIG. 13. Zones of impact around ramming noise for T2. Note the different range scale in (d) and (e).

noises, median spectra as well as 25th, 75th, and 95th percentiles were given. The 95th percentiles of propeller cavitation noise were taken to model ramming noise, the loudest part of cavitation noise occurring when the icebreaker rams an iceridge and is stopped by the ice.

Results showed that the *Henry Larsen* was audible to beluga whales in the Beaufort Sea over very long ranges of 35 (T1) to 53 km (T2) due to median bubbler noise and 52 (T1) to 78 km (T2) when ramming noise occurred. The range over which her noise interfered with beluga communication, however, was up to 24 km shorter. This was because of the different spectral characteristics of call and noise. The call was relatively low in frequency, with its 800-Hz and 1.6-kHz components most important in masking.[15] But it was the mid-frequencies of the noise (3–10 kHz) that remained audible over long ranges.

The ranges of masking were maximum, in the sense that the call was taken at its quietest recognizable level in the absence of noise. In other words, two communicating animals were modeled as being furthest apart. For animals closer together, received call levels would likely be louder, making the range over which ship noise could mask shorter. The relationship between animal–animal distance, ship–animal distance, and range of masking was illustrated elsewhere.[22]

Masking is a very complex and still poorly understood process and affects a variety of acoustic signals important to marine animals. We only examined the masking of intraspecies communication signals. The masking of echolocation signals, environmental cues, and predator and prey sounds was not analyzed. In none of the cases do we know the biological significance of masking. What are the long-term effects on an individual and a population? It is further unclear if and what type of means marine mammals might have to avoid masking. With respect to masking of their own sounds, this could be done by changing the volume and spectral characteristics of emitted signals. Humans adapt the loudness of their speech according to the loudness of ambient noise, the loudness with which they receive incoming speech, and the loudness with which they perceive their own signals.[23] Dolphins have been shown to echolocate louder and change the frequency spectrum of emitted clicks in the presence of noise.[24] Lesage *et al.*[25] measured the vocal behavior of beluga whales in the St. Lawrence in the absence and presence of a ferry and a small motorboat. The animals emitted calls repetitively, changed the types of calls used, and shifted the mean call frequency up during noise exposure. Whether animals manage to communicate the same information during noise exposure, or whether calls heard are simply ''alarm calls,'' is unknown.

Zones of disturbance were large and only slightly smaller than the predicted zones of audibility. This is in accordance with observed reaction distances. Cosens and Dueck[26] as well as Finley *et al.*[27] observed changes in beluga swimming behavior at distances of 40–60 km from an icebreaker in Lancaster Sound in the Canadian High Arctic. Cosens and Dueck[21] concluded that these animals avoid an icebreaker as soon as they detect it. Habituation or sensitization, however, affects the extent of the zone of disturbance. Beluga whales in the St. Lawrence Estuary approach large ships to much shorter distances.[28,29] One possible explanation is that these animals are more used to heavy traffic and habituated. On the other hand, this beluga population might be hearing impaired because of ongoing noise exposure or,

in fact, (PCB) water contamination or parasites affecting their hearing.

Data on what sound levels over what periods of time cause either temporary (TTS) or permanent threshold shift (PTS) in beluga whales do not exist. For an estimation of the range of hearing damage, we used TTS data from a bottlenose dolphin[17] and three pinniped species.[18] A TTS of 12–18 dB was modeled if beluga whales stayed within 40 m of median bubbler noise and within 120 m of ramming noise for over 30 min. Given the high mobility of beluga whales, this is unlikely. However, for a TTS of 4.8 dB to occur, an animal would ''only'' have to be within 1–2 km of median bubbler noise and 2–4 km of ramming noise for 20 min. This is conceivable. Hearing can be expected to return to normal within 24 h.[18] Unfortunately, it is not known if repeated exposure to TTS, particularly what noise dose, causes permanent hearing damage.

In summary, the *Henry Larsen* icebreaker studied in this project was audible to beluga whales over very long ranges of 35–78 km. Arctic beluga whales generally avoid icebreakers almost as soon as they detect them. The animals do not get close enough for potentially harmful effects to occur such as masking of their communication signals or damage to their auditory system. However, if the animals are engaged in important behavior such as mating, nursing, or feeding, they might not flee but put up with louder, possibly too loud, noise. Problems can arise particularly in heavily industrialized areas where a variety of noisy activities take place such as geophysical (seismic) exploration; oil drilling; mineral mining; offshore construction; helicopter, icebreaker, tanker, cargo, freighter, fishing (factory), and passenger vessel traffic; and ocean acoustics research. Summed noise levels could be very high and ongoing for long durations and cover large areas such that animals might either be permanently scared away from critical habitat or be adversely affected because they have nowhere to flee to. While this article assessed the impact of the Canadian Coast Guard vessel *Henry Larsen* in particular, projects focusing on critical locations and addressing a large variety of man-made noise are needed. The tools developed here can be used for a variety of ocean environments and differing animal species.

ACKNOWLEDGMENTS

We wish to thank the Canadian Coast Guard, Central and Arctic Region, in particular Patrice St-Pierre, for continued support and funding of this project. Thanks to John Richardson for his most helpful, comprehensive comments on this manuscript.

[1] W. J. Richardson, C. R. Greene, Jr., C. I. Malme, and D. H. Thomson, *Marine Mammals and Noise* (Academic, San Diego, CA, 1995).
[2] C. Erbe and D. M. Farmer, ''A software model to estimate zones of impact on marine mammals around anthropogenic noise,'' J. Acoust. Soc. Am. **108**, 1327–1331 (2000).
[3] C. Erbe and D. M. Farmer, ''Masked hearing thresholds of a beluga whale (*Delphinapterus leucas*) in icebreaker noise,'' Deep-Sea Res., Part II **45**, 1373–1388 (1998).
[4] B. R. Pelletier, *Marine Science Atlas of the Beaufort Sea: Geology and Geophysics*, Geological Survey of Canada, Ottawa, Miscellaneous Report 40 (1987).
[5] E. L. Hamilton, ''Geoacoustic modeling of the sea floor,'' J. Acoust. Soc. Am. **68**, 1313–1340 (1980).
[6] J. C. LaBelle, *Alaska Marine Ice Atlas* (Arctic Environmental Information and Data Center, University of Alaska, Anchorage, 1983).
[7] H. A. Laible and S. D. Rajan, ''Temporal evolution of under ice reflectivity,'' J. Acoust. Soc. Am. **99**, 851–865 (1996).
[8] C. Erbe, A. R. King, M. Yedlin, and D. M. Farmer, ''Computer models for masked hearing experiments with beluga whales (*Delphinapterus leucas*),'' J. Acoust. Soc. Am. **105**, 2967–2978 (1999).
[9] C. S. Johnsons, M. W. McManus, and D. Skaar, ''Masked tonal hearing thresholds in the beluga whale,'' J. Acoust. Soc. Am. **85**, 2651–2654 (1989).
[10] M. J. White, Jr., J. Norris, D. Ljungblad, K. Baron, and G. diSciara, ''Auditory thresholds of two beluga whales *Delphinapterus leucas*,'' Report by Hubbs/Sea World Research Institute for Naval Ocean System Center, Report 78-109, San Diego (1978).
[11] F. T. Awbrey, J. A. Thomas, and R. A. Kastelein, ''Low-frequency underwater hearing sensitivity in belugas *Delphinapterus leucas*,'' J. Acoust. Soc. Am. **84**, 2273–2275 (1988).
[12] C. S. Johnson, M. W. McManus, and D. Skaar, ''Masked tonal hearing thresholds in the beluga whale,'' J. Acoust. Soc. Am. **85**, 2651–2654 (1989).
[13] D. M. Farmer and Y. Xie, ''The sound generated by propagating cracks in sea ice,'' J. Acoust. Soc. Am. **85**, 1489–1500 (1989).
[14] Y. Xie and D. M. Farmer, ''Acoustical radiation from thermally stressed sea ice,'' J. Acoust. Soc. Am. **89**, 2215–2231 (1991).
[15] C. Erbe, ''Detection of whale calls in noise: Performance comparison between a beluga whale, human listeners and a neural network,'' J. Acoust. Soc. Am. **108**, 297–303 (2000).
[16] H. Fletcher, ''Auditory patterns,'' Rev. Mod. Phys. **12**, 47–65 (1940).
[17] W. W. L . Au, P. E. Nachtigall, and J. L. Pawloski, ''Temporary threshold shift in hearing induced by an octave band of continuous noise in the bottlenose dolphin,'' J. Acoust. Soc. Am. **106** (4, Pt. 2), 2251 (1999).
[18] D. Kastak, R. J. Schusterman, B. L. Southall, and C. J. Reichmuth, ''Underwater temporary threshold shift induced by octave-band noise in three species of pinniped,'' J. Acoust. Soc. Am. **106**, 1142–1148 (1999).
[19] L. G. L. and Greeneridge, ''Acoustic effects of oil production activities on bowhead and white whales visible during spring migration near Pt. Barrow, Alaska-1991 and 1994 phases: Sound propagation and whale responses to playbacks of icebreaker noise,'' OCS Study MMS 95-0051, Report for U.S. Minerals Management Service, Herndon, VA, USA, Natl. Tech. Info. Serv. Catalogue No. NTIS PB98-107667 (1995).
[20] C. R. Greene, Jr., ''Characteristics of oil industry dredge and drilling sounds in the Beaufort Sea,'' J. Acoust. Soc. Am. **82**, 1315–1324 (1987).
[21] S. E. Cosens and L. P. Dueck, ''Icebreaker noise in Lancaster Sound, N.W.T., Canada: Implications for marine mammal behavior,'' Mar. Mam. Sci. **9**(3), 258–300 (1993).
[22] C. Erbe, ''The masking of beluga whale (*Delphinapterus leucas*) vocalizations by icebreaker noise,'' Ph.D. Thesis, University of British Columbia, Canada, Department of Earth and Ocean Sciences (1997).
[23] N. R. French and J. C. Steinberg, ''Factors governing the intelligibility of speech sounds,'' J. Acoust. Soc. Am. **19**, 90–119 (1947).
[24] W. W. L. Au, *The Sonar of Dolphins* (Springer, New York, 1993).
[25] V. Lesage, C. Barrette, M. C. S. Kingsley, and B. Sjare, ''The effect of vessel noise on the vocal behavior of belugas in the St. Lawrence River Estuary, Canada,'' Mar. Mam. Sci. **15**(1), 65–84 (1999).
[26] S. E. Cosens and L. P. Dueck, ''Responses of migrating narwhal and beluga to icebreaker traffic at the Admiralty Inlet ice-edge, N. W. T. in 1986,'' *Port and Ocean Engineering under Arctic Conditions*, edited by W. M. Sackinger and M. O. Jeffries (Geophysical Institute, Univ. of Alaska, Fairbanks, 1988), pp. 39–54.
[27] K. J. Finley, G. W. Miller, R. A. Davis, and C. R. Greene, ''Reactions of belugas (*Delphinapterus leucas*) and narwhals (*Monodon monoceros*) to ice-breaking ships in the Canadian High Arctic,'' Can. Bull. Fish. Aquatic Sci. **224**, 97–117 (1990).
[28] L. Pippard, ''Status of the St. Lawrence River population of beluga *Delphinapterus leucas*,'' Can. Field Nat. **99**(3), 438–450 (1985).
[29] D. Sergeant, ''Present status of white whales *Delphinapterus leucas* in the St. Lawrence Estuary, Nat. Can. **113**(1), 61–81 (1986).

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.

Application of a multiple-beam model for lateral vibration analysis of a discretely supported rail at high frequencies

T. X. Wu[a)]
State Key Laboratory of Vibration, Shock and Noise, Shanghai Jiao Tong University, Shanghai 200030, China

D. J. Thompson[b)]
Institute of Sound and Vibration Research, University of Southampton, Southampton S107 1BJ, United Kingdom

(Received 19 May 1999; accepted for publication 1 March 2000)

In this paper a multiple-beam model of the rail, which allows all the essential deformations caused by the lateral vibration, has been applied for studying the lateral vibration behavior related to the discrete supports. Detailed investigation of the lateral vibration of a discretely supported rail has been carried out using a Green's function matrix and the superposition principle. Good agreement between the predictions and measurement data has been reached in terms of accelerance. © *2000 Acoustical Society of America.* [S0001-4966(00)01906-8]

PACS numbers: 43.40.At [PJR]

INTRODUCTION

Recently, a multiple-beam model has been developed by Wu and Thompson[1] for studying the lateral vibration behavior of railway track. Although much simpler than a finite-element (FE) approach, this model includes all the essential cross-sectional deformations up to at least 5 kHz. The results from this model have shown a very good agreement with an FE model in terms of the dispersion relation of free waves in the frequency region 50–5000 Hz. In this paper, the application of the multiple-beam model is extended to the discretely supported rail by using a Green's function matrix and the superposition principle. This approach was developed by Heckl[2] for a single rail beam and also used by Wu and Thompson[3] for a double-beam model of vertical track dynamics. Detailed study of the lateral vibration behavior related to the discrete supports is carried out based on the multiple-beam model of Ref. 1. Predictions from the model are compared with measurements and a good agreement is reached in terms of accelerance.

I. MODELING OF DISCRETELY SUPPORTED RAIL

A. Multiple-beam model

In studying the lateral vibration of a rail at high frequencies, the cross-sectional deformation should be taken into account, in particular rail-head bending and torsion, rail-foot bending and torsion, and the relative motion between the rail head and foot, web bending. A multiple-beam model has been developed in Ref. 1 which allows all these deformations. In this model the whole rail is divided into three parts: the head and the foot are represented by two infinite Timoshenko beams with torsion, and the web is replaced by numerous beams which connect the head and foot. In this simplified model the twisting stiffness and the bending stiffness of the web in the rail-axis direction vanish. Ignoring the bending stiffness will cause no problem as it is much smaller than the lateral bending stiffnesses of the rail head and foot. The effects of the web twisting stiffness are compensated by parameter adjustment. Parameters of the multiple-beam model are determined according to the cross section of a realistic rail. Detailed information about the multiple-beam model is given in Ref. 1.

B. Discretely supported rail

A discretely supported railway track may be simplified to an infinite rail with a finite number of discrete supports. This is because the supports at large distances from the point

[a)]Currently on leave at ISVR, University of Southampton, U.K.
[b)]Author to whom correspondence should be addressed. Electronic mail: djt@isvr.soton.ac.uk

at which the response is to be calculated can be neglected due to the wave propagation decay. In order to calculate the response of the rail, the discrete supports can be replaced by a series of point forces. Based on the Green's function and the superposition principle, the stationary response of the rail to a harmonic excitation can be obtained (Refs. 2, 3).

The Green's function $g_{ij}(z,z')$ for the infinite multiple-beam rail is defined as the response of the ith component of the displacement vector $\mathbf{q}(z)$ at z caused by a unit harmonic force/moment of the jth component of the excitation $\mathbf{F}$ at z', where

$$\mathbf{q}(z)=(v_h\ \psi_h\ \theta_h\ v_f\ \psi_f\ \theta_f)^T, \tag{1}$$

where v_h, ψ_h, and θ_h represent the head lateral displacement, slope, and rotation about the rail axis, respectively; v_f, ψ_h, and θ_f are the same motions but for the foot. Subscripts 1–6 in the Green's function g_{ij} correspond to each component in $\mathbf{q}(z)$; however, it is chosen that when $j=1$ or 4, the forces act at the top of the rail head or at the bottom of the rail foot rather than at their center, respectively. The Green's functions are calculated using the multiple-beam model; see Ref. 1.

Using the Green's function $g_{ij}(z,z')$ and the superposition principle, the lateral displacement and rotation of the rail head due to a harmonic force $Fe^{i\omega t}$ at $z=z_F$ are given by

$$v_h(z)=-\sum_{n=1}^{N} Z_{tn}\left[v_f(z_n)-\frac{h_f}{2}\theta_f(z_n)\right]g_{14}(z,z_n) - \sum_{n=1}^{N} Z_{rn}\theta_f(z_n)g_{16}(z,z_n)+Fg_{11}(z,z_F), \tag{2}$$

$$\theta_h(z)=-\sum_{n=1}^{N} Z_{tn}\left[v_f(z_n)-\frac{h_f}{2}\theta_f(z_n)\right]g_{34}(z,z_n) - \sum_{n=1}^{N} Z_{rn}\theta_f(z_n)g_{36}(z,z_n)+Fg_{31}(z,z_F), \tag{3}$$

where N is the number of the supports under consideration and h_f is the height of the rail foot. Z_{tn} and Z_{rn} are the translational and rotational dynamic stiffnesses at the nth support, respectively, and can be given as

$$Z_{tn}=\frac{k_{pn}(k_{bn}-m_s\omega^2)}{k_{pn}+k_{bn}-m_s\omega^2}, \tag{4a}$$

$$Z_{rn}=k_{prn}, \tag{4b}$$

where k_{pn} and k_{prn} are the lateral and rotational stiffness of the pad, respectively, at the nth support, k_{bn} is the lateral stiffness of the ballast, and m_s is the sleeper mass. The rotational inertia of the sleeper and the rotational stiffness of the ballast are regarded as infinitely large, compared with the rotational inertia of the rail cross section and the pad rotational stiffness. Damping is considered by introducing loss factors η_p into the pad stiffnesses k_{pn} and k_{prn}, η_b into the ballast stiffness k_{bn} and making them complex with the appropriate factor $(1+i\eta)$.

Equations (2) and (3) can be combined and represented in the following compact form:

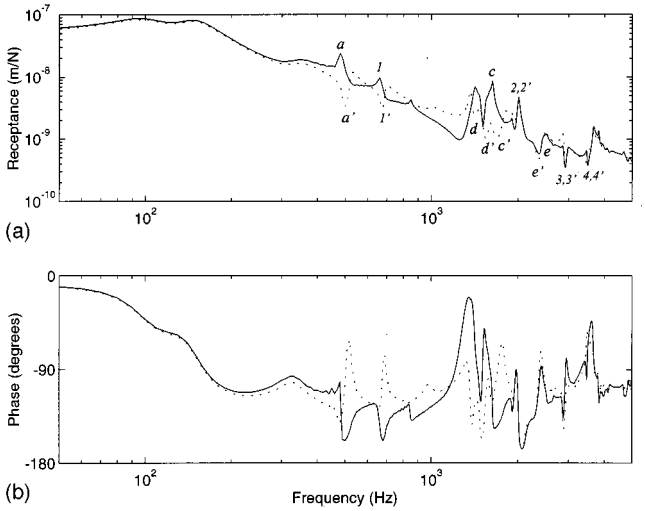


FIG. 1. Amplitude and phase of the point receptance for lateral vibration of the discretely supported rail. — For excitation acting at midspan, ··· for excitation acting above a sleeper (for meaning of symbols, see the text).

$$\mathbf{q}_h(z)=-\sum_{n=1}^{N}\mathbf{G}_{hf}(z,z_n)\mathbf{q}_f(z_n)+F\mathbf{G}_{hF}(z,z_F). \tag{5}$$

Similarly, the displacement and rotation of the rail foot are given by

$$\mathbf{q}_f(z)=-\sum_{n=1}^{N}\mathbf{G}_{ff}(z,z_n)\mathbf{q}_f(z_n)+F\mathbf{G}_{fF}(z,z_F). \tag{6}$$

At each support point $z=z_m$, the displacement and rotation of the rail foot can be written as follows:

$$\mathbf{q}_f(z_m)=-\sum_{n=1}^{N}\mathbf{G}_{ff}(z_m,z_n)\mathbf{q}_f(z_n)+F\mathbf{G}_{fF}(z_m,z_F), \quad m=1,2,...,N. \tag{7}$$

From Eq. (7) the displacement and rotation of the foot at each support point can be solved. Substituting them into Eqs. (5) and (6), one can obtain the displacement and rotation at any point on the rail head and foot.

II. RECEPTANCES OF DISCRETELY SUPPORTED RAIL

The point and cross receptances of the discretely supported rail for lateral vibration are calculated for an external excitation in the frequency range from 50 to 5000 Hz acting at the top of the rail head either above a sleeper or at midspan. For simplicity, uniform values of the pad and ballast stiffness are used at each support. A simplified version of UIC 60 rail is used in the model. The parameters for the rail and foundation used here are consistent with those in Ref. 1. Parameters for each support: $k_p=50$ MN/m, $k_b=80$ MN/m, $k_{pr}=0.654$ MNm, and $m_s=162$ kg, are from the values of track C in Ref. 4. Damping is added through loss factors: $\eta_r=0.01$ for the rail, $\eta_p=0.25$ for the pad, and $\eta_b=0.6$ for the ballast. The span length (the distance between two sleepers) could be variable in the model but is chosen as a constant value of $d=0.6$ m, and the number of supports is $N=80$.

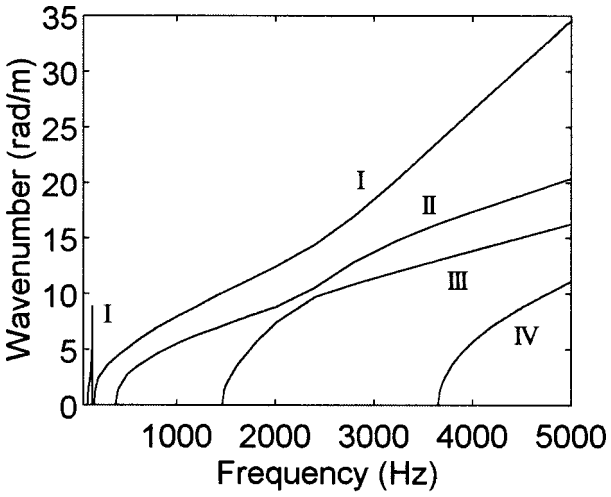


FIG. 2. Dispersion relation of waves in a continuously supported rail without damping for lateral vibration (from Ref. 1).

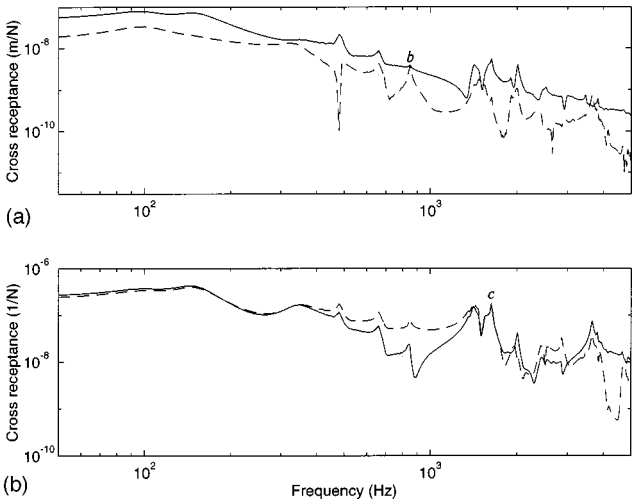


FIG. 3. Cross receptances for lateral vibration of the discretely supported rail with excitation acting at midspan at the top of the rail head. (a) Translational response, (b) rotational response. — For rail head, --- for rail foot.

A. Point receptance

The amplitude and phase of the lateral point receptance are shown in Fig. 1. To aid in interpreting the results, the dispersion relation of different waves in a continuously supported rail without damping is shown in Fig. 2, which is from Ref. 1. In Fig. 2 waves I, II, III, and IV represent the lateral bending, the torsion, the first- and second-order web bending, respectively. It can be seen from Fig. 1 that in the low-frequency region up to about 400 Hz the responses to the excitation acting both at midspan and above a sleeper are almost the same. Three well-damped resonances below 400 Hz and two sharp resonances at about 1400 and 3600 Hz can be seen in both of these curves. These resonances can be identified as the cut-on of the lateral bending wave (including its branch at 100–150 Hz), the torsional wave, and two web-bending waves as predicted in Ref. 1 by the continuously supported rail model. Two pinned–pinned resonances marked with a and a', c and c' can be seen clearly in Fig. 1. The $a-a'$ pinned–pinned resonance occurs at about 500 Hz, when the lateral bending wave length is equal to twice the span length (this corresponds to a wave number of 5.2 rad/m in Fig. 2). When the wavelength of the first-order web bending (rail head and foot rotation out of phase with each other) is equal to twice the span length, at about 1700 Hz, the $c-c'$ pinned–pinned resonance appears. The pinned–pinned resonance due to the torsional wave should occur at around 900 Hz, but it is not noticeable in Fig. 1. However, it can be identified in the cross-receptance plots below. The pinned–pinned resonance from the second-order web bending wave (rail head and foot rotation in phase with each other), which should occur at about 3900 Hz, does not appear because the high-frequency vibration of the rail is less affected by the supports. For the same reason the receptances at high frequencies are quite similar for the excitation acting either at midspan or above a sleeper.

When the wavelength is equal to the span length (this corresponds to a wave number of 10.4 rad/m in Fig. 2), the magnitude of the point receptance should have a minimum for excitation acting either at midspan or above a sleeper. For the lateral bending wave the frequency corresponding to this situation is about 1600 Hz, while for the torsional wave and the first-order web-bending wave the corresponding frequencies are about 2400 and 2600 Hz, respectively. The troughs near these frequencies can be observed in Fig. 1, and are marked with d, e, and d', e' for the excitation acting at midspan and above a sleeper, respectively. For the second web-bending wave no trough is observed because vibration at very high frequency is less affected by the supports. Unfortunately, for other peaks and troughs marked with 1–4 and $1'-4'$ no clear physical explanations can be found. They are possibly due to the superposition of different waves.

B. Cross receptance

The cross receptances are shown in Figs. 3 and 4. They are defined here as the response in the same forcing place but at some other positions or the rotational response to a force

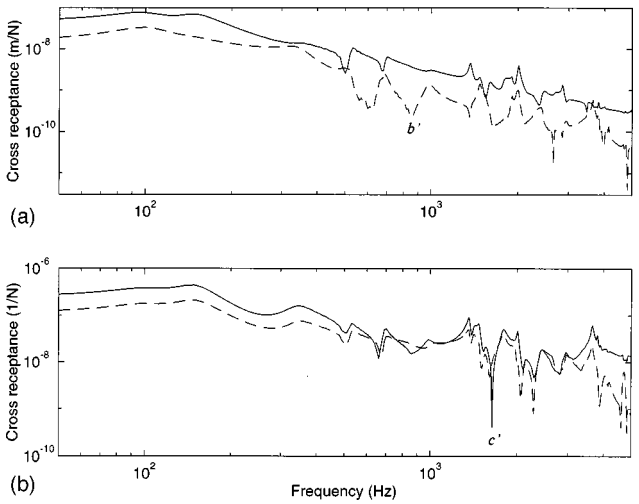


FIG. 4. Cross receptances for lateral vibration of the discretely supported rail with excitation acting above a sleeper at the top of the rail head. (a) Translational response, (b) rotational response. — For rail head, --- for rail foot.

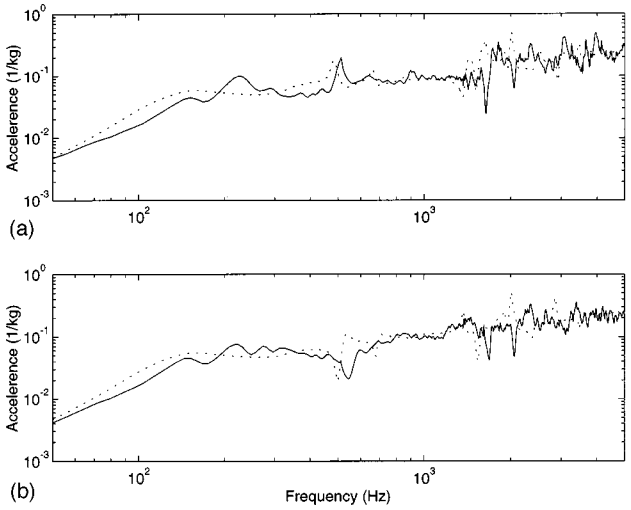


FIG. 5. Comparison of accelerances predicted from the discretely supported rail model and measured. Both excitation and response are at the center of the rail head. (a) Excitation acting at midspan, (b) excitation acting above a sleeper. — From predictions, ··· from measurements (from Ref. 4).

at the top of the head. In Figs. 3 and 4 the translational response is the lateral displacement in the center of the rail head or foot. At low frequencies both the lateral displacement and the rotation curves fluctuate smoothly. However, at high frequencies these curves fluctuate sharply, with many peaks and troughs, because of the effects of the discrete supports. The lateral displacements of the head are higher than those of the foot.

Although the cross receptances seem to be more complicated than the point receptances, some specific points can still be observed. The pinned–pinned resonance due to the torsional wave can be seen clearly at about 850 Hz in terms of the rail-foot displacements from Figs. 3(a) and 4(a), and is marked with b and b'. Another pinned–pinned resonance due to the first-order web-bending wave at about 1700 Hz can be seen much more sharply in terms of the rotations, and is marked with c and c'.

The decay of lateral vibration along the rail can be studied by means of the transfer receptance. This is not presented here because the results are similar to those from the continuously supported rail model in Ref. 1, except the pinned–pinned resonance phenomenon.

C. Comparison with experimental results

The measured and predicted lateral accelerances in the center of the rail head are compared in Fig. 5(a) and (b), which are for the excitation acting at midspan and above a sleeper, respectively. The measurement results are from Ref. 4, which was obtained from UIC 60 rail with concrete monobloc sleepers. A good agreement can be seen from Fig. 5 throughout the frequency region from 50 to 5000 Hz. The main deviation is that the response peaks in the predictions from the multiple-beam model are shifted towards lower frequencies, particularly those at about 1700 and 2000 Hz. This is mainly due to the simplification of the rail cross section. The connections between the web and head and between the web and foot of an actual rail have a tapered transition so that the real rail is stiffer than the simplified version and thus the response peaks appear at higher frequencies in the measurements.

III. CONCLUSIONS

In this paper the lateral vibration behavior of a discretely supported rail has been studied using the multiple-beam model, a Green's function matrix, and the superposition principle. Good agreement between the predictions and measurement data has been reached in terms of accelerance.

Compared with the continuously supported rail model, the predictions from the discretely supported rail model show that the track support has significant effects on the lateral vibration in the frequency region from about 400 to 3000 Hz which is of most importance for noise radiation. Many peaks or troughs of the receptance appear in this region. Among them are three pinned–pinned resonances corresponding to the bending wave, torsional wave, and the first web-bending wave, so that for the lateral vibration the pinned–pinned effects are more complicated than for vertical vibration due to the presence of multiple waves. Since these peaks reach relatively high levels, neglecting them may cause significant errors. All these suggest that a discretely supported rail model is more appropriate in this frequency range.

[1] T. X. Wu and D. J. Thompson, ''Analysis of lateral vibration behavior of railway track at high frequencies using a continuously supported multiple beam model,'' J. Acoust. Soc. Am. **106**, 1369–1376 (1999).
[2] M. A. Heckl, ''Railway noise—Can random sleeper spacing help?'' Acustica **81**, 559–564 (1995).
[3] T. X. Wu and D. J. Thompson, ''The effects of local preload on the foundation stiffness and vetical vibration of railway track,'' J. Sound Vib. **219**, 881–904 (1999).
[4] N. Vincent and D. J. Thompson, ''Track dynamic behavior at high frequencies. Part 2: Experimental results and comparisons with theory,'' Veh. Syst. Dyn. Suppl. **24**, 100–114 (1995).

Active control design for acoustic radiation using mixed-norm optimization

Robert L. Clark
Department of Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina 27708-0300

David E. Cox
NASA Langley Research Center, Hampton, Virginia 23681

(Received 24 May 1999; accepted for publication 20 April 2000)

Linear matrix inequalities were applied to design a mixed $\mathcal{H}_2/\mathcal{H}_\infty$ feedback control compensator for a structural acoustic system. The compensator was designed to minimize the $\mathcal{H}_2$ norm of radiation filters while constraining the $\mathcal{H}_\infty$ norm through the control path to be less than unity. A trade-off between minimizing sound power radiated and maximum rms gain in the control path resulted in a reliable and robust means of designing compensators for structural acoustic control. © *2000 Acoustical Society of America.* [S0001-4966(00)02609-6]

PACS numbers: 43.40.Vn [PJR]

I. INTRODUCTION

A mixed-norm design optimization was performed on a structural acoustic control problem to develop a closed-loop system that is optimal with respect to sound power radiated, while constrained to have known robustness properties. Recent developments in the use of linear matrix inequalities for control design problems were applied to find the optimal controller through convex optimization. A performance metric based upon the $\mathcal{H}_2$ norm, a measure of the rms signal energy, was selected. However, a robustness metric based upon the $\mathcal{H}_\infty$ norm, a measure of the maximum rms signal gain, was used to constrain the control signals. An experimental test structure for a transmission-loss experiment was constructed from a steel plate with piezoelectric transducers. Experimental results demonstrate that mixed-norm optimization is a viable approach in the design of compensators with robustness constraints based upon an infinity-norm and performance based on a two-norm.

II. OVERVIEW OF CONTROL THEORY

Optimal control problems involve the design of controllers that minimize a chosen norm of the system transfer function. As such, the designer has considerable freedom in defining closed-loop performance. For control of acoustic radiation, the $\mathcal{H}_2$ norm is attractive since it provides a measure of the signal energy and is physically meaningful as a control objective. However, the $\mathcal{H}_\infty$ norm provides a measure of maximum gain and can provide significant insight into the robustness of the controller. Most designs have focused on the choice of one of these norms; however, in recent years, the mixed-norm design problem has been studied.[1–4] With these methods, one can use the $\mathcal{H}_2$ norm to optimize performance and the $\mathcal{H}_\infty$ norm to enforce robustness constraints. A brief overview of robust control theory is provided in this section to help motivate the mixed-norm problem in the context of robust control design.

Consider the block diagram presented in Fig. 1. The physical system is represented by $P(s)$, with actuator inputs u, and measured outputs y, and controlled by the dynamic compensator $K(s)$. Closed-loop performance is obtained by minimizing $T_{pd}(s)$, the transfer function from disturbances, $d(s)$, to the performance signal $p(s)$, which represents acoustic radiation. For the purpose of robustness two additional signals are defined, $q(s)$ and $v(s)$, and a weighting filter, $W(s)$. This filter is used to represent an uncertainty on the input of the dynamic system, due to modeling errors or variations in actuator dynamics.

The control design objective is to find a compensator, $K(s)$, that minimizes the transfer function from disturbance to performance, and satisfies the constraint that the frequency weighted loop-gain, $T_{qv}(s)$, has an infinity norm less than one. With this constraint, stability of the plant is guaranteed for model errors on the input up to a level of $W(s)$, since by the small-gain theorem,[5]

$$\|T_{qv}(s)\Delta(s)\|_\infty < 1 \Rightarrow \text{Stability} \quad \forall \|\Delta(s)\|_\infty < 1. \tag{1}$$

Typically robust control designs seek to minimize the $\mathcal{H}_\infty$ norm of a composite transfer function, which includes performance and robustness paths, or as in the case of μ-synthesis, a scaled version of this path.[6] This yields controllers that provide both robustness and performance; however, the performance is only with respect to an $\mathcal{H}_\infty$ norm. In many systems the $\mathcal{H}_2$ norm is a more physically meaningful measure of performance. For example, in the model described here, $\|p(t)\|_2$ is the acoustic power radiated into a half-space.

Therefore, we seek a control design which minimizes radiated sound power as described by $\|T_{pd}\|_2$, but also has stability guarantees provided by the constraint $\|T_{qv}\|_\infty < 1$. Explicitly, the control problem is to find a stabilizing controller $K(s)$ such that

$$\min_{\|T_{qv}\|_\infty < 1} \|T_{pd}\|_2. \tag{2}$$

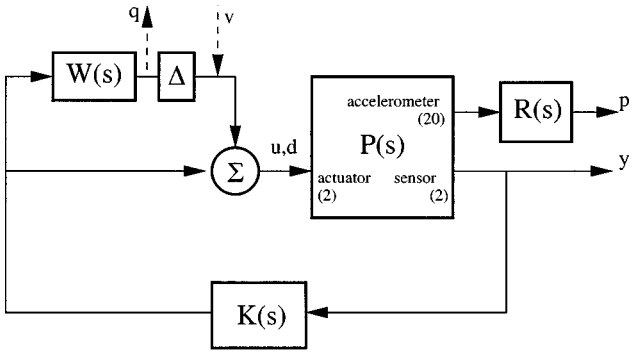


FIG. 1. Block diagram of closed-loop system with input uncertainty, $W(s)$, and radiation filters, $R(s)$.

This type of mixed-norm optimization problem can be posed as the optimization of a linear objective function subject to constraints expressed as a linear matrix inequality (LMI). Problems expressed in this fashion form convex optimization problems, and have unique solutions that can be consistently and efficiently solved using recent developments in interior point programming.[7] Details on the use of LMIs in control theory are outside the scope of this note, but can be found in Refs. 8 and 9.

III. DESCRIPTION OF THE EXPERIMENTAL DESIGN MODEL

An experimental test structure was used for the purpose of this demonstration. The test structure was machined from steel, measuring 18 in.-×16 in.-×0.1875 in. thick and mounted in a frame that approximated simply supported boundary conditions. The plate was instrumented with two 4×6 in.^{2} piezoelectric transducers that were used as both sensors and actuators and were positioned to minimize coupling of high-order modes.[10] Although not available for feedback, the system also contained 20 accelerometer measurements over an evenly spaced grid. A schematic diagram of the test structure, indicating the locations of the actuators and sensors, is given in Fig. 2.

The plant model, $P(s)$, was obtained through time-domain eigensystem realization[11] and contained 80 states, capturing a bandwidth up to 1500 Hz. The identified input paths included two piezoceramic actuators, and the identified output paths included a grid of 20 accelerometers and two piezoceramic sensors.

The radiation filters, $R(s)$, were designed using the grid of accelerometers and radiation modal expansion as outlined in Ref. 12, and applied to produce a performance output as illustrated in Fig. 1. The filters were reduced to four states to produce a measure of the first two radiation modes for the structural acoustic system. These two radiation modes were sufficient to capture 95% of the sound power over the bandwidth of interest.

Magnitudes of the frequency responses between each piezoceramic input and output for the identified system model are illustrated in Fig. 3. The colocated frequency response functions are characterized by alternating poles and zeros. The frequency response functions also provide an indication of modal participation as a function of actuator and

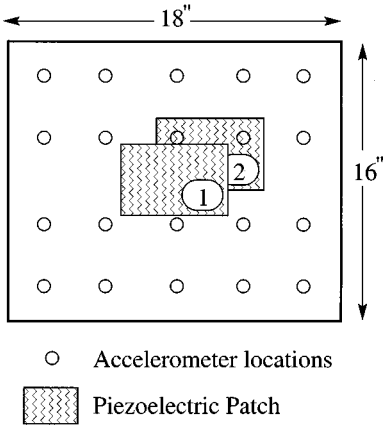


FIG. 2. Schematic diagram of test structure illustrating positions of actuators and sensors.

sensor location. The piezoelectric patch positioned on the center of the plate in Fig. 2 served as actuator and sensor 1, and the patch positioned off the central axis in Fig. 2 served as actuator and sensor 2. Actuator/sensor 1 does not couple well to the modes with even indices for the structural system studied [i.e., (1,2), (2,1), (2,2), etc.]. This choice is ''by design'' since we know that the modes with odd indices are more efficient acoustic radiators at low frequency. The second transducer was not symmetrically placed on the structure, but rather its position was optimized to improve high-frequency roll-off in the non-colocated transfer functions.[13] Additionally, we note that the frequency response between actuator 1 and sensor 2 is identical to that between actuator 2 and sensor 1 due to reciprocity.

For the purpose of this study, we chose to design a low-order dynamic compensator to achieve performance over a bandwidth up to approximately 800 Hz. A reduced-order 16-state model of the system was developed from the original 80-state model. The uncertainty model, $W(s)$, was used to bound the magnitude of the difference between the two models over the identified bandwidth. The magnitudes of the frequency response for the uncertainty model and that of the error due to model reduction are presented in Fig. 4. A high-

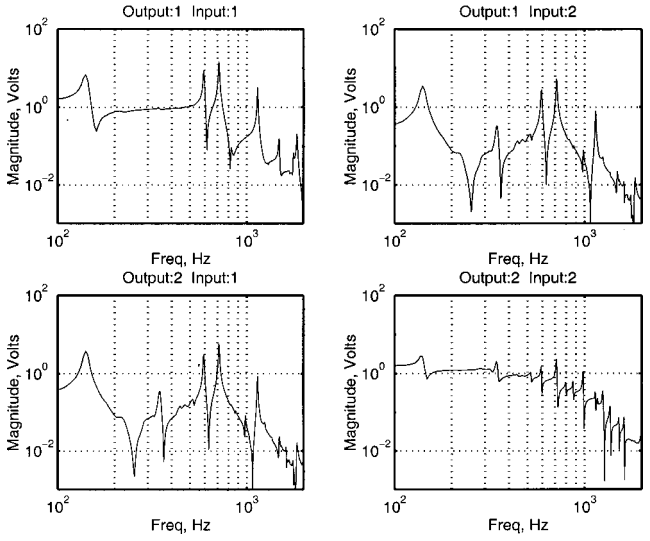


FIG. 3. Magnitude of frequency response functions between each piezoceramic actuator and sensor.

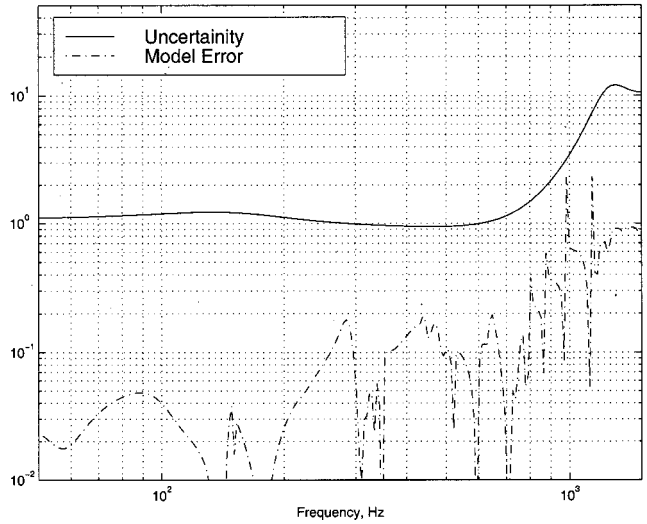


FIG. 4. Magnitudes of the frequency response for the uncertainty model used in the design model and that of the model error.

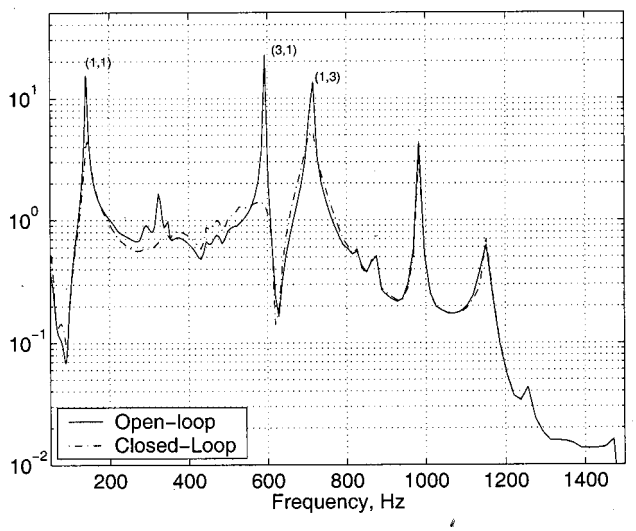


FIG. 5. Magnitude of predicted radiated sound power.

pass filter was utilized to bound uncertainty at higher frequencies and to impose a ''roll-off'' in the compensator over the bandwidth of unmodeled dynamics.

As illustrated in Fig. 4, the uncertainty model was set near 100% over the bandwidth identified for control ($<$800 Hz). One might question this high a level of uncertainty; however, the applied controls engineer must bear in mind the following:

(i) 100% input uncertainty is equivalent to stating that the maximum rms gain from the disturbance to control effort is less than unity. Here this is a practical design constraint as one would not want to produce a control signal that exceeds the original level of the disturbance input when entering through the same input path.

(ii) 100% input uncertainty provides a gain margin of 6 dB, which is a reasonable constraint for control system design and provides fault tolerance as either input can be disconnected and the system will remain stable.

(iii) The mixed norm problem only considers stability due to model errors; performance is independently optimized based on the nominal plant model. This allows for more aggressive controllers than in a H_∞ design, where stability constraints are contained in an augmented cost function.

Thus, the application engineer must take care in the interpretation of the language provided in the controls literature. While accepting a 100% uncertainty on the input might sound a bit excessive at first, upon casting this statement in terms of the fundamental physics, it poses a realistic bound for robust design.

IV. PREDICTED AND MEASURED RESULTS

The mixed-norm design problem, posed in Eq. (2), was solved using convex optimization routines in the Matlab LMI toolbox.[14] Although optimization on the augmented system took several hours to complete, solutions always converged. The identified model was augmented with analytic radiation filters to predict radiated sound power over the bandwidth of interest. The simulated open and closed loop results are presented in Fig. 5. As illustrated, approximately 10 dB of reduction results at a resonance frequency corresponding to the fundamental mode of the structure (141 Hz), and significant levels of reduction are predicted at the resonance frequencies corresponding to the (3,1) and (1,3) modes of the structure targeted by the radiation filters. The predicted maximum singular values of the constrained gain path are presented in Fig. 6. As illustrated, the uncertainty weighted gain from the disturbance input to the control effort remains below unity for the closed-loop system.

The resulting compensator was implemented on the test structure and the acoustic power for the closed-loop system was measured using the grid of 20 accelerometers. The full matrix relating the acceleration at the grid points to the radiated power was used to provide a measure of the radiated power.[12] As illustrated in Fig. 7, significant levels of reduc-

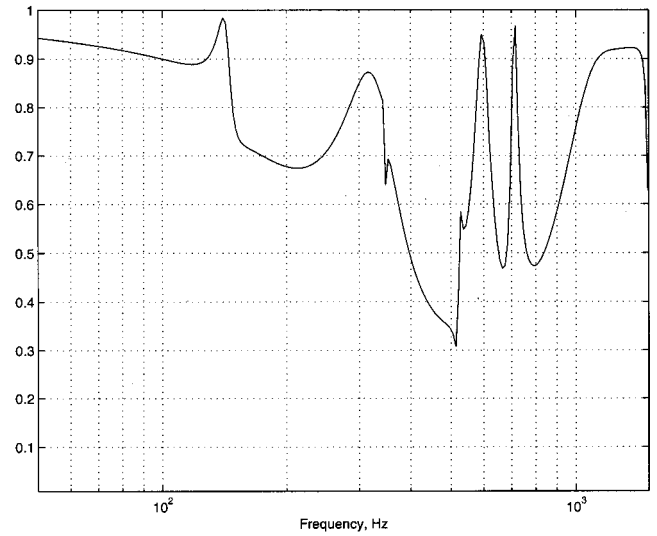


FIG. 6. Predicted maximum singular values of the constrained gain path.

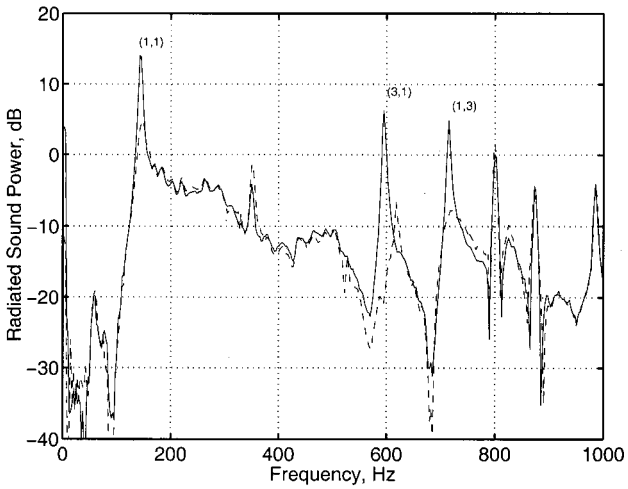


FIG. 7. Magnitude of measured sound power radiated, open and closed loop.

tion in sound power were achieved at resonance frequencies corresponding to the (1,1), (3,1), and (1,3) modes, as predicted. The total reduction in radiated power integrated over the bandwidth between 50 and 1000 Hz was 3.5 dB. Results are also presented in $\frac{1}{3}$ octave bands as illustrated in Fig. 8 to emphasize levels of reduction within specific bandwidths.

V. CONCLUSIONS

Linear matrix inequalities were used to pose and solve the mixed H_2/H_∞ feedback control problem applied to a structural acoustic system. The cost function was defined in terms of the outputs of radiation filters, and a compensator was designed to minimize the H_2 norm while constraining the H_∞ norm through the control path to be less than unity. The advantage of this mixed norm approach is that energy optimal performance does not need to be sacrificed to obtain controllers with robustness guarantees. As was discussed, constraining the control path to be insensitive to multiplicative input uncertainties of unity is equivalent to constraining the rms control signals to be no greater than that of the disturbance. Increasing the level of uncertainty further constrains the maximum rms control signal relative to the disturbance signal. While computationally time consuming, the mixed-norm design methodology provides a reliable and robust means of designing compensators for the experimental example provided.

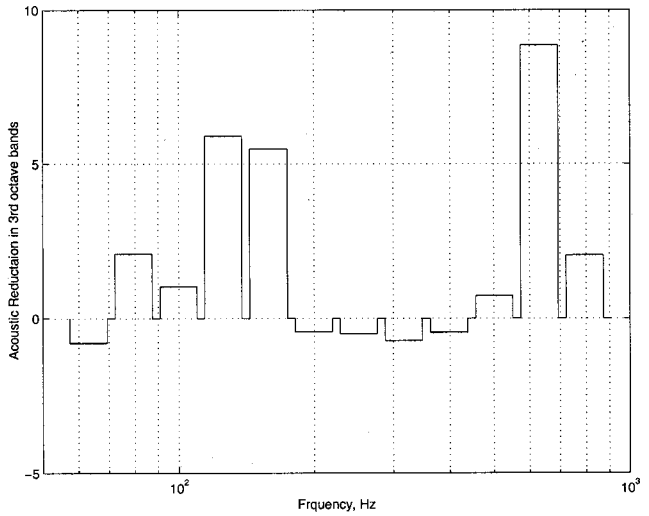


FIG. 8. Reduction in sound power radiated as a function of $\frac{1}{3}$ octave bands.

ACKNOWLEDGMENTS

The authors would like to gratefully acknowledge the support of the National Science Foundation under the CAREER Program Award: CMS-95-01470.

[1] D. S. Bernstein and W. M. Haddad, ''LQG Control with an H_∞ Performance Bound: A Riccati Equation Approach,'' IEEE Trans. Autom. Control **34**(3), 293–305 (1989).
[2] K. Zhou, K. Glover, B. Bodenheimer, and J. Doyle, ''Mixed H_2 and H_∞ Performance Objectives II: Optimal Control,'' IEEE Trans. Autom. Control **39**(8), 1575–1587 (1994).
[3] P. P. Khargonekar and M. A. Rotea, ''Mixed H_2/H_∞ Control: A Convex Optimization Approach,'' IEEE Trans. Autom. Control **36**(7), 824–837 (1991).
[4] C. Scherer, P. Gahinet, and M. Chilali, ''Multiobjective Output-Feedback Control via LMI Optimization,'' IEEE Trans. Autom. Control **42**(7), 896–911 (1997).
[5] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control* (Prentice-Hall, Upper Saddle River, NJ, 1995).
[6] A. Packard and J. Doyle, ''The Complex Structured Singular Value,'' Automatica **29**(1), 71–109 (1993).
[7] Y. Nesterov and A. Nemirovsky, *Interior-point Polynomial Methods in Convex Programming*, Vol. 13 of Studies in Applied Mathematics (SIAM, Philadelphia, PA, 1994).
[8] S. Boyd, V. Balakrishnan, E. Feron, and L. ElGhaoui, ''Control System Analysis and Synthesis via Linear Matrix Inequalities,'' in *Proceedings of the American Control Conference*, 1993, pp. 2147–2154.
[9] L. E. Ghaoui and S.-I. Niculescu, editors, *Advances in Linear Matrix Inequality Methods in Control* (SIAM, Philadelphia, 2000).
[10] J. S. Vipperman and R. L. Clark, ''Implications of using colocated strain-based transducers for active structural acoustic control,'' J. Acoust. Soc. Am. **106**, 1392–1400 (1999).
[11] J. N. Juang and R. S. Pappa, ''An Eigensystem Realization Algorithm for Modal Parameter Identification and Model Reduction,'' J. Guid. Control Dyn. **8**(5), 620–627 (1985).
[12] G. P. Gibbs, R. L. Clark, D. E. Cox, and J. S. Vipperman, ''Radiation modal expansion: Application to active structural acoustic control,'' J. Acoust. Soc. Am. **107**, 332–339 (2000).
[13] R. L. Clark and D. E. Cox, ''Band-Limited Actuator and Sensor Selection of Disturbance Rejection: Application to Structural Acoustic Control,'' J. Guid. Control Dyn. **22**(5), 740–743 (1999).
[14] P. Gahinet, A. Nemirovski, A. Laub, and M. Chilali, *LMI Control Toolbox Users Guide* (The Mathworks Inc., Natick, MA, 1995).

Finding the right cross-correlation peak for locating sounds in multipath environments with a fourth-moment function

John L. Spiesberger
Department of Earth and Environmental Science, 240 South 33rd Street, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6316

(Received 11 March 1999; revised 3 April 2000; accepted 23 May 2000)

To locate calling animals in reverberant environments from recordings on widely separated receivers, a fourth-moment ''Augmented-Template Correlation Function'' (ATCF) helps identify which of many peaks in each cross-correlation function is that corresponding to the difference in travel times for the first arrivals (reference-lag). This peak may not be the largest. The ATCF, by providing an approximate correlation between auto- and cross-correlation functions, can be orders of magnitude more efficient in selecting the reference-lag than the alternative of randomly selecting peaks. The ATCF's efficacy increases with the number of paths and their signal-to-noise ratios. © *2000 Acoustical Society of America.* [S0001-4966(00)00909-7]

PACS numbers: 43.60.Cg, 43.60.Gk [JCB]

I. INTRODUCTION

A method was recently introduced for locating calling animals recorded on widely separated receivers.[1,2] The wide-bandwidth calls traveled to the receivers along multipath. In the rural environments in which this method was applied, locations were not obtained by fitting a forward model to the data[3–9] because multipath reflected from obstacles whose locations were unknown, such as uneven topography, trees, rocks, and houses. The only paths that could be used to locate animals were the ones which arrived first because they were nearly straight.[1] The cross-correlation function contained many contending peaks for the desired lag, called a reference-lag, corresponding to the difference in travel times of the first arrivals. One of the methods used to narrow down the number of candidates for the reference-lag was based on a fourth-moment function called an ''Augmented-Template Correlation Function'' (ATCF). The purpose of this paper is to demonstrate that the ATCF is more effective in selecting the reference-lag than is possible by choosing peaks at random. This is the first quantitative assessment of this function.

The ATCF also plays a role in the blind solution of the impulse response of the channel at each receiver.[2,10] Blindly estimating the impulse response of a channel has applications in acoustic and electromagnetic communication systems, as well as tomographic imaging systems of environmental fields.[1,10,11]

II. AUGMENTED-TEMPLATE CORRELATION FUNCTIONS

The ATCFs are defined as

$$F_{ij}(T) \equiv \int_0^{\tau_{\max}} \breve{R}_{ii}(\tau)\breve{R}_{ij}(\tau - T)\,d\tau, \tag{1}$$

$$B_{ij}(T) \equiv \int_0^{\tau_{\max}} \breve{R}_{jj}(\tau)\breve{R}_{ji}(\tau - T)\,d\tau, \tag{2}$$

where the peak-selection function is[2]

$$\breve{R}_{ij}(\tau) \equiv \begin{cases} 0; & \text{if there is no peak at lag } \tau \text{ in } R_{ij}(\tau) \\ 1; & \text{if there is a peak at lag } \tau \text{ in } R_{ij}(\tau), \end{cases} \tag{3}$$

and the auto- and cross-correlation functions are

$$R_{ij}(p) \equiv \frac{1}{K} \sum_{k=1+p}^{K} r_i(k) r_j(k-p); \quad p \geqslant 0; \quad 1 \leqslant i < j \leqslant \mathcal{R}. \tag{4}$$

For $i \neq j$, the indices on k go from 1 to $K+p$ for p less than zero. The kth time sample at receiver i is $r_i(k)$. There are $\mathcal{R}$ receivers. For auto-correlation functions, $i=j$. The effectiveness of ATCFs comes from the fact that the relative travel times of paths at any receiver appear as similar patterns in the auto- and cross-correlation functions. When the auto- and cross-correlation functions are lined up properly, the ATCF tends toward large values. A complete discussion of this function appears on pp. 304–305 and Fig. 2 of Ref. 2.

An objective function used to estimate the efficacy of ATCFs in selecting the reference-lag is

$$O_{ij}(T) \equiv F_{ij}(-T) + B_{ij}(T), \tag{5}$$

where candidates for the reference-lag, $T = T_{FB_{ij}}$, are taken from the intersections of three sets,

$$\begin{aligned} T_{FB_{ij}} \in & \{T_{F_{ij}}(1), T_{F_{ij}}(2), \ldots T_{F_{ij}}(P_{F_{ij}})\} \\ & \cap \{T_{B_{ij}}(1), T_{B_{ij}}(2), \ldots T_{B_{ij}}(P_{B_{ij}})\} \\ & \cap \{\tau_{ij}(1), \tau_{ij}(2), \ldots \tau_{ij}(P_{\text{data}})\}. \end{aligned} \tag{6}$$

The number of candidate reference-lags from the F- and B-ATCFs are $P_{F_{ij}}$ and $P_{B_{ij}}$, respectively. The third set contains the P_{data} peaks occurring in the cross-correlation function $i-j$. The intersection with the third set guarantees that candidate reference-lags are taken from the set of picked peaks in the cross-correlation function. Thus, the reference-lag cannot be found in this study unless it is picked in the cross-correlation function. Since ATCFs tend to have large values when T is the reference-lag,[2] ranked candidates for reference-lags, $T_{FB_{ij}}$, are taken by ordering values of

$O_{ij}(T_{FB_{ij}})$ in monotonically decreasing order.

The peak-selection function utilizes peaks that exceed a certain signal-to-noise ratio. In this paper, a statistical peak-selection function is introduced which suppresses smaller peaks that are more likely due to noise than signal. Once the superiority of the statistical peak-selection function is established, further exploration of the efficacy of ATCFs will be based only on the statistical peak-selection function. The efficacy of the ATCFs will be investigated as a function of the number of paths and their signal-to-noise ratios.

III. STATISTICAL PEAK-SELECTION FUNCTION

Assume there are K data at receiver i. For simplicity, assume that the greatest lag p considered in the cross-correlation function, $R_{ij}(p)$, obeys $|p| \ll K$. Assuming noise samples at the receiver are mutually uncorrelated, then $R_{ij}(p)$ is, asymptotically with K for $|p| \ll K$, a mutually uncorrelated Gaussian random variable with mean zero and variance σ_{ij}^2 except for lag zero of the auto-correlation function which is not treated here. The Gaussian nature of the correlation values is a result of the central limit theorem.[12] The variance, σ_{ij}^2, is taken to be a constant, which is a good approximation when $p \ll K$. So, the probability density function of $R_{ij}(p)$ is

$$f_{R_{ij}}(R_{ij}(p)) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} \exp\left[-\frac{R_{ij}^2(p)}{2\sigma_{ij}^2} \right]; \quad p > 0. \tag{7}$$

One of many possible statistical peak-selection functions is

$$\check{R}_{ij}(p) = 1 - \exp\left[-\frac{R_{ij}^2(p)}{2\sigma_{ij}^2} \right], \quad \text{if } R_{ij}(p) \geqslant R_0, \tag{8}$$

and is zero if $R_{ij}(p) < R_0$. The peak-selection function has values from zero, for small peaks, to one, for big peaks.

In practical situations, the noise and reverberation may be correlated. The results in this paper may not be sensitive to the amount of correlation because the variance of the noise and reverberation, σ_{ij}^2, can be empirically estimated as a function of lag.

R_0 is chosen by specifying the probability that selected peaks are due to noise. Since $R_{ij}(p)$ is a random variable, so is $\check{R}_{ij}(p)$. The probability that a noisy peak exceeds R_0 is

$$P(\check{R}_{ij}(p) \geqslant R_0) = 2 \operatorname{erfc}(\sqrt{-2\ln(1-R_0)}), \tag{9}$$

where the complementary error function is

$$\operatorname{erfc}(x) \equiv \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-t^2/2)\, dt.$$

Given $P(\check{R}_{ij}(p) \geqslant R_0)$, R_0 is found numerically.

IV. CANDIDATE REFERENCE-LAGS CHOSEN AT RANDOM

The probability of randomly selecting the reference-lags among the candidates at $\mathcal{R}$ receivers is computed next. The number of cross-correlation functions that can be formed from $\mathcal{R}$ receivers is $\mathcal{R}(\mathcal{R}-1)/2$. For the cross-correlation between receivers i and j, suppose there are n_{ij} candidates for the reference-lag and one of these is correct. All physically possible reference-lags must have values between $-d_{ij}/c$ and d_{ij}/c, where d_{ij} is the distance between the receivers and c is the slowest speed at which sound propagates. If a peak is picked at random without replacement from each cross-correlation function, what is the probability, $p_f(f)$, that all the reference-lags are picked after making f random selections from each of the $\mathcal{R}(\mathcal{R}-1)/2$ cross-correlation functions? For example, $p_f(f=1)$ is the probability that all $\mathcal{R}(\mathcal{R}-1)/2$ reference-lags are chosen from the first selection.

To find this probability function, suppose for example there are three receivers. Then,

$$p_f(f=1) = \frac{1}{n_{12}} \frac{1}{n_{13}} \frac{1}{n_{23}}.$$

The probability that the reference-lags are picked among the first two selections is

$$p_f(f=2) = \frac{\min(2,n_{12})}{n_{12}} \frac{\min(2,n_{13})}{n_{13}} \frac{\min(2,n_{23})}{n_{23}},$$

where $\min(a,b)$ is the minimum value of a and b. So the general formula for $\mathcal{R}$ receivers is

$$p_f(f) = \prod_{i=1}^{j-1} \prod_{j=2}^{\mathcal{R}} \frac{\min(f,n_{ij})}{n_{ij}}. \tag{10}$$

Monte Carlo simulations will be used to measure the efficacy of ATCFs in comparison with this probability. So there is a need to define an average probability of $p_f(f)$ over many simulations,

$$\overline{p_f(f)} \equiv \frac{1}{Q(f)} \sum_{q=1}^{Q(f)} p_{f,q}(f), \tag{11}$$

where $p_{f,q}(f)$ is $p_f(f)$ for simulation q and $Q(f)$ is the number of simulations which have f selections.

V. SIMULATIONS

Three receivers are located on the circumference of a circle of radius 15 m. A source is placed at its center. Each receiver picks up five paths with the difference in arrival time between the last and first being 0.4 s. Arrival times are selected using a uniform random number generator. The speed of sound is 330 m/s. The signal-to-noise ratio of the first path correlated with itself, R_{ii}, in each auto-correlation function is set to 15 dB, i.e.,

$$15\,(\text{dB}) = 10 \log_{10} \frac{R_{ii}^2}{\sigma_{ii}^2}. \tag{12}$$

The amplitudes of the remaining paths decay following a spherical spreading of energy. The signal-to-noise ratio is set so that there is a high probability that the first arrivals will be picked in the cross-correlation functions, but not so high so that the later arrivals will all be picked. The value for R_0 is chosen so that there is a 0.2 probability that a selected peak is due to noise only. This corresponds to picking peaks with signal-to-noise ratios exceeding 2.2 dB. The correct reference lag appears among each set of picked peaks from each

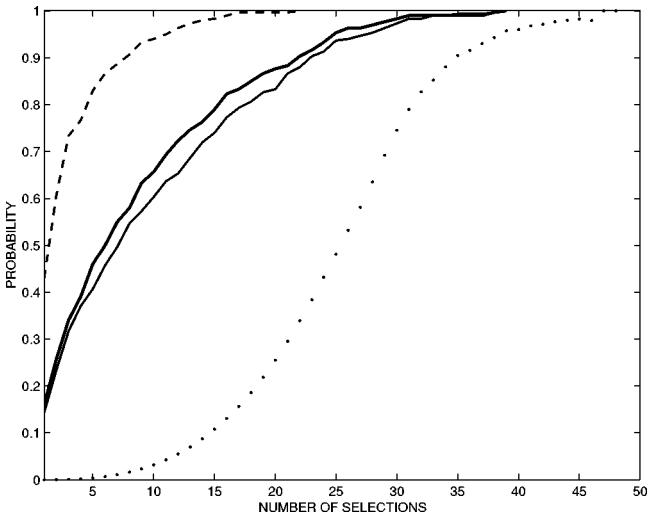


FIG. 1. Probability of choosing all three correct reference-lags as a function of the number of times one lag has been selected, without replacement, from each of the three pools of candidate reference-lags from each of the cross-correlation functions. The dotted line refers to random selection. The probability of obtaining the correct reference-lags using the Augmented-Template Correlation Functions (ATCFs) is shown as a function of the indicated number of selections using the statistical (thick line) and nonstatistical (thin line) peak-selection functions. Selections from the ATCFs are taken from the set $T_{FB_{ij}}$ [Eq. (6)] in the order of decreasing size of $O_{ij}(T_{FB_{ij}})$. The dashed line shows the efficacy of the ATCFs when the statistical peak-selection function is used with simulations having higher signal-to-noise ratios.

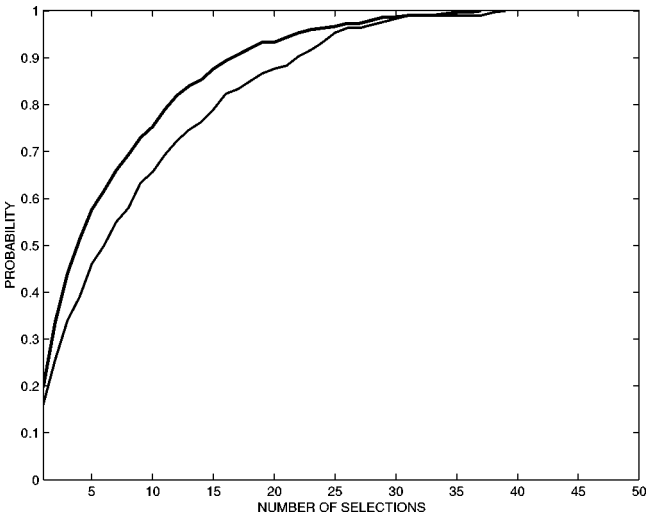


FIG. 2. Probability of selecting all three correct reference-lags using the ATCFs as in Fig. 1 as a function of the number of selections made from the candidate pools of reference-lags from the three cross-correlation functions. The thick and thin lines are based on simulations using ten and five paths, respectively, at each receiver.

cross-correlation function. Each simulation below uses 100 Monte Carlo runs.

ATCFs based on the statistical peak-selection function yield higher probabilities of selecting the reference-lags than the nonstatistical peak-selection function (thick and thin lines, Fig. 1).

The average number of candidate reference-lags found at each receiver for the statistically-based peak-selection function is 33. The reference-lags are found with much higher probabilities from the ATCFs than by random selection (thick and dotted lines, Fig. 1). For example, when the probability of selecting all three reference-lags using ATCFs is 0.8, the probability of randomly selecting all three reference lags is 0.1. The gain in probability is $0.8/0.1=8$.

The efficacy of ATCFs increases with the signal-to-noise ratio. For example, when the signal-to-noise ratio in Eq. (12) is raised to 40 dB, the probabilities of selecting the reference-lags increase significantly (dashed and thick solid lines, Fig. 1). When the probability of selecting all three reference-lags using ATCFs is 0.8, the probability of randomly selecting all three reference-lags is 0.004. The corresponding gain in probability is $0.8/0.004=200$.

ATCFs do better at larger signal-to-noise ratios because there are more signal-related peaks in the correlation functions, offering a more complete template of signal-related patterns to be used by the ATCFs. In fact, the average number of signal-related peaks that are picked from three auto- and three cross-correlation functions is 36 for the 15 dB simulation and 93 for the 40 dB simulation. The total number of signal-related lags in all the correlation functions is about 105, so less than half of these are picked for the 15 dB case. The improved efficacy of ATCFs is unrelated to the number of noisy peaks, because both the 15 and 40 dB simulations contain the same average number of noisy peaks, about 760, in all the correlation functions.

Reference 2 hypothesized that the efficacy of ATCFs should increase with the number of paths at each receiver. With more paths, the probability that noise could look like the more complicated templates of signal-related lags should decrease. Indeed, when the number of paths at each receiver is increased from five to ten, the probability of selecting all the reference-lags increases (Fig. 2).

VI. CONCLUSIONS

The difference in the travel times of the first arrivals at two receivers is useful for locating a naturally occurring sound such as a calling animal when the corresponding paths are known to be nearly straight.[1] Cross-correlation functions may not tell us what this difference is when there are many peaks of similar amplitudes due to multipath. A quantitative assessment of a fourth-moment function, called the Augmented-Template Correlation Function (ATCF), shows its effectiveness in identifying the desired difference compared to alternative methods. For example, simulations show that the ATCF is able to identify the difference in travel time with probabilities that are between 8 and 200 times that found by randomly selecting candidates for the reference-lags from the peaks in the cross-correlation functions. The performance of the ATCFs improves with the signal-to-noise ratio and the number of paths. These results are the first quantitative measure of the efficacy of this function.

It is important to emphasize that the ATCF alone may not provide a reliable enough estimate for the desired difference called the ''reference-lag.'' It is part of the algorithm used to blindly estimate the reference-lag. The reliability of the algorithm in simulations and field experiments comes also from the use of the lag and amplitude equations which link the auto- and cross-correlation functions.[1,2] The ATCF

is incorporated in two ways. First, it helps reduce the computational burden of that algorithm by providing guidance for reference-lag candidates. Second, it is an objective function used for selecting reference-lags.

ACKNOWLEDGMENTS

This work was supported by ONR Grant No. N00014-97-1-0613. During the review process of Ref. 2, an anonymous reviewer mentioned the possibility of letting the peak-selection function take on a continuum of values due to the presence of noise. This idea motivated the author to explore this idea. I thank the reviewers for the comments.

[1] J. L. Spiesberger, ''Locating animals from their sounds and tomography of the atmosphere: Experimental demonstration,'' J. Acoust. Soc. Am. **106**, 837–846 (1999).
[2] J. L. Spiesberger, ''Linking auto- and cross-correlation functions with correlation equations: Application to estimating the relative travel times and amplitudes of multipath,'' J. Acoust. Soc. Am. **104**, 300–312 (1998).
[3] W. S. Burdic, *Underwater Acoustic System Analysis*, 2nd ed. (Prentice-Hall, Englewood Cliffs, NJ, 1991).
[4] 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).
[5] M. D. Collins and W. A. Kuperman, ''Focalization: Environmental focusing and source localization,'' J. Acoust. Soc. Am. **90**, 1410–1422 (1991).
[6] J. L. Krolik, ''Matched-field minimum variance beamforming in a random ocean channel,'' J. Acoust. Soc. Am. **92**, 1408–1419 (1992).
[7] S. Li and P. M. Schultheiss, ''Depth measurement of remote sources using multipath propagation,'' IEEE J. Ocean Eng. **18**, 379–387 (1993).
[8] S. Narasimhan and J. L. Krolik, ''Fundamental limits on acoustic source range estimation performance in uncertain ocean channels,'' J. Acoust. Soc. Am. **97**, 215–226 (1995).
[9] A. Kristensen, A. Caiti, F. Ingenito, M. D. Max, J. M. Berkson, M. D. Collins, L. T. Fialkowski, N. C. Makris, E. McDonald, J. S. Perkins, and W. A. Kuperman, ''Environmental confirmation of focalization,'' in *Full Field Inversion Methods in Ocean and Seismo-Acoustics*, edited by O. Diachok *et al.* (Kluwer Academic, Dordrecht, 1995), pp. 153–158.
[10] ''Special issue on blind systems identification and estimation,'' Proc. IEEE **86**, 1903–2089 (1998).
[11] J. L. Spiesberger and K. M. Fristrup, ''Passive localization of calling animals and sensing of their acoustic environment using acoustic tomography,'' Am. Nat. **135**, 107–153 (1990).
[12] R. J. Serfling, *Approximation Theorems of Mathematical Statistics* (Wiley, New York, 1980).

A method for individual identification of echolocation signals in free-ranging finless porpoises carrying data loggers

Tomonari Akamatsu
National Research Institute of Fisheries Engineering, Ebidai, Hasaki, Kashima, Ibaraki 314-0421, Japan

Ding Wang and Kexiong Wang
Institute of Hydrobiology, The Chinese Academy of Sciences, Wuhan 430072, People's Republic of China

Yasuhiko Naito
National Institute of Polar Research, 1-9-10 Kaga, Itabashi, Tokyo 173-8515, Japan

(Received 1 June 1999; accepted for publication 5 June 2000)

Echolocation click events of a free-ranging juvenile and an adult finless porpoise (*Neophocaena phocaenoides*) were recorded with an acoustic data logger. Additionally, dive depth and swim speed of the juvenile were recorded with a behavior data logger. Echoes of echolocation signals from the water surface were clearly detected in shallow dives approximately less than 2 m. The delay time between a surface echo and a direct signal corresponded with the two-way transmission time for the animal's depth, indicating that the signals originated from the animal wearing the data loggers. The finless porpoises produced echolocation signals frequently and were thought to be able to detect their depth by listening to echoes from the water surface. © *2000 Acoustical Society of America.* [S0001-4966(00)01609-X]

PACS numbers: 43.80.Ka, 43.80.Jz, 43.66.Gf [WA]

INTRODUCTION

The individual identification of echolocation signals in cetaceans is desirable for observation of sonar behavior. A hydrophone array (Au *et al.*, 1998), an underwater mobile video/acoustic system (Dudzinski and Newborough, 1997), and acoustic data loggers (Tyack and Recchia, 1991; Parks *et al.*, 1999; Blomqvist and Amundin, 1999) were applied for the recording of dolphins' signals. However, none of the above studies was able to continuously identify the sound production of individuals for extended periods in the wild because dolphins typically move quickly away from the observation system and hard-to-exclude sounds from nearby conspecifics.

Dolphins have highly developed sonar abilities and have been studied extensively in captivity (Au, 1993). The use of sonar by dolphins and porpoises in the wild (Akamatsu *et al.*, 1998) is thought to be related to the underwater navigation and searching behavior. Here, we report a new and simple method to archive continuous click events individually.

I. MATERIALS AND METHODS

A. Study area and animals: Finless porpoises

Finless porpoises (*Neophocaena phocaenoides*) living in the Tian-e-zhou Baiji Nature Reserve (referred to as a seminatural reserve), in Shishou, Hubei, China, were caught with a seine net and transported to a fine-mesh net enclosure at an inlet in the reserve. The seminatural reserve was established for the protection of baiji (*Lipotes vexillifer*) and finless porpoises by the Chinese government. It was an old course of the Yangtze River (21 km long, 1 to 2 km wide, 20 m maximum depth). A juvenile female #1 and an adult male #2 (Table I) were equipped with acoustic and behavior data loggers (Fig. 1). When they were released, no other animals were seen around the release site.

B. Data loggers

The acoustic data logger consisted of a hydrophone (System Giken Co.), a preamplifier, a 10-kHz high-pass filter, a peak-level holding circuit, and an Integrated Circuit recorder (Sony ICD-80 with a sampling frequency of 4.8 kHz). The hydrophone had flat frequency response from 10 to 200 kHz within ± 5 dB with sensitivity of -220 dB/1V peak-to-peak *re* 1 μPa. The peak-level holding circuit kept the maximum voltage level of the signal for 0.5 ms, allowing the recording of the click events by the low sampling frequency system. During silent periods that lasted for more than 1 s, the ICD-80 recorder was automatically switched into the pause mode to allow for a prolonged data acquisition.

The peak-to-peak trigger level was 150 dB *re* 1 μPa. This trigger level was determined from experimentation with a captive finless porpoise using the identical data logger. The trigger-level calibration was done in a different test tank by projecting 100-kHz and 10-cycle tone bursts. A pressure-resistant rectangular aluminum case ($125\times73\times21$ mm, resists up to 300 m in depth) housed the entire electronic circuit.

An 8.0-cm-diameter rubber suction cup (Canadian Tire Corp.) was used to attach the system to the upper posterior side of the pectoral fin of a finless porpoise (Fig. 1). The suction cup attachment (Baird, 1998) of the present loggers was tested using captive finless porpoises, and both types of loggers remained attached for 8 h. The hydrophone was located approximately 30 cm behind the blowhole. This data logger was suited with flotation material (foam polystyrene)

TABLE I. Summary of dive behavior and sonar signal production of two finless porpoises.

Animal code of finless porpoises	Juvenile #1	Adult #2
Sex	Female	Male
Body length (m)	1.06	1.30
Captured	13 Dec. 1998	13 Dec. 1998
Released	15 Dec. 1998	16 Dec. 1998
Acoustic logger retrieval	16 Dec. 1998	17 Dec. 1998
Recording time (min.)	42.3	53.4
Behavioral logger retrieval	16 Dec. 1998	No
Recording time (min.)	1236	NA
Number of dives	199	117[a]
Average dive time (s)	11.6	26.3[a]
Maximum dive time (s)	40	141[a]
Average dive depth (m)	1.7	NA
Maximum dive depth (m)	11.5	NA
Average swim speed (m/s)	1.5	NA
Maximum swim speed (m/s)	2.9	NA
Number of click trains	615	1021
Number of click trains/min.	14.5	19.1
Average click train duration (s)	1.45	1.39
Dives with surface echo (%)	95	NA

[a]Acoustically determined from 53.4-min. recording.

and a VHF transmitter. When the logger came off, it floated on the water surface, and its location was determined by homing in on its VHF transmitter [Advanced Telemetry Systems Inc., USA, model (2)10-12(6V)].

The behavior data logger (Little Leonard, KS-PDT, 32 mm in diameter and 110 mm in length) was attached with an identical suction cup at the same position as the acoustic data logger but on the opposite side of the porpoise. This data logger was already applied to the observation of prey pursuit and capture behavior of king penguins (Ropert-Coudert *et al.*, 2000). The finless porpoise's dive depth (0.1-m resolution) and swim speed (0.05-m/s resolution) were recorded once every second. The body angle (0 deg for the horizontal direction) could be calculated by two velocity vectors, i.e., the swim speed and vertical depth change per second.

Interclick intervals were acquired with the analysis system described in Akamatsu *et al.* (1998). The end of a click train was defined as a period of no echolocation signals lasting more than 200 ms.

II. RESULTS

Both the acoustic and behavior data loggers of the juvenile finless porpoise #1 and the acoustic data logger of the adult #2 were retrieved the day after release (Table I). The acoustic data loggers recorded sonar signals from the juvenile #1 and the adult #2 for 42 min. 19 s and 53 min. 25 s, respectively. These recording durations included periods in which the recorder was in the pause mode. The behavior data logger on the juvenile #1 recorded dive depth and swim speed for 20 h, 36 min. In terms of the behavior data of #1, the number of dives, average and maximum values of dive time, dive depth, and swim speed during the simultaneous recording of the acoustic data (42 min. 19 s) were presented in Table I. The data loggers were seen attached in the correct position on the animals several hours after their release.

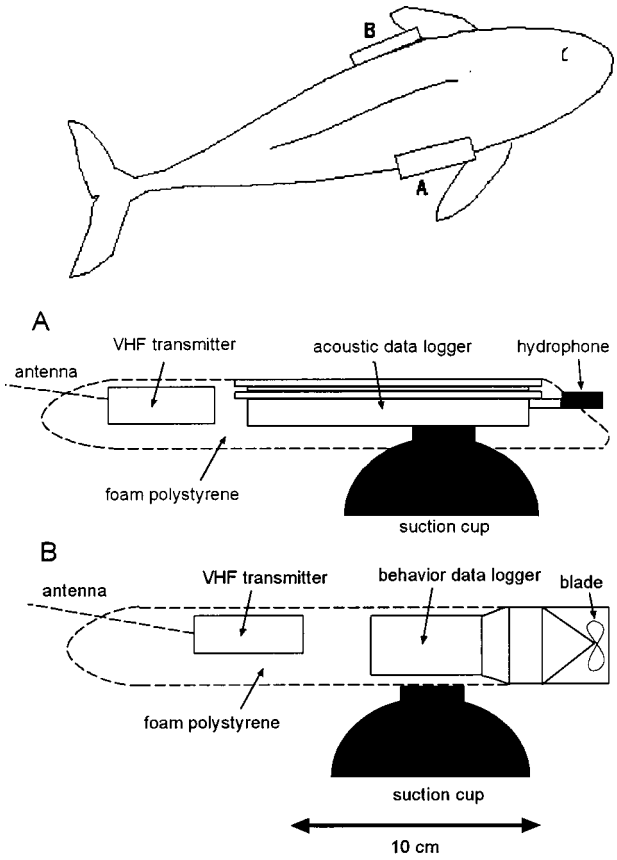


FIG. 1. A finless porpoise with acoustic (A) and behavior (B) data logger attached by a suction cup. A hydrophone picked up the sonar signals of the porpoise. The click events were recorded by an Integrated Circuit recorder housed in an aluminum case. Both data loggers were also suited with flotation material (foam polystyrene) and a VHF transmitter.

Surfacing time was defined as the time when the recorded depth was less than 0.5 m from the surface. Dive duration was determined by the interval between surfacing times. To exclude short successive surfacing as a dive, intervals less than 5 s were not used for the analysis. In total, 199 dives by the juvenile #1 were documented. The surface time was also recognizable by the water splash noise on the acoustic data logger recordings as shown at the top of Fig. 2. The 92% of surfacing time recorded by the behavior data logger of juvenile #1 was associated with the splash noise. Therefore, the surfacing and dive times of adult #2 could be recognized even without the behavioral data logger (Table I).

The acoustic data logger recorded 615 and 1021 click trains of #1 and #2, respectively (Table I). The finless porpoise produced click trains frequently, an average of once every 3 to 4 s.

Water surface echoes of the sonar signals were clearly detected by the acoustic data logger of the juvenile #1 (Fig. 2, middle). The lower part of Fig. 2 shows the simultaneous dive profiles. The delay time of the echo corresponded with the two-way sound transmission time between the recorded dive depth and the water surface (dashed line in the middle of Fig. 2). In total, 95% of the dives of #1 were associated with surface echoes (Table I). In the other dive sample of #1, the surface echo disappeared between 2.5 to 5 s (Fig. 3). At these times, the depth of the animal was 1.6 and 2.5 m, and

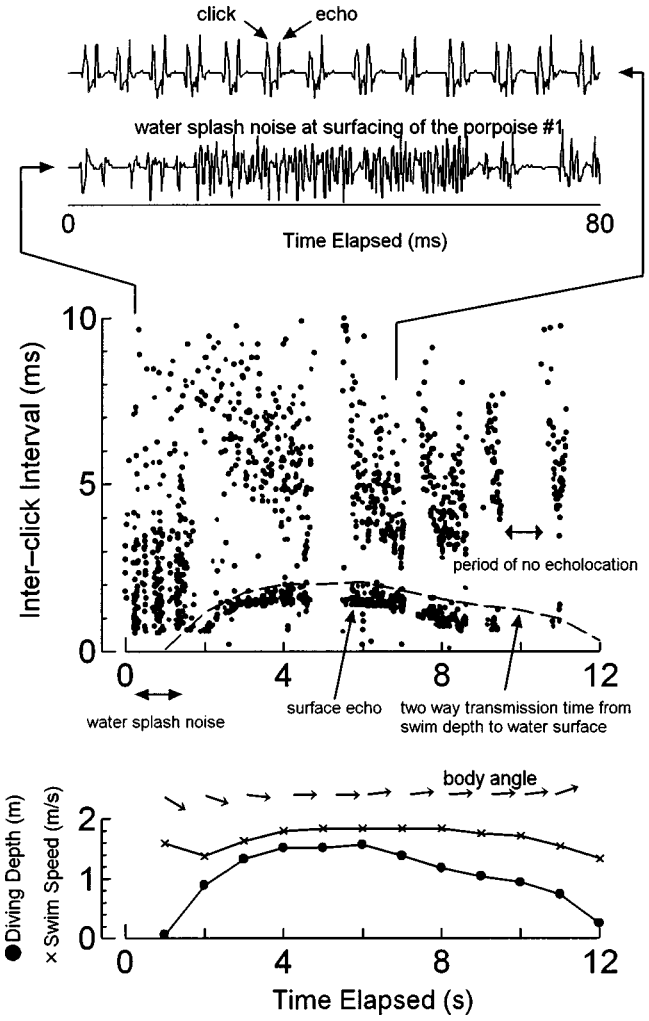


FIG. 2. Interclick intervals, dive depth, swim speed, and body angle of a free-ranging juvenile finless porpoise #1 during a 12-s dive. The waveform of peak-hold processed clicks and the water splash noise during surfacing were clearly different from each other as shown at the top. Interclick intervals less than 2 ms apart indicate surface echoes. The delay time corresponded with the two-way transmission time from the recorded dive depth to the water surface.

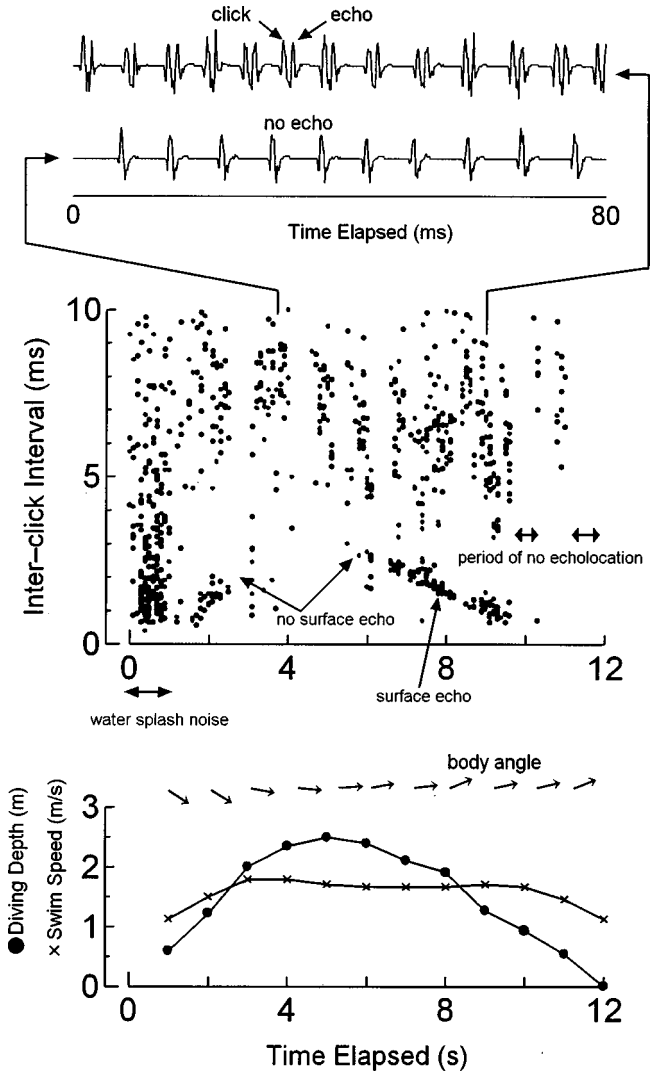


FIG. 3. Another sample of deeper dive of #1. The water surface echo disappeared in the middle of the dive. The waveform of the clicks without echoes shows a single pulse train structure.

the body angle was −31 and +7 deg, respectively. The recorded waveform shows a single pulse train during this time.

III. DISCUSSION

Shallow dives seemed to be common in the juvenile porpoise, as shown in Table I. The average and maximum dive depths of #1 were 1.7 and 11.5 m, respectively.

The acoustic data logger recorded pulse signals without any other animal nearby. This showed that the acoustic data logger could record the sound emissions of the animal wearing the logger. Correspondence of dive depth and delay time of surface echo signals also indicated that recorded sonar signals originated from the animal carrying the loggers. For the individual identification of sonar signals, simultaneous use of an acoustic data logger and a time–depth recorder was effective. Furthermore, the data suggest that the finless porpoise should be able to detect its depth acoustically in shallow water by listening to echoes of its own sonar signals from the water surface. Even with the possibility of not detecting subthreshold signals, the finless porpoises in the present study produced sonar signals frequently, once every 3 to 4 s. In a turbid water system, it may be natural to echolocate frequently.

The acoustic data logger positioned at the upper posterior side of the pectoral fin of the free-ranging finless porpoise detected off-axis signals having amplitudes of at least 150 dB peak-to-peak. The data logger also detected the surface echo of off-axis signals as shown in Figs. 2 and 3. When surface echoes disappeared at 2.5 s in Fig. 3, the body angle of the porpoise was −31 deg. Thus, the echo originated from the off-axis signal at an angle of 121 deg. The animal depth was 1.6 m at this time and the sound-pressure level change during the round trip to the surface was −10 dB. In this case, the off-axis signal should have a peak-to-peak sound-pressure level of 160 dB at 1 m from the animal, where the detection threshold level of the data logger was 150 dB.

The directional beam pattern and acoustic characteristics of a harbor porpoise (*Phocoena phocoena*), belonging to the same family as the finless porpoise, have already been reported (Au *et al.*, 1999). The peak-to-peak source level of

harbor porpoise clicks recorded in a floating net pen was 155–170 dB *re* 1 μPa. The 3-dB beamwidth in the horizontal plane of a harbor porpoise was 16 deg. The off-axis echolocation signals of the free-ranging finless porpoise seemed to have larger sound-pressure level than that of harbor porpoises in the net pen.

ACKNOWLEDGMENTS

We wish to thank Q. Zhao, Z. Wei, X. Wang, B. Yu, M. Kasamatsu, T. Morisaka, T. Tobayama, H. Katsumata, H. Hiruda, M. Nakamura, K. Sano, R. Shimizu, K. Kato, T. Fukuchi, A. Kojima, Y. Narita, and R. Baird. The Institute of Hydrobiology, The Chinese Academy of Sciences, Field Station of Tian-e-zhou Baiji Nature Reserve, National Institute of Polar Research of Japan, Kamogawa Sea World, Marine World Umino Nakamichi, OKI Sea Tech, Shizuoka OKI Electric, and System Giken Co. greatly supported our experiments. W.W.L. Au, Y. Hatakeyama, M. Amundin, K. Dudzinski, T.C. Sparkes, and two reviewers provided constructive criticism on this paper. This research was supported by the Science and Technology Agency of Japan, Japan International Cooperation Agency, Enoshima Aquarium, the Chinese Academy of Sciences (No. KZ951-A1-105), National Natural Science Foundation of China (No. 39670105), and Grant-in-aid for scientific Research (B) from the ministry of Education, Science and Culture (No. B09450172).

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.

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., Herzing, D. L., and Aubauer, R. (**1998**). ''Real-time measurement of the echolocation signals of wild dolphins using a 4-hydrophone array,'' Abstract for World Marine Mammal Science Conference, Monaco, January 1998, pp. 7–8.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer, New York).

Baird, R. (**1998**). ''Studying diving behavior of whales and dolphins using suction-cup attached tags,'' Whalewatcher **31**, 3–7.

Blomqvist, C., and Amundin, M. (**1999**). ''A mobile system for underwater recording of directional broad band pulsed sounds used in intra-specific communication in bottlenosed dolphins,'' Abstract for the 13th Biennial Conference on the Biology of Marine Mammals, p. 18.

Dudzinski, K. M., and Newborough, D. (**1997**). ''Concurrent recording of dolphin behaviours, frequency-modulated tones, and pulsed vocalisations (including echolocation clicks) underwater with a swimmer-propelled system,'' in Proceedings of the Institute of Acoustics 19, Pt. 9, pp. 199–206.

Parks, S., Tyack, P., Nowacek, D., and Johnson, M. (**1999**). ''Use of an archival recording tag to measure acoustic activity and behavior in north Atlantic right whales (*Eubalaena glacialis*),'' Abstract for the 13th Biennial Conference on the Biology of Marine Mammals, p. 144.

Ropert-Coudert, Y., Sato, K., Kato, A., Charrassin, J. B., Bost, C. A., Le Maho, Y., and Naito, Y. (**2000**). ''Preliminary investigations of prey pursuit and capture by king penguins at sea,'' Polar Biosci. **13**, 101–112.

Tyack, P. L., and Recchia, C. A. (**1991**). ''A datalogger to identify vocalizing dolphins,'' J. Acoust. Soc. Am. **90**, 1668–1671.