CODEN: JASMAN

ISSN: 0001-4966

The Journal of the Acoustical Society of America

Vol. 110, No. 3 September 2001

ACOUSTICAL NEWS—USA 1211

USA Meetings Calendar 1222

ACOUSTICAL STANDARDS NEWS 1233

Standards Meetings Calendar 1233

BOOK REVIEWS 1237

REVIEWS OF ACOUSTICAL PATENTS 1241

LETTERS TO THE EDITOR

On the effect of boundaries on radiation resistance of plates [40] K. Renji 1253

The effects of pressure loading on the dynamics of a curved piezostructure [40] James K. Henry, Robert L. Clark 1256

Comment on "Reducing seat dip attenuation" [J. Acoust. Soc. Am. 108, 2211–2218 (2000)] [55] David Lloyd Klepper 1260

Response to "Comment on 'Reducing seat dip attenuation'" [J. Acoust. Soc. Am. 110, 1260 (2001)] [55] W. J. Davies, T. J. Cox 1261

Improving object classification with biomimetic sonar [60] Roman Kuc 1263

Masked threshold difference due to masker harmonicity: Uniform modulation rates of auditory filter outputs versus periodicity in waveform fine structure [66] William C. Treurniet, Darcy R. Boucher 1267

GENERAL LINEAR ACOUSTICS [20]

On scattering from a bubble located near a flat air–water interface: Laboratory measurements and modeling George Kapodistrias, Peter H. Dahl 1271

Crack characterization using guided circumferential waves Christine Valle, Marc Niethammer, Jianmin Qu, Laurence J. Jacobs 1282

Coherent backscattering of elastic waves: Specific role of source, polarization, and near field B. A. van Tiggelen, L. Margerin, M. Campillo 1291

On the character of acoustic waves at the interface between hard and soft solids and liquids Christ Glorieux, Kris Van de Rostyne, Keith Nelson, Weimin Gao, Walter Lauriks, Jan Thoen 1299

Determination of Lamb mode eigenvalues V. Pagneux, A. Maurel 1307

Theory and experiments on the coupling of two Helmholtz resonators Torbjörn A. Johansson, Mendel Kleiner 1315

Sound propagation in rigid bends: A multimodal approach S. Félix, V. Pagneux 1329

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 110, NO. 3, SEPTEMBER 2001

CONTENTS—*Continued from preceding page*

UNDERWATER SOUND [30]

Tomographic inversion of geoacoustic properties in a range-dependent shallow-water environment — Patrick Pignot, N. Ross Chapman — 1338

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

Thin coating characterization by Rayleigh waves: An analytical model based on normal-mode theory — Filip W. Windels, S. Vanaverbeke, O. Leroy — 1349

STRUCTURAL ACOUSTICS AND VIBRATION [40]

Vibratory characteristics of multistep nonuniform orthotropic shear plates with line spring supports and line masses — Q. S. Li — 1360

Mode isolation: A new algorithm for modal parameter identification — Michael V. Drexel, Jerry H. Ginsberg — 1371

An efficient formulation of Krylov's prediction model for train induced vibrations based on the dynamic reciprocity theorem — Geert Degrande, Geert Lombaert — 1379

Experimental study of structure-borne sound transmission loss of mechanical joints — Leping Feng, Mingshu Liu, Anders Nilsson — 1391

Acoustic waves generated by pulsed microwaves in viscoelastic rods: Modeling and experimental verification — Christophe Bacon, Emmanuel Guilliorit, Bernard Hosten, Dale E. Chimenti — 1398

NOISE: ITS EFFECTS AND CONTROL [50]

Modifying modal characteristics of sound fields by state feedback control — Toshiya Samejima — 1408

ARCHITECTURAL ACOUSTICS [55]

Evaluation of decay times in coupled spaces: Bayesian parameter estimation — Ning Xiang, Paul M. Goggans — 1415

ACOUSTIC SIGNAL PROCESSING [60]

Linear prediction coding analysis and self-organizing feature map as tools to classify stress calls of domestic pigs (*Sus scrofa*) — Peter-Christian Schön, Birger Puppe, Gerhard Manteuffel — 1425

PHYSIOLOGICAL ACOUSTICS [64]

Middle-ear function with tympanic-membrane perforations. I. Measurements and mechanisms — Susan E. Voss, John J. Rosowski, Saumil N. Merchant, William T. Peake — 1432

Middle-ear function with tympanic-membrane perforations. II. A simple model — Susan E. Voss, John J. Rosowski, Saumil N. Merchant, William T. Peake — 1445

Wideband reflectance tympanometry in chinchillas and humans — Robert H. Margolis, Saurav Paul, George L. Saly, Patricia A. Schachern, Douglas H. Keefe — 1453

Maturation of the human cochlear amplifier: Distortion product otoacoustic emission suppression tuning curves recorded at low and high primary tone levels — Carolina Abdala — 1465

Origin of cubic difference tones generated by high-intensity stimuli: Effect of ischemia and auditory fatigue on the gerbil cochlea — Thierry Mom, Pierre Bonfils, Laurent Gilain, Paul Avan — 1477

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 110, NO. 3, SEPTEMBER 2001

CONTENTS—*Continued from preceding page*

PSYCHOLOGICAL ACOUSTICS [66]

A level of stimulus representation model for auditory detection and attention — Ervin R. Hafter, Kourosh Saberi — 1489

Melody recognition using three types of dichotic-pitch stimulus — Michael A. Akeroyd, Brian C. J. Moore, Geoffrey A. Moore — 1498

Peripheral auditory processing and investigations of the "precedence effect" which utilize successive transient stimuli — Klaus Hartung, Constantine Trahiotis — 1505

Loudness summation for pulsatile electrical stimulation of the cochlea: Effects of rate, electrode separation, level, and mode of stimulation — Colette M. McKay, Maria D. Remine, Hugh J. McDermott — 1514

Reconciling frequency selectivity and phase effects in masking — Andrew J. Oxenham, Torsten Dau — 1525

Spatial and temporal factors in auditory saltation — D. P. Phillips, S. E. Hall — 1539

SPEECH PRODUCTION [70]

Estimation of viscoelastic shear properties of vocal-fold tissues based on time–temperature superposition — Roger W. Chan — 1548

Effects of oscillation of a mechanical hemilarynx model on mean transglottal pressures and flows — Fariborz Alipour, Ronald C. Scherer — 1562

An inverse dynamics approach to face animation — Michel Pitermann, Kevin G. Munhall — 1570

The acoustic features of human laughter — Jo-Anne Bachorowski, Moria J. Smoski, Michael J. Owren — 1581

SPEECH PERCEPTION [71]

Challenging the notion of innate phonetic boundaries — Susan Nittrouer — 1598

A robust method to study stress "deafness" — Emmanuel Dupoux, Sharon Peperkamp, Núria Sebastián-Gallés — 1606

Minimum spectral contrast needed for vowel identification by normal hearing and cochlear implant listeners — Philipos C. Loizou, Oguz Poroy — 1619

On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception — Oded Ghitza — 1628

SPEECH PROCESSING AND COMMUNICATION SYSTEMS [72]

A probabilistic union model with automatic order selection for noisy speech recognition — Peter Jančovič, Ji Ming — 1641

MUSIC AND MUSICAL INSTRUMENTS [75]

Effects of relative phases on pitch and timbre in the piano bass range — Alexander Galembo, Anders Askenfelt, Lola L. Cuddy, Frank A. Russo — 1649

Vocal intensity in falsetto phonation of a countertenor: An analysis by synthesis approach — Kenneth Tom, Ingo R. Titze — 1667

BIOACOUSTICS [80]

Temperature dependence of ultrasonic enhancement with a site-targeted contrast agent — Christopher S. Hall, Jon N. Marsh, Michael J. Scott, Patrick J. Gaffney, Samuel A. Wickline, Gregory M. Lanza — 1677

Use of a dual-pulse lithotripter to generate a localized and intensified cavitation field — Dahlia L. Sokolov, Michael R. Bailey, Lawrence A. Crum — 1685

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 110, NO. 3, SEPTEMBER 2001

CONTENTS—*Continued from preceding page*

ERRATA

Erratum: "The matched-lag filter: Detecting broadband multipath signals with auto- and cross-correlation functions" [J. Acoust. Soc. Am. 109, 1997–2007 (2001)]	John L. Spiesberger	1696
Erratum: "Measured capacitance of a condenser microphone as a function of diaphragm displacement" [J. Acoust. Soc. Am. 108, 2134–2144 (2000)]	Xinche Yan, Malcolm J. Crocker, Li Jun Zeng	1696

CUMULATIVE AUTHOR INDEX 1697

ACOUSTICAL NEWS—USA

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

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

Preliminary Notice: 142nd Meeting of the Acoustical Society of America

The 142nd Meeting of the Acoustical Society of America will be held Monday through Friday, 3–7 December 2001, at the Greater Fort Lauderdale/Broward County Convention Center in Fort Lauderdale, Florida, USA. A block of rooms has been reserved at the headquarters hotel, the Fort Lauderdale Marina Marriott, which is adjacent to the convention center. A block of rooms has also been reserved at the Radisson Bahia Mar Beach Resort which is closer to the beach and is located about 2 miles from the Convention Center. Transportation between the Radisson and the Convention Center is being planned.

Information about the meeting can be found on the ASA Home Page at ⟨http://asa.aip.org/meetings.html⟩. Online registration is now available.

Technical Program

The technical program will consist of lecture and poster sessions. Technical sessions will be scheduled Monday through Friday, 3–7 December.

List of Special Sessions

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

Acoustical Oceanography (AO)

Turbulence and fine structure studies (Joint with Physical Acoustics)
Acoustic remote sensing of turbulent fine structure and microstructure in the water column and within the surface or bottom boundary layers, including measurements, model verification, and new techniques

Animal Bioacoustics (AB)

Acoustics and marine mammals (Joint with Underwater Acoustics, Acoustical Oceanography and Noise)
Fish audition and sound production
Interaction between marine mammals and the acoustics in their environment
Biomechanical, physiological, anatomical, and behavioral aspects of reception and production of sound by fishes
Session honoring William Watkins
Honoring Bill Watkins and his research on seals and large whales

Architectural Acoustics (AA)

Cruise ship acoustics (High seas acoustical adventures) (Joint with Noise)
Sound isolation, finish treatments, HVAC noise control and sound system design
Integration of synthesis techniques with ''acoustic'' music (Joint with Musical Acoustics)
A variety of controllers, computer programs, recording and playback methods, instruments, virtual orchestras
Speech intelligibility and the metrics used for its evaluation (Joint with Speech Communication)
For many types of architectural spaces, the primary mode of communication is via speech—how we insure that good speech intelligibility will result from our designs
Speech privacy design in office spaces
What architectural factors contribute to the attainment of good speech privacy, how is this performance evaluated and what are the consequences on occupant satisfaction/performance

Biomedical Ultrasound/Bioresponse to Vibration (BB)

Topical Meeting on Physics of Ultrasound in Relation to the Biology of its Therapeutic Effects (see below for details)

Education in Acoustics (ED)

Hands on demonstrations for high school students
15–20 minutes demonstration experiments for high school student participation
Low cost laboratory experiments
Laboratory experience is essential to learning acoustics. Low cost experiments will be presented
''Take fives:'' Sharing favorite ideas for teaching acoustics
Bring your favorite teaching tip—present it in 5 minutes

Engineering Acoustics (EA)

Celebration of Miguel Junger's Acoustics
(Joint with Physical Acoustics, Noise, Underwater Acoustics and Structural Acoustics and Vibration)
Papers celebrating the technical areas to which Junger has made important contributions
Ferroelectric single crystals: From manufacture to devices
New electroactive materials with high coupling factors ($\sim$0.9) are being used to increase bandwidth performance of devices
Small scale acoustics and acoustic devices
Acoustic transducers from the centimeter scale (for wireless phones and hearing aids), to the sub-millimeter (MEMS devices) size scale, and the special acoustics of their design

Musical Acoustics (MU)

Caribbean musical instruments and traditions
Exploration of acoustical properties of instruments and musical performance traditions of the Caribbean region
Musical acoustics in teaching science, engineering and technology (Joint with Education in Acoustics)
New and innovative ways of learning technical subjects through the common language of music
Musical technology (Joint with Education in Acoustics)
Aspects of technology including techniques for creating, recording, and reproducing sound, constructing instruments, and teaching music technology
Naturalness of synthesized sound (Joint with Speech Communication)
Problems and techniques in achieving naturalness in synthesized sound, especially in speech and singing

Noise (NS)

Community noise: Aircraft and other transportation sources
Current case studies comparing measured with modeled results
Computational modeling for community noise
Computational models have become powerful tools for predicting the impact of community noise sources and also for noise control design
Implementing the classroom acoustics standard (Joint with Architectural Acoustics and Signal Processing in Acoustics)
The recently written classroom acoustics standard has been approved. This session will discuss experiences implementing classroom acoustics in new and renovation projects
Noise in condominiums
(Joint with Architectural Acoustics and Structural Acoustics and Vibration)
Condominium noise control, especially between units

 0001-4966/2001/110(3)/1211/22/$18.00

Outdoor sound propagation in coastal areas (Joint with Physical Acoustics and Animal Bioacoustics)
The land–water interface presents unique challenges to the community noise consultant when working in coastal areas

Physical Acoustics (PA)

Applications of acoustic resonance
Having systems respond at acoustic resonance is a powerful method for measuring quantities that are otherwise difficult to detect
High amplitude effects in resonators
Nonlinear fluid dynamics in acoustical resonators

Signal Processing in Acoustics (SP)

Acoustic image reconstruction using tomographic techniques (Joint with Biomedical Ultrasound/Bioresponse to Vibration, Acoustical Oceanography and Underwater Acoustics)
The imaging of acoustic data from a variety of application areas (biomedical, soils, etc.) along with the underlying tomographic reconstruction techniques will be discussed
Theory and applications of acoustic time reversal (Joint with Physical Acoustics, Underwater Acoustics and Biomedical Ultrasound/Bioresponse to Vibration)
This session is intended for contributed papers related to the topical meeting on time reversal
Theory and basic experiments in time reversal
Theory and basic experiments of time reversal are reviewed with the goal of identifying open questions for further investigation
Time reversal in random and reverberating media
Properties of time reversal in random and reverberating media are examined, with an emphasis on areas requiring further research
Time reversal in signal processing
Time reversal from the perspective of signal processing and applications to a wide variety of acoustic problems
Time reversal in the ocean environment
Properties of time reversal in the ocean, where the wave guide character of propagation must be considered
Topical Meeting on Acoustic Time Reversal and Applications
See below for details

Structural Acoustics and Vibration (SA)

Methods of passive dissipation in structural acoustics
New and innovative methods of damping treatments and materials
Numerical methods in structural acoustics
Different aspects of numerical modeling for structural and fluid components

Underwater Acoustics (UW)

How did we get here? Insights into the history of underwater acoustics
A review of some of the insights contributing to the evolution of underwater acoustics
Ocean research and marine mammal regulatory issues (Joint with Acoustical Oceanography and Animal Bioacoustics)
A session of invited talks intended to inform scientific investigators about federal, state, and local regulations that may apply to ocean research using underwater sound, with emphasis on information about permitting and approval processes
Reverberation in shallow water
Characteristics of reverberation in shallow water and the latest advances in modeling

Other Technical Events

ASA's Role in Marine Mammal Bioacoustics

A panel discussion on ASA's Role in Marine Mammal Bioacoustics is being planned to follow the special sessions sponsored by Acoustical Oceanography, Animal Bioacoustics and Underwater Acoustics on (1) Ocean research and marine mammal regulatory issues and (2) Acoustics and marine mammals. The intent of this panel discussion will be to consider the best role for ASA on such complex issues as the increase in ambient noise levels and sonar signals and their effects on marine mammals. Please contact ASA's Executive Director, Charles Schmid (E-mail: charles@aip.org), for more details.

Topical Meeting on Acoustics Time Reversal and Applications

A one-day colloquium and discussion on the topic ''Acoustic Time Reversal and Applications'' sponsored by the Technical Committee on Signal Processing in Acoustics will be held on Monday, 3 December. After an introductory invited talk, five subtopic sessions will focus on basic theory and experiments, random and chaotic media, underwater applications, signal processing and communications, and nondestructive evaluation and medical applications. Each subtopic session will begin with a short overview invited paper, followed by discussion. Emphasis will be on determining the state of the art in each subtopic and identifying open problems. Contributed papers will not be presented Monday, but an opportunity for contributed papers will be provided in a session later in the week.

Topical Meeting on the Physics of Ultrasound in Relation to the Biology of Its Therapeutic Effects

This proposed meeting, sponsored by Biomedical Ultrasound/Bioresponse to Vibration (BB), consists of a one-day colloquium made up of four special sessions, each session containing a few invited presentations followed by a short panel-based discussion that includes audience participation. The emphasis is on how the physical effects generated by ultrasound in any form create associated biological effects that are therapeutic. Examples include (1) High-Intensity Focused Ultrasound or HIFU, typically used for hemostasis, frank tumor destruction, and destruction of kidney stones (i.e., lithotripsy); (2) Low-intensity (sometimes Focused) Ultrasound or LOFU, typically used for accelerating healing; (3) ultrasound with adjunctive agents such as chemicals, radiation, contrast agents, genes—such as ultrasound-assisted thrombolysis, ultrasound-activated/enhanced drug or gene delivery, etc. Note that the discussion of the biology of ultrasound for therapy is to be as valued and detailed in these special sessions as the discussion of the physics. The proposed four special sessions are as follows: (1) HIFU alone, (2) LOFU alone, (3) ultrasound with adjunctive agents, and (4) the imaging of acoustic lesions. Special sessions of contributed papers will compliment these invited sessions.

Exhibit

The instrument and equipment exposition, which will be conveniently located near the registration and meeting rooms, will open at the Fort Lauderdale/Broward County Convention Center with a reception on Monday evening, 3 December, and will close on Wednesday afternoon, 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. A special session on new and innovative techniques by exhibitors is also being planned. Contact the Exhibit Manager, Robert Finnegan, Advertising and Exhibits Division, American Institute of Physics, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502, Tel: 516-576-2433; Fax: 516-576-2481; E-mail: rfinneg@aip.org

History Lectures, Vern O. Knudsen Distinguished Lecture and Hot Topics Session

A ''Hot Topics'' session sponsored by the Tutorials Committee is scheduled. The technical committee on Architectural Acoustics will sponsor a Vern O. Knudsen Distinguished Lecture. 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 Acoustical Oceanography and Underwater Acoustics.

Gallery of Acoustics

The Technical Committee on Signal Processing in Acoustics will sponsor its fifth Gallery of Acoustics at the Fort Lauderdale 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 at-

tendees 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 $350.00 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: randall_w_smith@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.

Paper Copying Service

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

Tutorial Lecture

A tutorial presentation on Noise Outdoors will be given by Tony F. W. Embleton at 7:00 p.m. on Monday, 3 December. Lecture notes will be available at the meeting in limited supply. Those who register by 5 November 2001 are guaranteed receipt of a set of notes.

To partially defray the cost of the lecture a registration fee is charged. The fee is $15.00 for registration received by 5 November and $25.00 thereafter including on-site registration at the meeting. The fee for students with current ID cards is $7.00 for registration received by 5 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/lauderdale/lauderdale.html⟩ to register for the Tutorial Lecture.

Short Course on Applied Digital Signal Processing in Acoustics

This short couse will be held on Friday afternoon, 7 December, 1:00 p.m. to 6:00 p.m. and Saturday morning, 8 December, 8:00 a.m. to 1:00 p.m.

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

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

The registration fee is $300.00 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 5 November will be guaranteed receipt of instructional materials. There will be a $50.00 discount for registration made prior to 5 November. Full refunds will be made for cancellations prior to 5 November. Any cancellation after 5 November will be charged a $50.00 processing fee. Use the form in the call for papers to register for this short course or online at ⟨http://asa.aip.org/lauderdale/lauderdale.html⟩.

Special Meeting Features

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 15 October 2001 to Elaine Moran, ASA, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502, Tel: 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 automobile, 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 Investigator Travel Grant to help with travel costs associated with presenting a paper at the Fort Lauderdale meeting. This award is designed for young professionals who have completed the doctorate in the past 5 years (not currently enrolled as a student), who plan to present a paper at the Fort Lauderdale meeting. Each award will be of the order of $200. It is anticipated that the Committee will grant a maximum of five awards. 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. Deadline for receipt of applications is 26 October 2001. Note that applications will not be accepted by E-mail.

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 Elaine Moran, 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 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 Elaine Moran. Participants are responsible for the cost of their own meal.

Plenary Session, Awards Ceremony, Fellows' Suite, and Social Events

Complimentary buffet socials with cash bar will be held early on Tuesday and Thursday evenings at the Fort Lauderdale Marina Marriott Hotel. The Plenary session will be held on Wednesday afternoon at the Fort Lauderdale/Broward County Convention Center where Society awards will be presented and recognition of Fellows will be announced. A Fellows Hospitality Suite will be open on Thursday from 1:00 p.m. to 4:00 p.m. Refreshments will be provided.

Women in Acoustics Luncheon

The Women in Acoustics luncheon will be held on Tuesday, 4 December. Those who wish to attend this luncheon must register using the form in the call for papers or online at ⟨http://asa.aip.org/lauderdale/lauderdale.html⟩. The fee is $10 for preregistration by 5 November and $15 thereafter including on-site registration at the meeting. There is no fee for Students for preregistration received by 5 November and $5 thereafter including on-site registration at the meeting.

Acoustical Society Foundation Dinner will Feature Author of "The Perfect Storm"

The Acoustical Society Foundation, the College of Fellows and organizers of the Fort Lauderdale meeting have planned some special events for the meeting on Wednesday, 5 December. In cooperation with the Technical Committees on Engineering Acoustics, Physical Acoustics, Noise, Underwater Acoustics and Structural Acoustics and Vibration, a special celebratory invited session honoring long-time ASA Fellow Miguel Junger will take place in the morning. Following the Plenary Session that afternoon, a fundraising dinner will be held, featuring Miguel's son, Sebastian Junger, author of the popular book and movie, *The Perfect Storm*, as guest speaker. We invite all to attend these singular and enjoyable events. All proceeds from the dinner will be added to the ASA Endowment. Further details and ticket information will be available this fall. For more information, contact Bob Frisina, Acoustical Society Foundation; Tel: 716-275-8130, E-mail: rdf@q.ent.rochester.edu.

Transportation and Hotel Accommodations

Air Transportation

The closest airport is Fort Lauderdale Hollywood International Airport and is served by all major airlines including Delta, US Airways, American, Northwest, and Continental. The airport designation is FLL. Website: ⟨www.fll.net⟩. The alternative to arriving at Fort Lauderdale Hollywood International Airport is to arrive at either Miami International Airport (MIA, website: www.miami-airport.com) or Palm Beach International airport (PBI, website: www.pbia.org). Scheduled air transportation is available from MIA or PBI to Fort Lauderdale Hollywood Airport, although this is not very common. Check with the airlines for departure times to determine if such service would be convenient for your arrival time.

Ground Transportation

The Fort Lauderdale Hollywood International Airport is about 9 minutes by automobile (3 miles) southwest of the Fort Lauderdale Marina Marriott Hotel and about 12 minutes by automobile (4 miles) from the Radisson Bahia Mar Beach Resort. Transportation from the airport to the hotel may be by car, taxi or shuttle. Neither of the two hotels provide complimentary transportation to and from the Fort Lauderdale Hollywood International Airport. Convenient transportation from the Miami International Airport or Palm Beach International Airport to the Fort Lauderdale Marina Marriott Hotel or the Radisson Bahia Mar Hotel is provided by one of the shuttle van services. The cost of the van service is approximately USD $15 to USD $20 per person from MIA and approximately USD $30 to USD $40 per person from PBI. The trip to the hotel will take about 45 minutes from MIA and about 1 and 1/2 hours from PBI after you have contacted the shuttle service. There could be a 30-minute waiting time to get the shuttle service. An alternative to using the shuttle van service from MIA (30 miles SW) or PBI (60 miles NW) is to rent a car and drive to the hotel. Because of the cost, travel by taxi from MIA (USD $60) or PBI (USD $100) to either the Fort Lauderdale Marina Marriott Hotel or the Radisson Bahia Mar Hotel is not recommended for riders travelling alone.

Hotel Accommodations

All technical sessions and some other daytime functions will be held at the Greater Fort Lauderdale/Broward County Convention Center, which is directly across the street from the headquarters hotel, the Fort Lauderdale Marina Marriott Hotel. The evening functions and meetings will be held at the Fort Lauderdale Marina Marriott Hotel. The Marriott is approximately 2 miles from Fort Lauderdale beach. For those who are interested in hotel accommodation closer to the beach, the Radisson Bahia Mar Beach Resort also has a block of rooms reserved for the meeting. Transportation is being planned between the Radisson and the Convention Center. For either of the two hotels, please make your reservations directly with the desired hotel and ask for one of the rooms being held for the Acoustical Society of America. If attendees expect to stay in the hotel close to the convention center they are encouraged to make their reservations early. The reservation cutoff date for the special discounted ASA rates is 1 November 2001; after this date the special ASA rate will not be available.

The Fort Lauderdale Marina Marriott Hotel is the headquarters hotel. Room rates are $125.00 Single/Double/Triple/Quad plus 11% state and local taxes. To reserve a room contact the hotel directly at 1881 S.E. 17th Street, Fort Lauderdale, FL 33316; Tel: 954-463-4000, Toll-free reservations: 800-695-8284, Fax: 954-527-6701.

A limited number of rooms are available at the government rate.

The Radisson Bahia Mar Beach Resort is at the beach. Room rates are $129.00 Single/Double/$139.00 Triple/$149.00 Quad plus 11% state and local taxes. To reserve a room contact the hotel directly at 801 Seabreeze Boulevard, Fort Lauderdale, FL 33316; Tel: 954-764-2233, Toll-free reservations: 800-695-8284, Fax: 954-524-6912.

Remember to mention the Acoustical Society meeting when making reservations.

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 15 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 always the possibility of showers. Daytime temperatures in Fort Lauderdale in December average about 23 °C (75 °F). A sweater and jacket may be required for the evenings.

General Information

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: Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; E-mail: asa@aip.org

Accompanying Persons Program

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

The Galleria shopping mall ⟨www.galleriamall-fl.com⟩ is located 4 miles from the convention center, and courtesy shuttles are available from the Marina Marriott for those staying at that hotel. The Galleria, has two spectacular levels of world-class shopping at 150 stores.

Las Olas Boulevard ⟨www.lasolasboulevard.com⟩ in the heart of downtown Fort Lauderdale, is lined with chic boutiques and fine dining spots. Shops and galleries along this beautifully landscaped, gas-lighted street carry a range of goods from European antiques to clothing and stained glass lighting. Las Olas Boulevard can be reached by a short car or taxi ride by the water taxi which can be boarded from right outside the hotels.

Sawgrass Mills ⟨www.millscorp.com/sawgrass⟩ is approximately 13 miles from the Convention Center and offers over 270 brand-name and designer outlets plus specialty shops and restaurants.

Billie Swamp Safari ⟨www.seminoletribe.com⟩ at Big Cypress Reservation; Tel: 941-983-6101, Toll-free: 800-949-6101. Learn about the land, history, and culture of the Seminole Indians. Skim over glittering water-

scapes in an airboat, experience a swamp buggy eco-tour, paddle a canoe, walk the Swamp Trail and come face-to-face with the resident alligators.

Butterfly World ⟨www.introweb.com/butterfly⟩ on 3600 W. Sample Rd., Coconut Creek; Tel: 941-977-4400. Stroll through 3 acres of lush tropical gardens while thousands of live, brilliantly colored butterflies soar around. Waterfalls, fish, hummingbirds, an insectarium, orchids and roses complete the natural habitat.

International Game Fish Association (IGFA) Fishing Hall of Fame & Museum ⟨www.igfa.org⟩ on 300 Gulf Stream Way, Dania; Tel: 941-922-4212 is a three story museum that includes five galleries, an outdoor marina, a theater, and indoor/outdoor demonstration areas.

Just minutes from the beach, hotels, and convention center is the Riverwalk ⟨www.goriverwalk.com⟩ Arts and Entertainment District in the heart of downtown Fort Lauderdale. It extends from SW 7th Avenue to SW 2nd Avenue, and is Florida's most beautiful mile! It showcases lush tropical landscape and winding walkways linking downtown Fort Lauderdale's cultural attractions, restaurants, parks, and shops. The area includes the Broward Center for the Performing Arts, the Museum of Discovery and Science with its Blockbuster 3D IMAX Theater, the Florida Grand Opera, Old Fort Lauderdale, Museum of Art, the Fort Lauderdale Film Festival, and Stranahan House. All are within walking distance of one another along brick-lined Riverwalk, a meandering promenade that runs along the New River.

Museum of Discovery & Science and Blockbuster 3D IMAX Theater ⟨www.mods.org⟩ on 401 SW 2nd St., Fort Lauderdale (Tel: 941-467-6637 or 941-463-IMAX) is a dynamic hands-on exhibit and five-story IMAX theater.

Not only does Broadway come to Fort Lauderdale at the state-of-the-art Broward Center for the Performing Arts along the New River in downtown Fort Lauderdale's Riverwalk Arts and Science district, but more than 20 home-grown theater companies perform year-round.

Registration Information

The registration desk at the meeting will open on Monday, 3 December at the Greater Fort Lauderdale/Broward County Convention Center. To register use the form included in the call for papers or online at ⟨http://asa.aip.org/lauderdale/lauderdale.html⟩ or register online at ⟨http://www.aip.org/forms/asa/registration_form.html⟩. **If your registration is not received at the ASA headquarters by 16 November you must register on-site**.

Registration fees are as follows:

Category	Preregistration by 6 November	Registration after 6 November
Acoustical Society Members	$240.00	$290.00
Acoustical Society Members One-Day	$120.00	$145.00
Nonmembers	$290.00	$340.00
Nonmembers One-Day	$145.00	$170.00
Nonmember Invited Speakers (Note: The fee is waived for these speakers if they attend the meeting on the day of their presentation only)	$240.00	$290.00
Students (with current ID cards)	Fee waived	Fee waived
Emeritus members of ASA (Emeritus status preapproved by ASA)	$35.00	$45.00
Accompanying Persons (Spouses and other registrants who will not participate in the technical sessions)	$35.00	$45.00

Nonmembers who simultaneously apply for Associate Membership in the Acoustical Society of America will be given a $50.00 discount off their dues payment for the first year (2002) of membership. (Full price for dues: $100.00.) Invited speakers who are members of the Acoustical Society of America are expected to pay the registration fee, but **nonmember invited speakers** who participate in the meeting for one day only may register without charge. Nonmember invited speakers who wish to participate in the meeting for more than one day will be charged the member registration fee, which will include a one-year membership in the ASA upon completion of an application form.

NOTE: A $25.00 PROCESSING FEE WILL BE CHARGED TO THOSE WHO WISH TO CANCEL THEIR REGISTRATION AFTER 5 NOVEMBER.

Reports of Technical Committees 2000–2001

Acoustical Oceanography

I write this year-end report for Acoustical Oceanography (AO) with mixed emotions, as I am now chairman ex-officio of AO, and this is my last piece of ''chairman's business.'' I feel a bit let down, as being the AO TC chair was exciting, and put me in the center of a very active, motivated group of researchers, as well as in close touch with the ASA as a whole. I feel good, however, knowing that AO's new chair, Dr. Peter Worcester, will do a wonderful job leading the committee, and that AO will thrive under his leadership. Peter is a longtime colleague and shipmate of mine, and is one of the most knowledgeable and broad-based people in acoustics and oceanography that I know. I also feel good that AO seems to have done reasonably well under my stewardship, and that I did not incur any major disasters. (OK, I inadvertently sent the Navidad virus out with one of my Chairman's memos, but I believe that was the worst of it!) Anyway, before I settle back comfortably into my ''ordinary citizen'' status, let me discuss the last years events in AO with you.

The primary activity for a technical committee chairman is to organize the business that is transacted at the ASA meetings, and so let me start with our last two meetings, in Newport Beach and in Chicago. The technical sessions at both meetings were interesting and well attended. The Newport Beach meeting featured three special sessions on ''Bioacoustics'' (a more and more important topic as the years go on), a session on ''Acoustic Measurements of Sediment Transport,'' and joint special sessions in ''Detection and Classification of Bubbles'' (with BB and PA) and ''High Frequency Sediment Acoustics and Associated Sediment Properties'' (with UW). The Chicago meeting featured three special sessions on ''Acoustical Instrumentation for Water Column Measurements,'' two joint special sessions (with UW) on ''Inverse Methods for Sub-Bottom Surveys,'' and a joint special session (with UW) on ''Benchmarking Range Dependent Numerical Models.'' In addition, there were a number of excellent contributed sessions rounding out the technical agenda. I would like to thank all the session organizers for doing a wonderful job last year (and the two previous years, as well!)—the AO sessions were all first rate! One of AO's recent initiatives that has helped to improve our technical sessions is the ''Mini-Course in Oceanography,'' in which we have two oceanographers, generally from outside the ASA, attend the technical sessions to present one-hour mini-tutorials about various areas of oceanography (which are generally related topically to our sessions). This initiative has been well received, and I am pleased to know that Peter Worcester wants to keep it as an ongoing one during his chairmanship. Last year, we were lucky to have George Rose and Peter Wiebe talk at Newport Beach (on fisheries acoustics and zooplankton acoustics, respectively), and Larry Mayer and Manell Zakharia talk at Chicago (on applications of multibeam, wideband, and correlation sonars for various water column investigations).

Also part of our technical sessions is our ''Best Student Paper Awards.'' At each meeting, we give awards (of $250, $150, and $100) to the three papers that our judging panel deems best in AO. This year's Newport Beach winners were Kelley Benoit-Bird for ''Acoustic Investigration of the Hawaiian Mesopelagic Boundary Community,'' Alex De Robertis for ''Three Dimensional Acoustic Tracking of Krill With a Multibeam Sonar,'' and Joshua D. Wilson for ''Acoustic Detection and Classification of Hurricanes.'' Our congratulaions to all of them. (The Chicago winners have yet to be announced.)

Let me turn from the meetings to another topic, some of the Awards and Prizes that AO has both received and given this last year. The biggest piece of news has to be the receipt of the ASA's Gold Medal by Hank Medwin, AO's founder. Hank has always been one of the ASA's stars, in research, teaching, and citizenship, and this prestigious career award is well deserved! Way to go, Hank! On the ''early career'' side of things, our congratulations go to John Colosi, who won the prestigious Wood Medal this year. John's work on the long-range acoustics has gained him recognition worldwide, and it is great to see him receive this major award. Finally, let me mention the Medwin Prize that was initiated this last year. The first award of that Prize, which was initiated and funded by Hank and Eileen Medwin, will go to Tim Leighton for his work on bubbles (highly fitting!). Unfortunately, due to some family health problems, Tim will be receiving his prize at a future meeting, as he could not do so in Chicago. We congratulate Tim on being a fitting and very deserving first recipient of the

Medwin Prize, and send him our prayers and best wishes for his family's health. Finally, let me mention Fellows. Though I do not have the list of recent AO Fellows close at hand (mea culpa), I know we have done well in having Fellows elected, and I hope to see this trend continue in the future. There are still a number of people in our committee that deserve this recognition, and it takes our mutual efforts to see that they get it.

I will leave off with a tiny bit of philosophizing about three topics concerning AO's present and future. First, there is the ongoing debate over the ''specialization'' that has gone on in the area of ocean acoustics. The ASA now has committees in Underwater Acoustics, Acoustical Oceanography, Animal Bioacoustics, and Signal Processing, all of which have large components of underwater sound work. Some people in the ASA regard this as ''fragmentation.'' However, I personally regard it as ''inevitable evolution'' of the kind that has been seen in many areas of science. But, that aside, we still want and need to have these related groups interact, and would all like to be able to hear talks in all these areas at the meetings. This is a logistical problem more than anything else, and should be carefully addressed in the years to come. A second problem for all users of underwater sound is marine mammal regulations. We need to both protect our marine life and do our scientific work, and to do so, we need to work carefully with the many groups concerned with regulating underwater sound. Currently, I think that the people concerned with underwater sound in the ASA are doing a good job addressing this. However, we cannot let down our guard! (As an aside, Peter Worcester is extremely knowledgeable in this area, which should be of great benefit to AO.) My final piece of folk-philosophy is a question that my former graduate student, Rich Pawlowicz, used to ask facetiously once in a while: ''whither goest thou, Mr. Oceanographer?'' A little self-examination as to where the future might lie is often a good thing, and perhaps we as a committee could consider doing this in the near future. Just a thought! Well, let me formally sign off here. Thanks for putting up with me for the past 3 years—it was great fun!

JAMES F. LYNCH
Chair 1998–2001

Architectural Acoustics

TCAA had a very productive year during 2000 with many presentations from meetings in Atlanta and Newport Beach (jointly with Noise–Con 2000) and the activities supported by Technical Initiatives from the Technical Council.

In summary the major events for the year included the ASA meetings in Atlanta and Newport Beach, completion and publication of the Classroom Acoustics booklet, continued work in support of the standards group on Classroom Acoustics, work with the International Standards Organization (ISO) and the Student Design Competition.

Sessions presented at the Atlanta meeting included Electronic Enhancement and the Traditionally Unamplified Art Form; Stadium and Arena Acoustics; Acoustical Design of Learning Spaces; Recent Authors of Books on Architectural Acoustics; History of Architectural Acoustics with Ewart Wetherill; Computer Modeling and Measurement Techniques for Large Room Acoustics. Also at the Atlanta meeting was a poster session on Classroom Acoustics. Our Technical Committee organizer for the Atlanta meeting was Scott Pfeiffer. Session chairs and co-chairs included Scott Pfeiffer, Jack Randorff, Brandon Tinianow, William Cavanaugh, Ron Freiheit, and Paul Calamia. TCAA member Mendel Kleiner was welcomed as a Fellow to the ASA. Dorie Najolia attended the student council meeting in Atlanta as the TCAA representative. TCAA, Noise and NCAC jointly sponsored a Student Reception. One of the special events of the Atlanta meeting was a book fair which included a number of authors from Architectural Acoustics.

The Newport Beach meeting was held jointly with Noise–Con. A number of joint sessions along with Noise were part of the Noise–Con meeting; Heating, Ventilation, and Air Conditioning Noise; International Noise Standards; Environmental Noise Focused on Combined Noise Sources. The regular ASA sessions included Building Renovation and Sound Scattering and Absorption; Integration of Synthesis Techniques and ''Acoustical'' Music; Amphitheater Acoustic Design and Sound Control for Nearby Communities; Building Acoustics Test Standards; Speech Privacy in the Built Environment; Measures of Auditorium Acoustics. Along with the sessions there was a Technical Tour of the Orange County Performing Arts Center hosted by Jerald Hyde. Representing the Architectural Acoustics Technical Committee at the organizing meeting for the Newport Beach meeting were Angelo Campanella and Lily Wang. The session chairs and co-chairs for the Newport Beach meeting were Gary Siebein, Courtney Burroughs, Paul Schomer, Brigitte Schulte-Fortkamp, Richard Campbell, Tony Hoover, Dana Hougland, Les Blomberg, David Marsh, Angelo Campanella, and Ken Roy. ASA welcomed TCAA members Soren Bech, Peter D'Antonio, and Mark Holden as Fellows. TCAA, Noise and NCAC jointly sponsored a Student Reception. TCAA along with TCEA under the direction of Neil Shaw and Allan Devantier and in association with American Loudspeakers Manufacturer's Association, JBL Professional, JBL Consumer, Infinity Systems and Revel held a Student Design competition for the design of a loudspeaker system. First Honors and $1000 was awarded to Shawn Devantier from the University of Victoria, British Columbia. Commendations and $500 each was awarded to to Ara Baghdassarian from the University of California, Northridge; Geoff Christopherson of the University of Southern California and Dave Tremblay of the University of Colorado.

Technical Initiatives supporting TCAA were used for the following projects in 2000: Publication of the Classroom Acoustics booklet developed by Bob Coffeen, Ben Seep, and Robin Glosemeyer.

TCAA also maintained its participation in Continuing Education for architects by providing AIA credits for presentations such as the Classroom Acoustics Workshop.

Many of the members from TCAA participated in the Standard Working Group in the development of guidelines for acoustics properties of properly design classrooms. The group received funding from the Access Board to develop this standard with the goal of completing the ANSI standard by 2001.

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

Sixteen members attended the B&B technical committee (TC) meeting in December, 2000, in Newport Beach, CA. Special sessions in Newport Beach were organized by Tim Leighton and Ron Roy, while Mark Prausnitz chaired a lengthy session and Larry Crum gave the ''Hot Topics'' lecture on Biomedical Ultrasound. Carr Everbach organized and chaired the Topical Meeting on ''Physics of Echo-contrast Agents,'' which had a maximum attendance of over 50. Panelists for the Topical Meeting included John Allen, Charlie Church, Bob Apfel, Kathy Ferrara, Larry Crum, Tom Matula, Mike Averkiou, Jim Miller, Diane Dalecki, Lars Hoff, Andy Hurley, Junru Wu, Wes Nyborg, and Jim Greenleaf. Mike Bailey was the TPOM representative for Newport Beach and arranged the papers into sessions.

Technical Initiatives from B&B for the year 2001 include webpage development (please send links to your website to Shira Broschat for inclusion on the B&B homepage), a student paper contest in Chicago, travel expenses of some invited speakers, and funds for inviting IEEE–UFFC members to join ASA B&B. Sandy Poliachik will continue as our liaison to the student council until the Fort Lauderdale meeting.

We continued our plan, since the Berlin meeting, of putting most of our Special Sessions in the Spring ASA meeting each year, to avoid conflicts with AIUM or IEEE.

The Chicago TC meeting in June, 2001, was attended by 29 people. Special Sessions were organized by Doug Mast, Christy Holland, Brian Fowlkes, and Pei Zhong (the Brad Sturtevant Memorial Session in Lithotripsy). Glynn Holt, Paul Carson, Bill O'Brien, and Carr Everbach chaired sessions, and Tom Royston organized the B&B student paper contest. Robin Cleveland ably chaired the ''Hot Topics'' session and Brian Fowlkes was the hero of the meeting for being the TPOM representative (arranging the papers into sessions and juggling room choices). At the Chicago meeting, B&B had 78 papers in eight sessions, all of which were well attended (typically 40–60 attendees per session).

In Chicago, Floyd Dunn and Ed Carstensen received commendations for 50 years of membership in the ASA, and Christy Holland, Shira Broschat, and Carr Everbach, were made new Fellows. Tim Leighton was announced as the recipient of the first Medwin Prize.

The student paper contest in Chicago was very successful, with 13 entries and 18 judges. First prize went to Oliver Kripfgans (University of Michigan) for ''Acoustic vaporization of single droplets;'' second prize went to Sandy Poliachik (University Wash/APL) for ''Role of high intensity focused ultrasound induced cavitation on platelet activation;'' and third prize

went to Mark Haun (University of Illinois) for ''Efficient three-dimensional cylindrical-geometry ultrasound imaging.''

In keeping with our plan to use the Fall meetings for innovative formats, we have currently planned a one-day Topical Meeting on ''The Physics of Ultrasound in Relation to the Biology of its Therapeutic Effect,'' organized by Pierre Mourad and Larry Crum. For the Pittsburgh meeting, we are currently planning the following Special Sessions: Flow Characterization and Monitoring; Scattering Theory in Biological Media; Ultrasonic Field Characterization; Ultrasound Propagation in Bone; Acoustic Microscopy. Additional good ideas include a grant-writing workshop and a special session on ''Funding Biomedical Engineering Research'' which would bring in funders such as Whitaker Foundation or Howard Hughes Medical Institute to explain what they do.

We are already planning the Cancun, Mexico meeting, that will take place 2–6 December 2002, at the Fiesta Americana. Special sessions may include the following: Ultrasound Meditated Drug Delivery and Gene Transfection, Diagnostic Ultrasound Applications, Lithotripsy, Acoustics in Multiphase Media. Due to scheduling issues, we will need to finalize these plans in Fort Lauderdale. We hope to see you at this or another upcoming meeting.

E. CARR EVERBACH
Chair 1999–2002

Engineering Acoustics

The Engineering Acoustics Technical Committee (EATC) met at each of the two meetings of the Society, in Newport Beach and in Chicago. In Newport Beach, the EATC sponsored three special sessions and two sessions of contributed papers. The special sessions and their organizers were Ultrasonic Sensors and Motors–Harold Robinson; Acoustic Measurements and Materials Characterization–Lowell Smith; International Comparison of Calibration Methods and Measurements–George Wong. In addition, EATC was a co-sponsor of the Loudspeaker Student Design Competition, with Architectural Acoustics. Julien Bernard won the EATC Student Paper Award for the paper entitled ''Design of actively tuned flexural piezoceramic bar/disk transducers.''

The EATC met in Newport Beach on the evening of Tuesday December 5, with 27 members and friends participating. One item of particular interest to the committee was the issue of continued support by the Society of the activities of ISO Technical Committee 108. For many years, the society has maintained the Secretariat of TC108, which deals with Mechanical Vibration and Shock. The Secretariat currently maintains 80+ standards and has 140+ additional standards in review, most of which will probably be issued in the next year or so. The society estimates its costs at ~$100k/year to maintain this Secretariat. Mahlon Burkhard, George Wong, and Victor Nedzelnitsky spoke in favor of continuing support, while seeking outside funding to cover at least a part of the costs. It was suggested that the companies that are regulated by the TC108 standards should be a willing source of funding for the Secretariat if they understood the issue. It was further suggested that the income associated with running the Secretariat will increase significantly as many new standards come online and we collect the revenue from publication and sales. A poll of the members on interest in continuing to maintain the Secretariat at least until it can be assured that there is another US entity who could manage it was unanimously in favor (20 in favor, 0 opposed).

In Chicago, the EATC sponsored three special sessions and three sessions of contributed papers. The special sessions and their organizers were Transducers for Underwater Vehicle Sonars–Thomas Howarth; Modern Magnetostrictive Materials–Harold Robinson and Stephen Butler; Magnetostrictive Transducers–Stephen Butler and Harold Robinson. The first of these special sessions was co-sponsored with the Technical Committee on Underwater Acoustics.

The EATC met in Chicago on the evening of Tuesday June 5, with 26 members and friends participating. Members were happy to hear that the Executive Council has agreed to continue the support of ISO TC108 at least through this year while continuing to look for sources of funding outside the Society. Interested committee members and friends are asked to recommend companies, and people within those companies, who might be able to support these activities.

EATC would like to express its thanks to Robert Finch who completed 6 years of service as the chair of the Subcommittee on Awards and our representative to the Medals and Awards committee of the Society. Bob completes his tenure at this meeting having successfully submitted candidates for the award of Silver Medal in Engineering Acoustics. Mahlon Burkhard has agreed to act as Chair of the Subcommittee for the next term. Many thanks to Bob for his years of service, and to Mahlon for his activities in the future.

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

STEVEN C. THOMPSON
Chair 2000–2003

Musical Acoustics

2000–2001 was another eventful year for the Technical Committee on Musical Acoustics (TCMU). The December meeting in Newport Beach featured some special events in addition to the invited and contributed paper sessions sponsored by the committee. Julius Smith presented the Tutorial Lecture on Virtual Musical Instruments. Dean Ayers presented a Hot Topics talk, ''Recent studies of the lip reed: Low-tech experiments and stroboscopic observations.'' A special session on historical brass instruments and brass band performance organized by Dean Ayers was followed by a performance by the Americus Brass Band. Other Musical Acoustics special sessions at Newport Beach were a session on modal analysis organized by Dan Russell and a session on Asian musical instruments and traditions organized by Tom Rossing and Jim Cottingham. The musical activities concluded on Friday with a technical tour of the Fiske Musical Instruments Museum at The Claremont Colleges.

There were three Musical Acoustics special sessions at the Chicago meeting. A session on experimental musical instruments organized by Jim Cottingham included invited speakers representing a variety of disciplines and points of view. The session concluded with an extended demonstration and performance on ceramic flutes by instrument maker/performer Susan Rawcliffe. A special session on new synthesis techniques organized by Jim Beauchamp included a hands-on demonstration of a continuum fingerboard. The session, ''3-D Spatialization for Music Applications,'' organized by Gary Kendall concluded with a concert illustrating diverse approaches to music incorporating spatial features. The contributed paper session at Chicago chaired by Uwe Hansen included papers on a wide range of topics. Concluding this session were three papers on the acoustics of the Baltic psaltery, which included demonstrations on several instruments.

Efforts to promote student involvement in musical acoustics continue through the student paper competition and joint sponsorship of student socials. Rachel Romond continues to serve as the Musical Acoustics representative on the ASA Student Council. There were nine entrants in the student paper competition in Newport Beach and six entrants in Chicago. The winner of the Best Student Paper Award in Musical Acoustics at Newport Beach was Jeffrey Moffitt, of Baldwin–Wallace College, who presented ''Self-sustained oscillation as a function of blowing pressure in a one mass lip reed model.'' The award winning paper at the Chicago meeting was ''Normal modes of a finite element lip reed model,'' presented by Daniel Ludwigsen of Brigham Young University.

Technical initiative support enabled Uwe Hansen to conduct another in the series of educational workshops on acoustics for teachers. A technical initiative begun in 2000 and continuing into 2002 is a project being conducted by Peter Hoekje which involves preparation of a set of demonstrations in musical acoustics to be available on CD-ROM and on the web.

The representatives to the Technical Program Organizing Meetings in 2000–2001 were Dean Ayers for the Newport Beach meeting and Jim Beauchamp for the Chicago meeting. The Chicago meeting marked the end of the term of service for Roger Hanson as the TCMU representative on the Medals and Awards Committee. Jim Beauchamp has been appointed to replace him on this committee. Representation from the TCMU on other ASA committees continues with Uwe Hansen on Membership Committee and Ian Lindevald on the Committee on Standards. Stephen McAdams was appointed at the Chicago meeting to work with Dean Ayers as an additional *JASA* Associate Editor for Musical Acoustics. Members appointed or reappointed to serve on the technical committee for terms expiring in 2004 include R. Dean Ayers, Judith C. Brown, Courtney B. Burroughs, Robert D. Collier, George F. Emerson, Bozena Kostek, Barry Larkin, Daniel O. Ludwigsen, Thomas D. Rossing, Julius O. Smith, and William J. Strong.

The web site for the TCMU continues to maintain current information on musical acoustics at ⟨http://www.public.coe.edu/~jcotting/tcmu⟩.

JAMES P. COTTINGHAM
Chair 1999–2002

Noise

As the first order of business in this report, the Technical Committee on Noise would like to commend its outgoing Chair, Richard J. Peppin, for 3 years of outstanding service. During Rich's tenure the membership of TCN grew and became stronger. He promoted and encouraged many technical programs and special sessions of high quality. His leadership was clearly evident, and due to his contributions TCN prospered and is in a great position to continue to benefit its members, and the ASA. Thanks Rich!

The Technical Committee on Noise met twice in the past year: in the fall of 2000 in Newport Beach, CA, and in the spring of 2001 in Chicago, IL. Both meetings had a great turnout.

The ASA meeting in Newport Beach was particularly well attended, as it was a joint meeting with the Institute for Noise Control Engineering (INCE). This was a very large and successful meeting for the noise community, with over 25 Special Sessions with more than 130 papers. As Leo Beranek and Bill Lang pointed out in a distinguished lecture on the History of the ASA in Noise Control at the meeting, the ASA and INCE must continue to combine efforts on both future meetings and publications in order to stay strong and serve their constituencies. The Newport Beach meeting could serve as a model for future collaboration between the ASA and INCE.

It was also noted that the meeting of the National Council of Acoustical Consultants (NCAC) held in conjunction with the ASA meeting was the largest ever. Most members agreed that the joint meeting with INCE was a significant factor.

There are many people to thank for the success of the Newport Beach meeting, including the Program Organizers, Rich Peppin, Brigitte Schulte-Fortkamp, George Maling, Alan Marsh, and John Van Houten; the Session chairs, Bruce Walker, James Thompson, Alan Marsh, Gordon Bricken, Rob Greene, Greg Tocci, Gary Siebein, Dave Lubman, Glenn Warnaka, John Van Houten, Ken Cunefare, Bob Hellweg, Lou Sutherland, Brigitte Schulte-Fortkamp, Joe Pope, Frank Brittain, Paul Schomer, Dan Raichel, Paul Donavan, Bob Putnam, and Tor Kihlman, and, of course, all of the paper presenters.

The ASA meeting in Chicago offered a number of fine sessions on a variety of topics from soundscapes and community noise policy to automotive and aircraft noise source measurement and control. TCN also cosponsored several sessions with architectural and engineering acoustics. Thanks go to the Program Organizers, Angelo Campanella and Bennett Brooks; Session chairs, Brigitte Schulte-Fortkamp, Dan Raichel, Angelo Campanella, Bob Hellweg, Gilles Daigle, Dave Lubman, and Joe Pope, and the many paper presenters.

The positive position of TCN is evidenced by the enthusiastic support by members for developing many noteworthy special sessions and tutorials for the upcoming meetings in Fort Lauderdale, Pittsburgh, and Cancun. These meetings should prove to be very interesting and rewarding.

An area of continued visibility is the Standard on Classroom Acoustics. The draft standard has now been approved by the ANSI S-12 Committee on Noise, and final comments are currently being resolved. It is expected that the standard will be published later this year. That should help to contribute to quieter classrooms, and a more educationally friendly environment in our schools.

Also, as many know, the new ASA Standards Manager, Susan Blaeser, has moved the Standards office to Melville, Long Island, near the ASA main office, from its previous Wall Street location. This should allow better support and coordination of its extensive standards work. In related news, the ASA Executive Council committed to continue its support for the Secretariat of the ISO Technical Committee TC 108 on Shock and Vibration. It was recognized that support of US participation and leadership in this area of international standards is crucial to the Society. The ASA Committee on Standards is exploring ways of increasing funding for this effort. One way is to encourage TCN members to seek support from their organizations for membership in one or more of the ANSI S committees. The cost is truly small relative to the significant benefits.

TCN would like to thank Noral Stewart for his leadership and efforts to further the Outstanding Papers by Young Presenters awards. Many candidates presented papers, and Noral organized a large contingent of judges. As of this writing, the results are still being tabulated. Any awards for the Chicago meeting will be declared in forthcoming TCN announcements. The latest award was recently presented to Alan Kennedy for his paper at the Columbus meeting.

Technical Initiative funding for TCN activities continues in the areas of the younger presenter awards, web page maintenance, the student reception, and the development of the online library (see www.nonoise.org). New areas include support for a classroom construction TV documentary (with TCAA and TCSC), and funds to distribute the classroom acoustics booklet in PDF format.

In other action, TCN membership elected to cosponsor a Miguel Junger Retrospective session. Also, TCN is actively seeking candidates for the position of Student Representative to the ASA Student Council. Some travel support is being offered. Candidates and the sponsoring faculty member should contact the TCN Chair.

Finally, on a personal note, I would like to express my heartfelt appreciation to the TCN membership for giving me the opportunity to serve you as Chair. It is my hope that I can meet the high expectations for this position established by my predecessors. With your help, TCN can continue to provide an outstanding forum for the evolving exchange of ideas in the noise community.

BENNETT M. BROOKS
Chair 2000–2003

Physical Acoustics

We continue to have nice, well-attended technical committee meetings. Forty-three people attended in Newport Beach, with seventy-six in Chicago.

I believe that the pleasant homey atmosphere of the Acoustical Society is due in large part because it is largely a society of volunteers. Our thanks go out to them. Sameer Madanshetty agreed to continue serving as our representative on the ASA Committee on Standards, Steven Garrett agreed to continue serving as our representative on the Membership Committee. James Sabatier represents us on the Medals and Awards Committee. Preston Wilson has been representing students in physical acoustics at the Student Council. Larry Wilen and Anthony Atchley were our TPOM representatives in Chicago, and I did it in Newport Beach. Kevin Bastyr and Matt Poese have helped us with refreshments at our TC meetings. Special sessions were organized by Jim Sabatier and George Mozurkewich. Thank you all.

Mark J. Marr-Lyon won the first place prize ($300) for his presentation ''Passive stabilization of capillary bridges in air with acoustic radiation pressure and the equilibrium shape of bridges,'' and Ray Scott Wakeland received the second place prize ($200) for ''Numerical model of heat exchangers in oscillating flow with oscillating pressure,'' in our first Best Student Paper in Physical Acoustics competition in Chicago. It was a difficult decision for the judges since so many of the papers were excellent. Picking winners was much like splitting hairs, but in the end the subcommittee decided to follow the original rules and not split the awards up further. The effort to judge the papers was organized by Keith Wilson, and he was assisted by James Chambers, Kerry Commander, Kenneth Gilbert, Bart Lipkins, Philip Marston, Ralph Muehleisen, David Chambers, Sandra Collier, and Steven Kargl.

At the Newport Beach meeting, Gregory W. Swift received the Silver Medal in Physical Acoustics for his work in thermoacoustics.

As a technical initiative, we cosponsored with Eindhoven University of Technology the First International Workshop in Thermoacoustics, held April 23–25, 2001, at s'Hertogenbosch, The Netherlands. It was great! It was a nice mix of the usual ASA thermoacoustics people with their counterparts in Europe and Asia. It was also a nice mix of the usual work seen in the ASA with other related work in pulse tubes, Stirling cycles, and combustion that we normally do not see. The principal force behind this meeting was Jos Zeegers of Eindhoven University, and he was assisted by Fons de Waele, Charles Schmid, Greg Swift, and myself. There was the feeling at the close of the workshop that we should do this again after a few years.

We decided in Newport Beach to have Socials for Students in Physical Acoustics every other meeting since they can be difficult to organize and expensive to run the way we have been doing them—off the meeting site in a brew pub or tavern. Our social in Chicago had to be cancelled because the tavern at which it was to be held went out of business just before the meeting. We then discussed at the Chicago TC meeting whether we should join in with one of the other two agglomerations of technical committees

that hold socials in the meeting hotel. This would be a lot less work for the organizers, as it is difficult to scope out the meeting area during the TPOM before the main meeting and often we do not have anyone local at the meeting site. But Matt Poese spoke, it seemed for the rest of the students present, that they would much prefer what we were doing at the off site locations. So we agreed to give it another shot at the upcoming Fort Lauderdale meeting.

ROBERT M. KEOLIAN
Chair 1999–2002

Psychological and Physiological Acoustics

As expected, the ASA meeting held in Newport Beach was not well attended by P&P members. There were three podium or poster sessions and no special sessions. This is in keeping with the policy of P&P to deemphasize the fall meeting. Thanks to Bruce Berg for serving as P&P Technical Program Chair and to Monita Chaterjee, Abeer Alwan, and Bob Shannon for their contributions to the organization of the meeting. At the P&P open meeting, the major agenda items were the reports from the Associate Editors and plans for special sessions for Chicago. Five sessions were suggested and the chair indicated that 3–4 would be chosen by polling the members of the Technical Committee. It was also noted that at the Chicago meeting there would be two History lectures, one given by Murray Sachs, the other by Ira Hirsh. The results of a poll of the members of the P&P Technical Committee on whether there should be awards for ''best paper'' and for ''best student presentation'' were reported. All 18 committee members responded: the vote was 16 against, one for, and one neutral. This reaffirms P&P's long-standing opposition to such awards. The reasons for this opposition, which is strenuous, are varied but generally center around the attitude that ASA papers in P&P are already expected to be excellent, that the negative consequences of such awards would more than offset any potential improvement in quality, and that the political and practical aspects of implementing the awards make them unfeasible.

The Chicago meeting was very well attended and there was considerable P&P activity: six full podium or poster sessions, two special sessions, and two history presentations. In addition, there were several sessions that were co-sponsored by P&P, including a special session honoring Harry Levitt and a session on the dynamics of speech production and perception. Thanks to Ann Clock Eddins for organizing the session on neuroimaging and to Ruth Litovsky for the session on behavior and physiology. Both of these P&P special sessions were interesting, well attended, and, by all reports, successful. Murray Sachs presented a fine, scholarly history of physiological acoustics in ASA and Ira Hirsh gave a unique lecture on the history of psychological acoustics in ASA. With so many sessions, avoiding conflict is difficult. Thanks to Stan Sheft and Bill Shofner for organizing the sessions.

The P&P open meeting was somewhat rowdy but accomplished its important goals. There was brief discussion of special sessions for the Pittsburgh meeting (Spring 2002) and six potential sessions were presented to the Technical Council on Friday, June 8. In keeping with P&P policy, there are no special sessions planned for the Fall meetings in Fort Lauderdale (2001) or in Cancun (2002). P&P representatives on ASA committees (Peggy Nelson, ASACOS, and Marjorie Leek, Medals and Awards) described the function of these committees. The new P&P representative on the ASA Membership Committee is Ervin Hafter. He replaces Tino Trahiotis who was thanked for his brief tenure on Membership. Although any Fellow can nominate someone for Fellowship in ASA, P&P requests that nominations for P&P Fellows be sent via our representative on the Membership Committee. This will help assure that this important process is orderly and that no deserving person is overlooked. Although he was terminated from Membership by the P&P Chair for recent lapses, thanks to Dennis McFadden for his role in overseeing the initial stages of implementing this procedure and, far more importantly, for helping to correct P&P's negligence in nominating deserving persons for Fellowship in ASA. Congratulations to the most recent P&P Fellows: Paul Abbas, Soren Bech, Michael Gorga, John Grose, Virginia Richards, and Mario Ruggero.

The P&P Associate Editors gave their reports and there appears to be no major changes in editorial activity. Sid Bacon's term as Associate Editor ends July 1. He was thanked for his excellent service. After again stressing the crucial importance of the Associate Editors, the chair, who also should be terminated for incompetence as well as for tackiness, announced that Neal Viemeister will replace Bacon as Associate Editor.

Six new members of the Technical Committee were elected during the meeting: John Culling, Torsten Dau, Armin Kohlrausch, Robert Lutfi, Christine Mason, and Brian Moore. Their terms begin after the meeting in Spring, 2002. Thanks to the people who agreed to be on the ballot and to the outgoing members, who prepared the ballot: Bob Carlyon (Election Chair), Prudence Allen, David Dolan, Brent Edwards, Nina Kraus, and John Rosowski. The meeting concluded with a nice reception hosted by the Parmly Hearing Institute and its director, Bill Yost. Highlights of the Chicago meeting occurred during the Plenary Session/Awards Ceremony where Bill Hartmann was awarded the Helmholtz–Rayleigh Interdisciplinary Silver Medal and Andrew Oxenham was awarded the R. Bruce Lindsay Award. Bill has been a valuable contributor to P&P and, as his fine acceptance speech suggested, his heart is somewhere within the body of psychoacoustics, probably between the ears. Andrew's Lindsay award, which is for substantial contributions for youngsters under 35, was especially significant for P&P because it attests to the youthful vigor and excellence of P&P.

P&P representation in ASA is very good. Bill Hartmann was anointed President of ASA, Jan Weisenberger became Vice President, Bill Yost is Vice President Elect and is on the Executive Council, and Donna Neff continues her service on the Executive Council. Our congratulations and thanks to all four. Finally, P&P continues with its standard Technical Initiatives, which include travel support for invited speakers, student receptions, and homepage maintenance. Suggestions for innovative uses for these funds, such as workshops, satellite meetings, etc., are welcome (E-mail: nfv@tc.umn.edu).

NEAL F. VIEMEISTER
Chair 1999–2002

Signal Processing in Acoustics

Signal and image processing methods are the technological engines that provide much of the driving impetus to extract the desired information from both simulated and measured acoustical data within each of the individual acoustical application areas represented by the various Technical Committees (TCs). For instance, these applications range from processing animal bioacoustic sounds, structural vibrations, underwater acoustics, speech, biomedical ultrasound, as well as acoustical imaging and tomography. Over this year the Technical Committee on Signal Processing (SP) in Acoustics has sponsored many technical sessions jointly with the TCs and individually on Processing Tutorials discussing critical technologies in signal processing that directly impact many of the TC areas of high interest. The Gallery of Acoustics sponsored and administered annually by SP has become a staple at the ASA meetings. Our Technical Liaisons assure that we are continuously interacting at all levels with the TCs. The upcoming Signal Processing Challenge Problem on the ASA web site along with our own SP web site provides a venue for signal processing colleagues to interact with us through the ASA.

We were elected to the status of a full Technical Committee by the Executive Council based on the recommendation of the Technical Council (11 TCs voted yes, 1 abstention). We are now ''officially'' the TECHNICAL COMMITTEE on Signal Processing in Acoustics with 262 members at the time of the vote. We encourage any signal processors to choose SP as their committee. Since we are now a TC, we have appointed David Havelock as our representative on the Membership Committee, Stan Erhlich as our representative on the Medals and Awards Committee, and David Evans as our representative on the Standards Committee. Also, Dave Havelock, the previous chair of SP, constructed a ''brand new, shiny'' web site for us (with links under ASA—check it out). Brian Ferguson, one of our SP members was elevated to the prestigious grade of ASA Fellow at the Chicago meeting.

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 Technical Committee on Signal Processing in Acoustics (SP) 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 Newport Beach and Chicago meetings.

At the 140th ASA meeting in Newport Beach, California, the SP sponsored or co-sponsored eight sessions: one contributed session and seven special sessions. These sessions were entitled

''Signal Processing Techniques,'' (SP), Chair: David I. Havelock; ''Vibration and Noise in Structures I '' (SA/SP/PA), Chair: Sean F. Wu; ''Vibra-

tion and Noise in Structures II,'' (SA/SP/PA), Chair: Scott D. Sommerfeldt; ''Acoustical Imaging,'' (SP) Chair: David H. Chambers; ''Signal Processing for Speech,'' (SC/SP), Chair: Shrikanth S. Narayanan; ''Acoustic Time Reversal and Acoustic Communications,'' (UW/SP) Chair: David R. Dowling; ''Novel Optical Techniques for Measuring Surface Vibration,'' (PA/N/Stds S2/SP/SA), Co-Chairs: James M. Sabatier and Benjamin A. Bard; ''Blind Deconvolution and Source Separation in Acoustics'' (SP) Chair: Leon Sibul.

The ''Gallery of Acoustics'' sponsored by SP and compiled, set up, and displayed by Randall W. Smith and Preston S. Wilson was one of the highlights of the meeting. Our winner for the Newport Beach meeting was Michel Versluis, Barbara Schmitz, Anna von der Heydt, and Detlef Lohse, University of Twente, The Netherlands with the entry entitled, On the Sound of Snapping Shrimp. There will be galleries set up in the Fort Lauderdale meeting. The Gallery provides ''creative'' acousticians the opportunity to display the results of their efforts in a more creative light.

The 141st ASA meeting in Chicago, Illinois featured more joint SP sponsored/co-sponsored sessions with 14 total sessions: 3 contributed and 11 special. These sessions were ''Signal Processing for Underwater Acoustics,'' (SP) Chair: Brian G. Ferguson; ''Signal Processing for Airborne Acoustics,'' (SP) Chair: Charles F. Gaumond; ''Nondestructive Testing Signal Processing,'' (SP) Chair: David H. Chambers; ''Role of Signal Processing in Understanding Echolocation,'' (AB/SP) Chairs: James A. Simmons; ''Novel Imaging Techniques in Biomedical Ultrasound,'' (BB/SP) Chair: Christy K. Holland; ''Bayesian Signal Processing Approach in Acoustics,'' (SP/UW) Co-Chairs: James V. Candy and Ning Xiang; ''Coded Signals for Acoustics Applications,'' (SP/AA/BB/UW) Co-Chairs: Ning Xiang and Wing T. Chu ''Beamforming and Adaptive Aberration Correction,'' (BB/SP) Chair: T. Douglas Mast; ''Machinery Prognostics,'' (SA/EA/N/SP) Chair: David C. Swanson; ''A Microphone Array for Hearing Aids,'' (SP) DISTINGUISHED LECTURE: Professor Bernard Widrow, Stanford University, Chair: James V. Candy; ''Information Theory Analysis of Animal Acoustic Communication,'' (AB/SP) Chair: John R. Buck; ''Underwater Acoustic Communications I: Algorithm Design and Analysis for Realistic Propagation Conditions,'' (UW/SP) Chair: James C. Presig; ''Underwater Acoustic Communications II: Algorithm Design and Analysis for Realistic Propagation Conditions,'' (UW/SP) Chair: James C. Presig; ''Speech Analysis and Synthesis Techniques and Their Applications to Acoustics,'' (SC/SP) Co-Chairs: Jose A. Diaz and Emily A. Tobey; ''Signal Processing for Auditory and Speech Systems,'' (SP) Chair: Robert Bilger.

We had our first, ''Best Young Presenter Award'' which turned out to be a tie with the winners splitting the cash prize. Our winners were Patrick J. Wolfe for the presentation ''A Bayesian Framework for Perceptually Motivated Audio Signal Enhancement,'' and Corey I. Cheng for the presentation, ''Error Analysis of HRTF's Measured with Complimentary (Golay) Codes.''

We also offered a short course in both Newport Beach and Chicago entitled, ''Applied Acoustical Digital Signal Processing.'' It was presented by James V. Candy and will be offered again in Fort Lauderdale.

In a concerted effort to enhance the communications between the SP and the other Technical Committees, Technical Liaisons (TL) were created for each TC. A Technical Liaison is a member of the SP who represents the interests of the group to other Technical Committees (e.g., Underwater Acoustics). The TL would attend the SP meeting as well as his particular Technical Committee meeting. He would primarily facilitate the CO-SPONSORING of joint sessions with SP (our charter) and the Committee assuring no technical overlap with other sessions. The following SP members are the current Technical Liaisons: Acoustical Oceanography—James Miller; Animal Bioacoustics—Brian Ferguson; Architectural Acoustics—Deborah Grove/Ning Xiang; 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/Dave Chambers; Psychological and Physiological Acoustics—Open; Speech Communication—Jose Diaz; Structural Acoustics and Vibration—Stuart Bolton/Dave Swanson; Underwater Acoustics—Ed Sullivan/Geoff Edelson/Jean-Pierre Hermand.

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 participating in Acoustic 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 turnaround policy for short articles (six pages) or letters. It is strictly complimentary to the normal full-length JASA articles.

The efforts of David Evans on the Standards Committee are also important to the SP. We would also like to recognize our recording secretary, Charles Gaumond, and Dave Havelock and Martin Bartlett for creating, updating, and maintaining our web site, another task that requires much effort but yields little return. Surf to our site and get all of the current tidbits of information and links to other processing web sites. Our web address is http://www.arlut.utexas.edu/asaspweb/

JAMES V. CANDY
Chair 2000–2003

Speech Communication

Dr. Diane Kewley-Port has been elected to Chair the Speech Technical Committee for the next 3 years.

Productivity for the Speech Technical Committee has been excellent 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 Catherine T. Best, Rene Carre, Robert D. Frisina, Megan M. Hodge, Jody E. Kreiman, Charissa R. Lansing, Diane L. Meador, Christopher A. Moore, Shrikanth S. Narayanan, John J. Ohala, Dwayne Paschall, Astrid Schmidt-Nielsen, Samuel A. Seddoh, Anu Sharma, Winifred Strange, Elaine T. Stathopoulos, Emily A. Tobey, Mariko Aoki, Shari R. Baum, Suzanne E. Boyce, Christopher S. Campbell, Randy L. Diehl, David Dorado, Bruce Gerratt, John H. L. Hansen, Jean-Paul Haton, Sarah Hawkins, Mark S. Hedrick, Sun-Ah Jun, Klaus J. Kohler, Nancy S. McGarr, Ian E. Rogers, Robert Ruiz, Caroline L. Smith, Greg S. Turner, Negalapura Viswanath, Carol Y. Espy-Wilson, Marios S. Fourakis, John W. Hawks, William F. Katz, Patricia A. Keating, Paul A. Luce, Deborah M. Rekart, Juergen Schroeter, Janet W. Stack, Mario A. Svirsky, Rosalie M. Uchanski, and Gary G. Weismer. The ex officio members of the Committee are Keith R. Kluender, Anders Lofqvist, Douglas D. O'Shaughnessy, Christopher W. Turner, Maureen L. Stone, and James M. Hillenbrand. If you are interested in becoming a member of the Committee, please contact Diane Kewley-Port.

Web Page Improvements

Bob Port continues to do a great job developing the web page for the Speech Communication Committee (asa.aip.org). 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 Chicago meeting student reception was excellent with approximately 30 students participating in the event. The reception for students at the Newport Beach meeting was ''standing-room'' only with over 50 attendees from Speech Communication, Psychological and Physiological Acoustics, Animal Bioacoustics, and Musical Acoustics.

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 Newport Beach went to Angelique Grosgeorges and the outstanding student presentation in Chicago went to Om D. Deshmukh for his paper on ''Robust speech event detection using strictly temporal information.''

Melissa Epstein from the University of California Los Angeles continues to serve as our representative to the ''ASA Student Council,'' an initiative started by our past 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 ARLO and is now our representative to ASACOS.

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 Ken Grant, Roger Chan, Gary Weismer, and Jim Hillenbrand, to review the selection criteria and mechanics of selection. The subcommittee has revised the application form and streamlined the selection procedures.

Special Initiatives

Peggy Nelson and Sig Soli are preparing a booklet describing the rationale for spending additional monies on acoustically sound classrooms. Any member who has an idea for a special ASA technical initiative in Speech Communication should contact one of the Committee members or Diane Kewley-Port.

Medals, Awards, and Fellows

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

EMILY TOBEY
Chair 1998–2001

Structural Acoustics and Vibration

At the Newport Beach meeting (Fall 2000), the structural acoustics and vibration TC sponsored three special sessions. ''Diagnostics of vibration and noise in structures'' was organized by Sean Wu of Wayne State University, ''Acoustic microsensors'' was organized by Vasu Varadan of Penn State University, and ''Novel optical methods for measuring surface vibration'' (joint with physical acoustics) was organized by James Sabatier of the University of Mississippi. The TC also sponsored a Student Paper Award competition. The first place paper at Newport Beach was given by J. D. Clark from Duke University (Rob Clark–advisor) and the second place paper was given by Nassif Rayess from Wayne State University (Sean Wu–advisor).

There were again three special sessions sponsored by the TC at the Chicago meeting (Spring 2001). ''Complexity and random matrix theory I & II'' were organized by Richard Weaver of the University of Illinois, ''Machinery prognostics'' (joint with signal processing in acoustics, noise, and engineering acoustics) was organized by David Swanson of Penn State University, and ''David G. Crighton memorial session I, II, & III'' (joint with physical acoustics) were organized by Paul Barbone of Boston University. In addition, Yves Berthelot of Georgia Tech University represented our committee admirably in the ''Hot Topics'' session. Yves gave a very nice presentation on ''Hot topics in structural acoustics: Optical imaging of surface vibrations—Some new developments and applications.'' The Student Paper Award competition was again sponsored, and we would like to thank Greg McDaniel of Boston University for his efforts with this competition. The first place paper at Chicago was given by Paulo Alves from the State University of Campinas in Brazil (Jose Arruda–advisor) and the second place paper was given by Sondipon Adhikari of Cambridge University (J. Woodhouse–advisor). In addition, there were special sessions planned in Chicago for the next three meetings that promise to be very interesting.

There has been considerable discussion in ASA and in the SAV TC over the past year regarding ISO TC 108 (dealing with shock and vibration standards). It has been debated whether ASA should continue its support of this standards committee. There has been a strong encouragement from the SAV TC to continue support, and the Executive Council has approved continued support for this year. However, if we wish to see support continue, it would behoove us to work with Susan Blaeser to identify potential organizations who would be willing to offer some financial support for these efforts. We have been encouraged to forward information on potential contributors to Susan at E-mail: sblaeser@aip.org.

Our TC has also been affected by the restructuring of the JASA associate editors. We now have four associate editors for structural acoustics: Jerry Ginsberg, Andrew Norris, Richard Weaver, and Earl Williams. We would like to express our appreciation to Paul Remington and Courtney Burroughs who have served long and well as associate editors for structural acoustics. Finally, Sabih Hayek is our representative for structural acoustics on ASA's Medals and Awards Committee, and Courtney Burroughs is our representative on the Membership Committee, which makes decisions regarding Fellows of the Society. If any of you wish to nominate an individual for any of these honors, please contact Sabih or Courtney.

SCOTT D. SOMMERFELDT
Chair 2000–2003

Underwater Acoustics

The main activities of the Underwater Acoustics Technical Committee (UWTC) centered around the Newport Beach and Chicago meetings.

The UWTC representatives to the Newport Beach Technical Program Organizing Meeting (TPOM) were Kevin Smith and David Dowling. This meeting featured two special sessions on High-Frequency Sediment Acoustics and Associated Sediment Properties and Processes. These sessions were co-sponsored by the Acoustical Oceanography Technical Committee (AOTC) and were chaired by Eric Thorsos and Kevin Williams. Underwater Acoustics also co-sponsored a special session with Signal Processing and Speech Communication on Blind Deconvolution and Source Separation in Acoustics. Leon Sibul was the chairman for this session.

At the Newport Beach meeting, the UWTC joined with the Acoustical Oceanography, Psychological and Physiological Acoustics, Speech Communication, and Signal Processing Committees to sponsor a well-attended student reception. This worked out very nicely—by combining our resources, we were able to have great refreshments and a much livelier crowd. The UWTC also held a student paper competition. The winners were Prunima Ratilal, Allen Reed, and Karim Sabra. The chairman of the judges committee was William Siegmann.

At the Technical Committee meeting in Newport Beach we discussed the change in status of the Interdisciplinary Technical Group on Signal Processing (ITG–SP). The ITG–SP had requested to become a new Technical Committee (TC). Some concern was expressed that this new technical committee would draw members away from the UWTC, similar to what happened when the Acoustical Oceanography TC was formed. Most people felt that the UWTC, the AOTC, and the ITG–SP were coordinating their activities very well and that this new TC would be a benefit to the Society. A majority voted to support the ITG–SP in their application for TC status. It will take some effort on the part of these three technical committees to coordinate their activities.

Joseph Lingevitch and Stewart Glegg represented the UWTC at the Chicago TPOM. This meeting featured a special session on Underwater Acoustic Communications: Algorithm Design and Analysis for Realistic Propagation Conditions. The session was co-sponsored with the Signal Processing TC and was chaired by James Preisig. The UWTC also co-sponsored four other special sessions with a variety of other Technical Committees including Signal Processing, Acoustical Oceanography, Biomedical Ultrasound/Bioresponse to Vibration, and Architectural Acoustics.

As for student activities at Chicago, we co-sponsored a reception with the same group of Technical Committees that got together for Newport Beach. This reception was even better than Newport Beach. The food might not have been as good, but it was a large and lively group. The UWTC has selected Geoffrey Edelmann to represent us on the Student Council. Geoffrey is a student at the University of California, San Diego, working with Bill Kuperman at the Scripps Institution of Oceanography. The UWTC also held a student paper competition in Chicago. The winners were Karim Sabra, Luc Lenain, and Preston Wilson. The chairman of the judges committee

was again William Siegmann. Mohsen Badiey has been nominated to chair this committee for the Fort Lauderdale meeting.

Looking forward to the Fort Lauderdale meeting, we have three special sessions planned. One is being organized by Kevin LePage—the subject is reverberation in shallow water. Ralph Goodman is organizing a session on insights into the evolution of underwater acoustics. He has some distinguished invited speakers already committed to the meeting. Since the UWTC was scheduled to present a historical lecture at Fort Lauderdale, this session will admirably fill that requirement. Finally, Ellen Livingston is organizing a special session on marine mammal regulatory issues. We plan to coordinate this with a similar session being sponsored by the Animal Bioacoustic TC. Also, an ''open meeting'' is planned to discuss what role the ASA might play organizing the discussion of this sensitive subject.

JOHN S. PERKINS
Chair 2000–2003

USA Meetings Calendar

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

2001

4–6 October	Ninth Annual Conference on the Management of the Tinnitus Patient, Iowa City, IA [Rich Tyler, Tel.: 319-356-2471; E-mail: rich-tyler@uiowa.edu; WWW: www.medicine.uiowa.edu/otolaryngology/news/news].
7–10 October	2001 IEEE International Ultrasonics Symposium Joint with World Congress on Ultrasonics, Atlanta, GA [W. O'Brien, Electrical and Computer Engineering, University of Illinois, 405 N. Mathews, Urbana, IL 61801; Fax: 217-244-0105; WWW: www.ieee-uffc.org/2001].
29–31 October	NOISE–CON 01, The 2001 National Conference and Exposition on Noise Control Engineering, Portland, Maine, USA. [Institute of Noise Control Engineering, P.O. Box 3206 Arlington Branch, Poughkeepsie, NY 12603, USA; Tel: 914-462-4006; Fax: 914-462-4006, E-mail: omd@ince.org; WWW: users.aol.com/inceusa/ince.html].
15–18 November	American Speech Language Hearing Association Convention, New Orleans, LA [American Speech-Language-Hearing Association, 10801 Rockville Pike, Rockville, MD 20852; Tel: 888-321-ASHA; E-mail: convention@asha.org; WWW: professional.asha.org/convention/abstracts/welcome.asp].
3–7 December	142nd Meeting of the Acoustical Society of America, Ft. Lauderdale, FL [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

2002

21–23 February	National Hearing Conservation Association Annual Conference, Dallas, TX [NHCA, 9101 E. Kenyon Ave., Ste. 3000, Denver, CO 80237; Tel.: 303-224-9022; Fax: 303-770-1812; E-mail: nhca@gwami.com; WWW: www.hearingconservation.org/index.html].
10–13 March	Annual Meeting of American Institute for Ultrasound in Medicine, Nashville, TN [American Institute of Ultrasound in Medicine, 14750 Sweitzer Lane, Suite 100, Laurel, MD 20707-5906; Tel.: 301-498-4100 or 800-638-5352; Fax: 301-498-4450; E-mail: conv_edu@aium.org; WWW: www.aium.org].
3–7 June	143rd Meeting of the Acoustical Society of America, Pittsburgh, PA [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].
2–6 December	Joint Meeting: 144th Meeting of the Acoustical Society of America, 3rd Iberoamerican Congress of Acoustics and 9th Mexican Congress on Acoustics, Cancun, Mexico [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org/cancun.html].

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

Ordering information: Orders must be paid by check or money order in U.S. funds drawn on a U.S. bank or by Mastercard, Visa, or American Express credit cards. Send orders to Circulation and Fulfillment Division, American Institute of Physics, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2270. Non-U.S. orders add $11.00 per index.

Some indexes are out of print as noted below.

Volumes 1–10, 1929–1938: JASA, and Contemporary Literature, 1937–1939. Classified by subject and indexed by author. Pp. 131. Price: ASA members $5.00; Nonmembers $10.00.

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

Volumes 21–30, 1949–1958: JASA, Contemporary Literature and Patents. Classified by subject and indexed by author and inventor. Pp. 952. Price: ASA members $20.00; Nonmembers $75.00.

Volumes 31–35, 1959–1963: JASA, Contemporary Literature and Patents. Classified by subject and indexed by author and inventor. Pp. 1140. Price: ASA members $20.00; Nonmembers $90.00.

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

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

Volumes 45–54, 1969–1973: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 540. Price: $20.00 (paperbound); ASA members $25.00 (clothbound); Nonmembers $60.00 (clothbound).

Volumes 55–64, 1974–1978: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 816. Price: $20.00 (paperbound); ASA members $25.00 (clothbound); Nonmembers $60.00 (clothbound).

Volumes 65–74, 1979–1983: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 624. Price: ASA members $25.00 (paperbound); Nonmembers $75.00 (clothbound).

Volumes 75–84, 1984–1988: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 625. Price: ASA members $30.00 (paperbound); Nonmembers $80.00 (clothbound).

Volumes 85–94, 1989–1993: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 736. Price: ASA members $30.00 (paperbound); Nonmembers $80.00 (clothbound).

Volumes 95–104, 1994–1998: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. Price: ASA members $40.00 (paperbound); Nonmembers $90.00 (clothbound).

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

The Technical and Administrative Committees listed below have been appointed by the Executive Council. These appointments, with such changes as may be made by the President from time to time, will be in effect until the Spring meeting of the Society in 2002.

Technical Committees 2001–2002

Acoustical Oceanography

	Term to
Peter F. Worcester, *Chair* to 2004	2004
Ching-Sang Chiu	2004
Bruce D. Cornuelle	2004
Orest I. Diachok	2004
Kenneth G. Foote	2004

Andrew M. Forbes	2004
Boris G. Katsnelson	2004
James F. Lynch	2004
Jungyul Na	2004
Gopu R. Potty	2004
Stephen A. Reynolds	2004
Mark V. Trevorrow	2004
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
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
David M. Rubenstein	2002

Ex officio:
David L. Bradley, Associate Editor of JASA
John C. Burgess, Associate Editor of JASA
Stanley L. Chin-Bing, Associate Editor of JASA
William L. Siegmann, Associated 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
Whitlow W. L. Au	2004
Robert J. Dooling	2004
Robert Hickling	2004
David A. Helweg	2004
Peter M. Narins	2004
Andrea M. Simmons	2004
James A. Simmons	2004
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
D. Vance 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

Ex officio:
Whitlow W. L. Au, Associate Editor of JASA and member of Medals and Awards Committee
David L. Bradley, Associate Editor of JASA
Stanley A. Chin-Bing, Associate Editor of JASA
William L. Siegmann, Associate Editor of JASA
Charles R. Greene, member of Membership Committee
Ann E. Bowles, member of ASACOS

Architectural Acoustics

	Term to
K. Anthony Hoover, *Chair* to	2004
Bennett M. Brooks	2004
Steven M. Brown	2004
Richard H. Campbell	2004
M. David Egan	2004
Jesse J. Ehnert	2004
Donna A. Ellis	2004
Ronald A. Freiheit	2004
Richard M. Guernsey	2004
Mark A. Holden	2004
Dana S. Hougland	2004
Jerald R. Hyde	2004
David W. Kahn	2004
Martha M. Larson	2004
Gary S. Madaras	2004
Charles T. Moritz	2004
Paul B. Ostergaard	2004
Dennis A. Paoletti	2004
Neil A. Shaw	2004
Richard H. Talaske	2004
Ewart A. Wetherill	2004
George E. Winzer	2004
Michael R. Yantis	2004
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

Ex officio:
Jerry H. Ginsberg, Associate Editor of JASA
Mendel Kleiner, Associate Editor of JASA
Andrew N. Norris, Associate Editor of JASA
Richard L. Weaver, Associate Editor of JASA
Earl R. Williams, 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
Diane Dalecki	2004
J. Brian Fowlkes	2004
Seyed H. R. Hosseini	2004
Kullervo Hynynen	2004
James F. Greenleaf	2004
T. Douglas Mast	2004
Thomas J. Matula	2004
Douglas L. Miller	2004
Pierre D. Mourad	2004
Sandra I. Polichik	2004
Ronald A. Roy	2004
Thomas J. Royston	2004
Pei Zhong	2004
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

Ex officio:
Laurel H. Carney, Associate Editor of JASA
Floyd Dunn, Associate Editor of JASA
Neal F. Viemeister, 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
Mahlon D. Burkhard	2004
James Christoff	2004
Fernando Garcia-Osuna	2004
Charles S. Hayden	2004
Dennis F. Jones	2004
Jan F. Lindberg	2004
Yushieh Ma	2004
Elizabeth A. McLaughlin	2004
Alan Powell	2004
Roger T. Richards	2004
Harold C. Robinson	2004
Kenneth D. Rolt	2004
Neil A. Shaw	2004
James Tressler	2004
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

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

Musical Acoustics

	Term to
James P. Cottingham, *Chair* to 2002	2002
R. Dean Ayers	2004
Judith C. Brown	2004
Courtney B. Burroughs	2004
Robert D. Collier	2004
George F. Emerson	2004
Bozena Kostek	2004
Barry Larkin	2004
Daniel O. Ludwigsen	2004
Thomas D. Rossing	2004
Julius O. Smith	2004
William J. Strong	2004
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
Steven E. McAdams	2002
Gary P. Scavone	2002
Gabriel Weinreich	2002

Ex officio:
R. Dean Ayers, Associate Editor of JASA
John C. Burgess, Associate Editor of JASA
Stanley L. Ehrlich, Associate Editor of JASA
Stephen E. McAdams, Associate Editor of JASA
James W. Beauchamp, member of Medals and Awards Committee
Uwe J. Hansen, member of Membership Committee
Ian M. Lindevald, member of ASACOS

Noise

	Term to
Bennett M. Brooks, *Chair* to 2003	2003
Martin Alexander	2004
John P. Barry	2004
Leo L. Beranek	2004
Arno S. Bommer	2004
Anthony J. Brammer	2004
James O. Buntin	2004
Robert J. Cook	2004
Jim R. Cummins	2004
Kenneth A. Cunefare	2004
Joseph M. Cuschieri	2004
Paul R. Donavan	2004
William D. Gallagher	2004
Gerald C. Lauchle	2004
Anna Maria Monslave	2004
Miomir Mijic	2004
Thomas Norris	2004
John P. Seiler	2004
Noral D. Stewart	2004
Michael R. Stinson	2004
Alice H. Suter	2004
Louis C. Sutherland	2004
Jiri Tichy	2004
Dennis Walton	2004
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

Ex officio:
Stanley L. Ehrlich, Associate Editor of JASA
Michael R. Stinson, Associate Editor of JASA and 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
David T. Blackstock	2004
David A. Brown	2004
John A. Burkhardt	2004
Kerry W. Commander	2004
Bruce C. Denardo	2004
Logan E. Hargrove	2004
D. Kent Lewis	2004
Julian D. Maynard	2004
George Mozurkewich	2004
Lev. A. Ostrovsky	2004
Andrea Prosperetti	2004
Neil A. Shaw	2004
Victor W. Sparrow	2004
Richard Stern	2004
Roger M. Waxler	2004
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

Ex officio:
Yves H. Berthelot, Associate Editor of JASA
Dale E. Chimenti, Associate Editor of JASA
Floyd Dunn, Associate Editor of JASA
Mark F. Hamilton, Associate Editor of JASA
Michael S. Howe, Associate Editor of JASA
Steven G. Kargl, Associate Editor of JASA
Martin Ochmann, Associate Editor of JASA
Ronald A. Roy, Associate Editor of JASA
Louis C. Sutherland, Associate Editor of JASA
Lonny L. Thompson, 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
Bruce G. Berg	2004
Joan M. Besing	2004
Monita Chatterjee	2004
Robert D. Frisina	2004
Andrew J. Oxenham	2004
Pierre L. Divenyi	2003
Richard L. Freyman	2003
Wesley L. Grantham	2003
Virginia M. Richards	2003
Barbara G. 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

Ex officio:
Leslie R. Bernstein, Associate Editor of JASA
Laurel H. Carney, Associate Editor of JASA
Marjorie R. Leek, Associate Editor of JASA and member of Medal and Awards Committee
Brenda L. Lonsbury-Martin, Associate Editor of JASA
Neal F. Viemeister, Associate Editor of JASA
Ervin R. Hafter, member of Membership Committee
Peggy B. Nelson, member of ASACOS

Signal Processing in Acoustics

	Term to
James V. Candy, *Chair* to 2003	2003
Leon Cohen	2004
Jose A. Diaz	2004
Deborah M. Grove	2004
David I. Havelock	2004
Alan W. Meyer	2004
Richard J. Ruhala	2004
Roger W. Schwenke	2004
David C. Swanson	2004
Preston S. Wilson	2004
Geoffrey S. Edelson	2003
Stanley L. Ehrlich	2003

Brian Ferguson	2003
William M. Hartmann	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 W. Smith	2002

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

Speech Communication

	Term to
Diane J. Kewley-Port, *Chair* to	2004
Catherine T. Best	2004
Rene Carre	2004
Robert D. Frisina	2004
Megan M. Hodge	2004
Jody E. Kreiman	2004
Charissa R. Lansing	2004
Diane L. Meador	2004
Christopher A. Moore	2004
Shrikanth S. Narayanan	2004
John J. Ohala	2004
Dwayne Paschall	2004
Astrid Schmidt-Nielsen	2004
Samuel A. Seddoh	2004
Anu Sharma	2004
Winifred Strange	2004
Elaine T. Stathopoulos	2004
Emily A. Tobey	2004
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
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

Ex officio:
Keith R. Kluender, Associate Editor of JASA
Anders Lofqvist, Associate Editor of JASA
Douglas D. O'Shaughnessy, Associate Editor of JASA
Christopher W. Turner, Associate Editor of JASA
Gary R. Weismer, member of Medals and Awards Committee
Maureen L. Stone, member of Membership Committee
James M. Hillenbrand, member of ASACOS

Structural Acoustics and Vibration

	Term to
Scott D. Sommerfeldt, *Chair* to	2003
Paul E. Barbone	2004
Jeffrey E. Boisvert	2004
Robert L. Clark	2004
Joseph M. Cuschieri	2004
Jerry H. Ginsberg	2004
John A. Fahnline	2004
Stephen A. Hambric	2004
Peter C. Herdic	2004
Harry Himelblau	2004
Timothy W. Leishman	2004
Jerome E. Manning	2004
Philip L. Marston	2004
James G. McDaniel	2004
Douglas M. Photiadis	2004
Karl M. Reichard	2004
Angie Sarkissian	2004
Victor W. Sparrow	2004
Eric E. Ungar	2004
Nickolas Vlahopoulos	2004
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

Ex officio:
Jerry H. Ginsberg, Associate Editor of JASA
Andrew N. Norris, Associate Editor of JASA
Richard L. Weaver, Associate Editor of JASA

Earl R. Williams, Associate Editor of JASA
Courtney B. Burroughs, member of Membership Committee
Sabih I. Hayek, member of Medals and Awards Committee
Louis A. Herstein, member of ASACOS

Underwater Acoustics

	Term to
John S. Perkins, *Chair* to 2003	2003
Ahmad T. Abawi	2004
Michael G. Brown	2004
Dennis B. Creamer	2004
Christian P. de Moustier	2004
Stanley E. Dosso	2004
George V. Frisk	2004
Stewart A. L. Glegg	2004
Joseph F. Lingevitch	2004
Nicholas C. Makris	2004
Zoi-Heleni Michalopoulou	2004
Marshall H. Orr	2004
Gregory J. Orris	2004
James C. Preisig	2004
Daniel Rouseff	2004
William L. Siegmann	2004
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
Anatoliy N. Ivakin	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
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
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

Ex officio:
David L. Bradley, Associate Editor of JASA and member of Medals and Awards Committee
John C. Burgess, Associate Editor of JASA
Stanley A. Chin-Bing, Associate Editor of JASA
William L. Siegmann, Associate Editor of JASA
Peter H. Rogers, member of Membership Committee
Arnie L. Van Buren, member of ASACOS

Administrative Committees 2001–2002

Archives and History

	Term to
Henry E. Bass, *Chair* to 2004	2004
David T. Blackstock	2004
William W. Lang	2004
Julian D. Maynard	2004
Rosalie M. Uchanski	2004
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

Books

	Term to
Mohsen Badiey, *Chair* to 2002	2002
Stanley L. Chin-Bing	2004
Gordon E. Martin	2004
Victor W. Sparrow	2004
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
Emily A. Tobey	2002

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

Education in Acoustics

	Term to
Uwe J. Hansen, *Chair* to 2003	2003
David A. Brown	2004
Robert D. Collier	2004
Corinne M. Darvennes	2004
Margaritis S. Fourakis	2004

Carole E. Gelfer	2004
Douglas R. Jones	2004
Sharon Y. Manuel	2004
Philip L. Marston	2004
Ralph Muehleisen	2004
Andrew A. Piacsek	2004
Daniel R. Raichel	2004
Sally G. Revoile	2004
Thomas D. Rossing	2004
Ronald A. Roy	2004
Dawn R. Schuette	2004
Scott D. Sommerfeldt	2004
William Thompson, Jr.	2004
Robert A. Walkling	2004
George S.K. Wong	2004
William 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

Ethics and Grievances

	Term to
Robert E. Apfel, *Chair* to 2003	2003
Mardi C. Hastings	2004
David Feit	2004
William J. Cavanaugh	2003

Investments

	Term to
John V. Bouyoucos, *Chair* to 2002	2002
Ira J. Hirsh, *Chair* to 2004	2004
Lawrence A. Crum	2004
Ira Dyer	2002

David Feit, Treasurer, *ex officio*
Henry O'Hara, *ex officio*

Long Range Planning

	Term to
Fredericka Bell-Berti, *Chair* to 2002	2002
Paul A. Baxley	2004
Kenneth E. Gilbert	2004
Brigitte Schulte-Fortkamp	2004
Scott D. Sommerfeldt	2004
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

Richard Stern, President Elect, *ex officio*

Medals and Awards

		Term to
D. Vance Holliday, *Chair* to 2002		2002
James W. Beauchamp	Musical Acoustics	2004
Michael J. Buckingham	Acoustical Oceanography	2004
Mahlon D. Burkhard	Engineering Acoustics	2004
Stanley L. Ehrlich	Signal Processing in Acoustics	2004
Gary G. Weismer	Speech Communication	2004
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	Noise	2003
Steven M. Brown	Architectural Acoustics	2002
David L. Bradley	Underwater Acoustics	2002
Whitlow W. L. Au	Animal Bioacoustics	2002
Sabih I. Hayek	Structural Acoustics and Vibration	2002

Meetings—June 2001–December 2001

Dana S. Hougland, *Chair* to 2002
Sergio Beristain, Fall 2003, Cancun
Yves H. Berthelot, Spring 2000, Atlanta
Courtney B. Burroughs, Spring 2002, Pittsburgh
Joseph M. Cuschieri, Fall 2001, Ft. Lauderdale
Samir N. Y. Gerges, Fall 2003, Cancun
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
Janet M. Weisenberger, Vice President
James E. West, Fall 2003, Cancun
William A. Yost, Vice President Elect

Membership

		Term to
Joseph W. Dickey, *Chair* to 2003		2003
Christopher Feuillade	Acoustical Oceanography	2004
Steven L. Garrett	Physical Acoustics	2004
David I. Havelock	Signal Processing in Acoustics	2004
Peter H. Rogers	Underwater Acoustics	2004
Gregory C. Tocci	Architectural Acoustics	2004
Uwe J. Hansen	Musical Acoustics	2003
Thomas R. Howarth	Engineering Acoustics	2003
Charles R. Greene	Animal Bioacoustics	2003
Maureen Stone	Speech Communication	2003
Ervin R. Hafter	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

Public Relations

	Term to
Paul A. Baxley, *Chair* to 2003	2003
E. Carr Everbach	2004
Charles Gaumond	2004
Christy K. Holland	2004
David Lubman	2004
Andrew A. Piacsek	2004
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

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
David T. Blackstock	2004
Mark F. Hamilton	2004
Patricia K. Kuhl	2004
Alan Powell	2004
Sigfrid D. Soli	2004
Jont B. Allen	2003
Robert C. Bilger	2002
James F. Lynch	2002
James H. Miller	2002
Allan J. Zuckerwar	2002

Richard Stern, President-Elect, *ex officio*
Allan D. Pierce, Editor-in-Chief, *ex officio*

Regional Chapters

Roger L. Richards, *Chair* to 2004

Elmer L. Hixson	Austin
Angelo J. Campanella	Central Ohio
Dean E Capone	Central Pennsylvania
John W. Kopec	Chicago
Ernest M. Weiler	Cincinnati
Edwin H. Toothman	Delaware Valley
Gary W. Siebein	Florida
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 A. 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
William J. Cavanaugh	2004
Floyd Dunn	2004
Elaine Moran	2003
Charles E. Schmid	2003
William M. Hartmann	2002
Richard H. Lyon	2002

Prizes and Special Fellowships

	Term to
Wayne M. Wright, *Chair* to 2002	2002
Anthony A. Atchley	2004

Constantine Trahiotis	2004
Fredericka Bell-Berti	2003
James E. West	2003
Uwe J. Hansen	2002

Standards

Executive Committee
Daniel L. Johnson, Chair (Standards Director)
Paul D. Schomer, Vice Chair
Susan B. Blaeser, 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
Janet M. Weisenberger, 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
David J. Evans, Signal Processing in Acoustics
James M. Hillenbrand, 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)
Susan B. Blaeser
George S. K. Wong

Tutorials

	Term to
Joseph Pope, *Chair* to 2002	2002
Ann E. Bowles	2004
Gary W. Elko	2004
Beverly A. Wright	2004
Fredericka Bell-Berti	2003
Robin O. Cleveland	2003
George V. Frisk	2003
Yves H. Berthelot	2002
Uwe J. Hansen	2002
Ralph T. Muehleisen	2002

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

Women in Acoustics

	Term to
Nancy S. Timmerman, *Co-chair*	2003
Lawrence A. Crum	2004
Martha M. Larson	2004
Penelope Menounou	2004
Sandra L. Poliachik	2004
Brigitte Schulte-Fortkamp	2004
Elvira B. Viveiros	2004
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

William A. Yost, *ex officio* as Vice President Elect

JASA Editorial Board

June 2004	
W. W. L. Au, Bioacoustics—Animal	2004
Y. H. Berthelot, Ultrasonics, Physical Effects of Sound	2004
J. C. Burgess, Acoustic Signal Processing	2004
D. E. Chimenti, General Linear Acoustics	2004
F. Dunn, Bioacoustics—Biomedical	2004
J. H. Ginsberg, Structural Acoustics and Vibration	2004
M. F. Hamilton, Nonlinear Acoustics	2004
M. S. Howe, Aeroacoustics	2004
M. Kleiner, Architectural Acoustics	2004
S. E. McAdams, Music and Musical Acoustics	2004
M. Ochmann, General Linear Acoustics	2004
D. D. O'Shaughnessy, Speech Processing	2004
W. L. Siegmann, Underwater Acoustics	2004
L. C. Sutherland, Aeroacoustics	2004
L. L. Thompson, General Linear Acoustics	2004
N. F. Viemiester, Psychological Acoustics	2004
R. L. Weaver, Structural Acoustics and Vibration	2004
E. R. Williams, Structural Acoustics and Vibration	2004
June 2003	
R. D. Ayers, Music and Musical Instruments	2003
L. E. Bernstein, Psychological Acoustics	2003
D. L. Bradley, Underwater Sound	2003

L. H. Carney, Physiological Acoustics	2003
S. A. Chin-Bing, Underwater Sound	2003
S. A. Kargl, Ultrasonics, Physical Effects of Sound	2003
A. Lofqvist, Speech Production	2003
B. L. Lonsbury-Martin, Physiological Acoustics	2003
J. J. McCoy, Mathematical Acoustics	2003
E. Moran, Acoustical News—USA	2003
T. D. Rossing, Education in Acoustics	2003
R. A. Roy, Ultrasonics, Physical Effects of Sound	2003
V. W. Sparrow, Computational Acoustics	2003
R. Stern, Electronic Archives and References	2003
M. R. Stinson, Noise, Its Effects and Control	2003

June 2002

S. B. Blaeser, Acoustical News—Standards	2002
S. L. Ehrlich, Applied Acoustics; Transduction; Acoustical Measurements	2002
K. R. Kluender, Speech Perception	2002
P. L. Marston, Acoustical Reviews—Books	2002
W. G. Mayer, Acoustical News—International	2002
A. N. Norris, Structural Acoustics and Vibration	2002
D. L. Rice, Acoustical Reviews—Patents	2002
C. W. Turner, Speech Perception	2002
G. S. K. Wong, Acoustical Standards News	2002

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 Emergence of the Speech Capacity

D. Kimbrough Oller

Lawrence Erlbaum Associates, Mahwah, New Jersey, 2000. 441 pp. Price: $99.95 hardcover ($39.95 softcover). ISBN: 0805826289 hc (0805826297 sc).

Since ancient times, scholars have puzzled over the question of how speech develops in children. There are several recent books on this venerable topic, including two that should be mentioned here: *The Discovery of Spoken Language* by Peter W. Jusczyk (2000) and *How Language Comes to Children* by Benedicte de Boysson-Bardies (1999). D. Kimbrough Oller is one of the major contemporary contributors to the study of infant vocalizations and their role in the development of speech. His publications, spanning more than 20 years, have introduced important methodologies and theoretical concepts that have helped to shape the current understanding in this field of study. This book of 15 chapters is an extended discussion of Oller's theory of *infraphonology*. This term is not so much defined in the book as it is explicated over several pages to show how this perspective illuminates some longstanding problems in relating infant vocalizations, such as babble, to the development of the spoken language. Oller likens the theoretical promise that infraphonology holds for the study of human vocal development to the roles that atomic physics and thermodynamics have had in chemistry, and embryology and genetics have had in botany. This is a tall order, as the various ''infrastructural systems'' that Oller identifies in the physical and biological sciences are supported by a great deal of experimental data and they have played central roles in their respective scientific arenas.

If infraphonology meets Oller's expectations, it would indeed be a giant advance in the ontogeny and phylogeny of speech. The most succinct definition of infraphonology that appears in this book is found in an illustration that describes it in these words: ''specifies principles generating the entire class of potential well-formed operational units and specifies properties of utilization and function of such units'' (p. 12). The example *par excellence* of such an operational unit is the canonical syllable. Toward the end of the book, when Oller challenges himself to prove the innovative value of infraphonology, he points to this unit as the prime evidence. He defines the ''notion of canonical syllable'' as ''a vocal sound sequence consisting of a smoothly voiced nucleus or vowel-like sound, produced with the mouth open in any of a wide variety of possible postures, along with a quick movement of the mouth, either opening to the nucleus or closing from it'' (p. 62). A more elaborated technical definition of about two pages in length is given later, but even this longer discussion relies on concepts that are rather loosely introduced.

In the early chapters, Oller explains why the ''myths of babbling and the tradition of transcription'' should give way to an approach that emphasizes protophones. True to the general plan in this book, Oller does not offer a succinct definition of a protophone but rather contrasts it with other concepts in a lengthy discussion and a table of descriptors. According to the index, the term protophone is defined as ''categories of vocalization'' on pp. 193–194 and as a ''signaling value'' on pp. 197–200. But nowhere in these five pages does Oller actually define what a protophone is. Although one is eventually rewarded by a careful reading of Oller's treatment of this concept, some readers might prefer a more deliberate effort at formal and explicit definitions of at least the major terms. Comparative discussions surely have a heuristic value, but definitions help to sharpen distinctions and identify the primary attributes of a term.

The ensuing chapters discuss infraphonology and protophones in more detail, delineate canalized patterns of babbling, consider primate vocalizations according to infrastructural concepts, propose stages of vocal evolution in humans, and compare fixed vocal signals across humans and other modern primate species. The intellectual embrace of the book is considerable, and the concept of infraphonology provides a satisfactory coherence to its ambitious scope. There is much to be learned from Oller's reflections and proposals even if one does not accept all of his assumptions and conclusions.

The book is devoted almost entirely to a presentation of infraphonology rather than to a systematic appraisal of opposing points of view. Occasionally, however, Oller refers to competing theories or perspectives. For example, he contends that a focus on the purely physical parameters of vocalization (e.g., fundamental frequency, amplitude, resonance peak) will not lead to valuable interpretations of human babbling or nonhuman vocalizations because the physical measures in themselves have no intrinsic basis for interpretation. In making this argument, he neglects the role of self-organizational processes in which a system so endowed can recognize and exploit regularities in environmental stimuli, which is a topic of much interest in contemporary research on the development of speech and language. Some physical properties of speech that are emphasized in other contemporary accounts of speech development are given short shrift in Oller's book. For example, Oller gives only passing mention to the topic of prosody, including rhythm, intonation, and stress patterning. Of these, only rhythm receives a serious, albeit brief, commentary. Interestingly, both of the other two books mentioned earlier in this review give prosody a major place in the earliest stages of language learning. Reading all three books is recommended to those who want an overview of contemporary thinking about the way in which infants learn spoken language.

A common approach to the study of human communication is to compare and contrast it with the communication systems of nonhuman species. Oller follows this popular pattern, devoting a considerable part of this book to this topic, but with a studiously primate fixation. Relatively little is said about other vocalizing species, including songbirds, that have often been taken as an important animal correlate (if not model) of human vocalization. In this sense, Oller joins those who fully expect that human communication has some vital precursors in nonhuman primates.

One of the fresh contributions in Oller's book is his delineation of 18 infrastructural properties of communication, with special reference to the communicative functions in humans and nonhuman primates. Taking some inspiration from the famous linguist Charles Hockett, who wrote an influential treatise on the design features for language, Oller works toward an improved description in which the design features are both hierarchical and dimensional (i.e., not binary) and both more elaborated and more cognitive than Hockett's. Each of the 18 infrastructural properties is examined in some detail in a chapter that runs to 100 pages. This chapter has a pivotal position between the first eleven chapters that describe the vocalizations of human infants and primates and the final three chapters that develop a theoretical perspective on vocal evolution, fixed vocal signals, and possible directions in infraphonological theory.

Oller's book is recommended to both specialists and nonspecialists. It contains a wealth of ideas that appeal to broad and narrow interests alike. Fortunately, it is not so technical as to frustrate those who are new to this field of study. But an effort to define critical terms succinctly and clearly

would have helped to make the way easier for all readers and perhaps would ensure a smoother launching of the theory of infraphonology.

RAY D. KENT
Waisman Center
University of Wisconson–Madison,
Madison, Wisconsin 53505-2280

Fundamentals of Modeling and Analyzing Engineering Systems

Philip D. Cha, James J. Rosenberg, and Clive L. Dym

Cambridge University Press, New York, 2000.
466 pp. Price: $100.00 hard cover ISBN: 0-521-59443X; $44.00 soft cover ISBN: 0-521-59463-4.

The focus of the text is on lumped parameter modeling techniques and the analysis of linear time-invariant (LTI) single input output (SISO) systems. State-space concepts and feedback control of SISO systems are also covered but briefly. The authors state that their target audience is sophomores in engineering. The material on LTI SISO systems is no doubt directed at students at this level.

The exposition of the linear-graph approach to model and analyze LTI systems cannot be faulted. In fact the approach is at the heart of the discipline of Acoustics. The textbook runs for 466 pages and is split into 10 chapters. The material in the text is similar to that covered in *Introduction to System Dynamics* by J. L. Shearer, A. T. Murphy, and H. H. Richardson, Addison–Wesley, 1967 or more recently a book by the same title by D. Rowell and D. Wormley, Prentice–Hall, 1996. Both books follow the linear graph approach. The gold standard for this material is Harry F. Olson's (ASA Silver medallist), *Dynamical Analogies*, Van Norstrand, 1943.

Chapter 1 outlines the fundamental concepts of modeling and relationship of the developed models to ordinary differential equations. In addition the notion of a linear operator is introduced albeit without introducing mathematical formality.

The causality and shift invariance are not addressed at the outset. This state-of-affairs forces the authors to backtrack in later chapters. The governing laws of system motion are addressed making use of a control volume analogy akin to the Reynolds transport theorem.

Chapter 2 examines lumped-parameter modeling of ideal translational mechanical and ideal electrical systems. For electrical systems KCL and KVL are introduced as the laws governing interconnection of electrical elements. For mechanical systems the application of Newton's second law in conjunction with free-boundary diagrams are used. D'Alembert's principle is used to introduce the nodal force balance concept. However I believe its introduction does not enhance the development.

In Chapter 3 generalized lumped-parameter elements are discussed and ''across'' and ''through'' variables are defined. As a departure from the choice of power variables, the authors have chosen to use force as the through variable and displacement rather than the velocity as the across variable. This choice has consequences. For example the interconnection of a mechanical system to an electrical system using an ideal transformer is problematic if one uses the displacement as the across variable. The importance of the inertial reference frame in describing the motion of a mass in a mechanical system and its impact on the choice of the across variable is not clearly discussed. Later in the chapter the authors are forced to introduce this concept in order to develop lumped-parameter models of rotational systems. Fluid and thermal systems are also cast in terms of a lumped-parameter element formalism.

Chapter 4 examines the solution methods for first-order system models. Of particular note is the presentation of the solution of unforced and forced response for said systems. Without introducing the shift invariance the authors have to rely on what I consider a weak development of the convolution theorem to describe the response of a first-order system to an arbitrary input. An exposition in a similar vein is outlined in Chapter 5 for second-order systems.

In Chapter 6 the Laplace transform is introduced. This material should have been introduced prior to Chapter 4. The authors outline the properties of unilateral Laplace transforms in a traditional manner. However in keeping with the linear system vein of the book they could have mentioned that the Laplace transform is a linear operator. The inversion of the transforms is done using partial fraction expansions and table look up. This is a common approach in most texts at this level.

In Chapter 7 the frequency response of LTI systems is examined. The relationship of the response in the Laplace domain and the sinusoidal state response is defined. The concept of the transfer function is introduced and the Bode plot techniques are presented. I believe that Chapter 7 is a natural break point in the text.

I believe the material covered in Chapters 8–10 is better covered in many of the textbooks on classical control theory.

In Chapter 8 state space formulation of LTI systems is briefly presented. The relationship of ordinary differential equations to the state equations is examined and a brief introduction to eigenvalues and eigenvectors is given. I found the presentation lacking in depth and detail for a beginning student.

Chapter 9 introduces pole-zero plots and the role of dominant poles in the transient response of LTI systems.

Chapter 10, the final chapter, deals with feedback systems. Block diagrams are examined. Frequency response based methods such as the phase and gain margin techniques are examined and applied to the design of controllers.

In summary the material in this text is classical in nature. However the presentation style is not inspired and relies on examples and figures to make points of conceptual significance. Some sections read more like a workbook than a textbook. I believe the text would make a reasonable supplementary text. It may be used as a primary text if the instructor augments the presentation given in the text.

CHARLES THOMPSON
Department of Electrical and Computer Engineering
University of Massachusetts Lowell
Lowell, Massachusetts 01854

BOOKS RECEIVED

Dictionary of Acoustics. Christopher L. Morfey (Editor). Academic Press, London, 2001. 430 pp. $64.95 *hc*. ISBN: 0125069405

Handbook of Elastic Properties of Solids, Liquids, and Gases. M. Levy, H. E. Bass, and R. R. Stern (Editors-in-Chief). Academic Press, San Diego, 2000. Four volume set. 1728 pp. $1500.00 *hc*. ISBN: 0124457606

Acoustic and Electromagnetic Equations: Integral Representations for Harmonic Problems. Jean-Claude Nedelec. Springer-Verlag, New York, 2001. 325 pp. $64.95 *hc*. ISBN: 0387951555

Boundary Element Acoustics: Fundamentals and Computer Codes. T. W. Wu (Editor). WIT Press, Billerica, MA, 2001. 256 pp. $149.00. ISBN: 1853125709

Boundary Elements in Acoustics: Advances and Applications. O. von Estorff (Editor). WIT Press, Billerica, MA, 2000. 488 pp. $246.00 *hc*. ISBN: 1853125563

Hearing Aids. Harvey Dillon. Thieme Medical Publishers, New York, 2001. 522 pp. $59.00 *pb*. ISBN: 1588900525

Microphone Arrays: Techniques and Applications. M. Brandstein and D. Ward (Editors). Springer-Verlag, New York, 2001. 398 pp. $99.00 *hc*. ISBN: 3540419535

Advances in Network and Acoustic Echo Cancellation. J. Benesty, T. Gansler, D. R. Morgan, M. M. Sondhi, and S. L. Gay (Editors). Springer-Verlag, New York, 2001. 234 pp. $69.95 *hc*. ISBN: 3540417214

Nonlinear and Nonstationary Signal Processing. W. J. Fitzgerald, R. L. Smith, A. T. Walden, and P. C. Young (Editors). Cambridge University Press, New York, 2001. $95.00 *hc*. ISBN: 0521800447

Strain Solitons in Solids and How to Construct Them. Alexander M. Samsonov. CRC Press, Boca Raton, FL, 2001. 248 pp. $69.95 *hc*. ISBN: 0849306841

Structure-Borne and Flow Noise Reductions. Sung-Hwan Ko, Sangwoo Pyo, and Woojae Seong. Seoul National University Press, Seoul, Korea, 2001. 327 pp. *hc*. ISBN: 8952102584

The Energy of Nature. E. C. Pielou. University of Chicago Press, 2001. 256 pp. $25.00 *hc*. ISBN: 0226668061

Turbulence Structure Vortex Dynamics. J. C. R. Hunt and J. C. Vassilicos. Cambridge University Press, New York, 2001. 304 pp. $80.00 *hc*. ISBN: 0521781310

Independent Component Analysis: Principles and Practice. S. Roberts and R. Everson (Editors). Cambridge University Press, New York, 2001. 352 pp. $69.95 *hc*. ISBN: 0521792983

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
DAVID PREVES, *Songbird Hearing, Inc., 5 Cedar Brook Drive, Cranbury, New Jersey 08512*
WILLIAM THOMPSON, JR., *601 Glenn Road, State College, Pennsylvania 16803*

6,178,140

43.20.Rz METHOD FOR THREE-DIMENSIONAL BEAMFORMING IN DIRECTION FINDING SYSTEMS

Heiko Schlieter, assignor to L-3 Communications Elac Nautik GmbH
23 January 2001 (Class 367/103); filed in Germany 19 May 1998

The outputs of identical planar groups of transducers that lie on opposite sides of two orthogonal axes are combined in various ways that are said to result in improved side lobe suppression and enhanced beam focusing.—WT

6,181,643

43.30.Vh INTERFEROMETER WITH A SINGLE PROJECTOR ARRAY AND A SINGLE RECEIVER ARRAY

Richard Ove Nielsen, assignor to The Boeing Company
30 January 2001 (Class 367/105); filed 19 July 1999

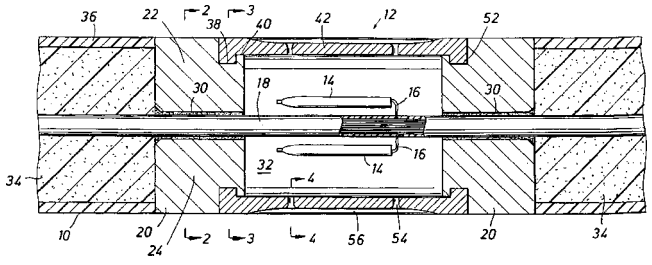
An interferometer location determination sonar system is realized using a single projector array, rather than two distinct projector arrays, plus a single receiver array, by functionally partitioning the projector array into two parts, wherein some elements of the array may belong to both parts, but such that there is a nonzero separation distance between the acoustic centers of the two parts. The echo returns sensed by the receiver are suitably processed to provide location information.—WT

6,188,646

43.30.Yj HYDROPHONE CARRIER

John Luscombe *et al.*, assignors to Syntron, Incorporated
13 February 2001 (Class 367/154); filed 29 March 1999

One element of a multielement towed seismic streamer is shown in cross section. Each element may contain a number of transducer elements **14** (two such are shown in this figure). These transducer elements may be realized as a pair of piezoelectric disks or plates mounted on opposite ends of some support structure and housed within the oil-filled tube structure **14**. The hydrophone assemblies are positioned within a free-flooded cylindrical chamber **32**, preferably made of titanium. Communication with the surrounding acoustic medium is accomplished via a series of openings **54** in the outer wall of the chamber. Central strength member **18** supports adjacent elements and spacers **20** and **34** as well as housing the electrical leads **16** from each hydrophone. Potting material **30** seals the openings where central member **18** passes through the ends of chamber **32**. Other embodiments of the sensors **14** are discussed.—WT



6,160,757

43.38.Hz ANTENNA FORMED OF A PLURALITY OF ACOUSTIC PICKUPS

Wolfgang Tager and Gregoire Le Tourneur, assignors to France Telecom S.A.
12 December 2000 (Class 367/119); filed in France 10 September 1997

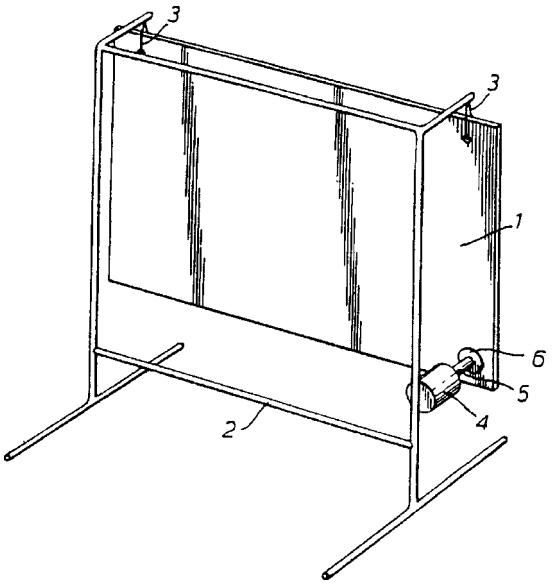
Although the invention is described in terms of microphone arrays it could also be applied to directional loudspeaker arrays. The patent includes considerable background information as well as sophisticated mathematical analysis. Optimized U-shaped and Pi-shaped arrays are developed that might be described as line/endfire hybrids. Anyone interested in broadband array design will find the patent interesting and informative.—GLA

6,058,196

43.38.Ja PANEL-FORM LOUDSPEAKER

Kenneth Harry Heron, assignor to The Secretary of State for Defense in Her Britannic Majesty's Government of the United Kingdom of Great Britain and Northern Ireland
2 May 2000 (Class 381/152); filed in the United Kingdom 4 August 1990

Loudspeakers that behave like sounding boards predate coil-driven cones. Panel form loudspeakers using moving coils to generate bending waves go back at least 30 years. However, it is only in the last ten years or so that distributed mode loudspeakers (DMLs) have been designed using sophisticated scientific analysis. The invention is a skinned, multiresonant



panel with a honeycomb core. It is excited by a single electromechanical driver. Design constraints for a preferred embodiment are given in detail.—GLA

6,144,746

43.38.Ja LOUDSPEAKERS COMPRISING PANEL-FORM ACOUSTIC RADIATING ELEMENTS

Henry Azima *et al.*, assignors to New Transducers Limited
7 November 2000 (Class 381/152); filed in the United Kingdom 9 February 1996

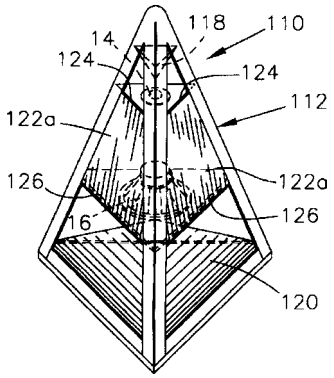
This distributed mode loudspeaker design (DML) makes use of a multiresonant panel resiliently mounted within a peripheral frame, allowing piston operation at low frequencies.—GLA

6,152,257

43.38.Ja AUDIO SPEAKER

Thomas L. Denham, assignor to Thomas L. Denham; Denham Pyramidal Corporation
28 November 2000 (Class 181/155); filed 4 May 1999

Known prior pyramid-shaped speaker systems include woofers and tweeters vertically arranged plus four-sided pyramidal sound reflectors. The



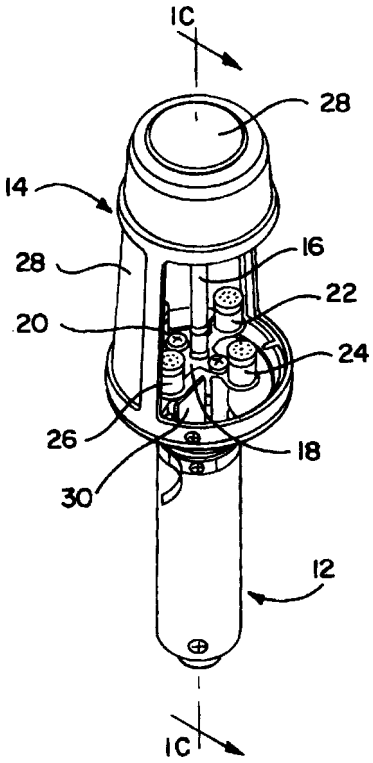
inventor argues that a three-sided pyramid is inherently superior.—GLA

6,084,973

43.38.Kb DIGITAL AND ANALOG DIRECTIONAL MICROPHONE

Jacquelynn Green *et al.*, assignors to Audio Technica U.S., Incorporated
4 July 2000 (Class 381/92); filed 22 December 1997

The patent describes an interesting, sophisticated electroacoustical device. A highly directional ''shotgun'' microphone **16** is combined with additional reference capsules **24** and some fancy digital processing. The goal is



to generate an output signal in which sounds originating from in front of the directional microphone are enhanced and sounds originating from other directions are suppressed.—GLA

6,151,399

43.38.Kb DIRECTIONAL MICROPHONE SYSTEM PROVIDING FOR EASE OF ASSEMBLY AND DISASSEMBLY

Mead C. Killion *et al.*, assignors to Etymotic Research, Incorporated
21 November 2000 (Class 381/313); filed 31 December 1996

The patent includes a well-written overview of hearing aid microphones, more particularly the usage of directional microphones with in-the-ear hearing aids. The invention makes use of several clever tricks to realize the benefits of such a combination while minimizing attendant drawbacks. A nondirectional capsule is combined with a directional capsule to allow user-selected directivity and noise reduction.—GLA

6,157,724

43.38.Lc APPARATUS HAVING LOUDSPEAKERS CONCURRENTLY PRODUCING MUSIC SOUND AND REFLECTION SOUND

Fukushi Kawakami, assignor to Yamaha Corporation
5 December 2000 (Class 381/63); filed in Japan 3 March 1997

Yamaha has previously patented methods of combining electronic musical instruments with room reverberation enhancement. The invention described here is an all-in-one variant. Three loudspeakers and three microphones are housed in the cabinet of an electronic keyboard. Loudspeaker–microphone pairs amplify generally reflected room sound. At the same time, the loudspeakers reproduce audio signals from the keyboard.—GLA

6,141,424

43.38.Si HEADPHONE CAPABLE OF DIRECTLY CONVERTING DIGITAL AUDIO SIGNAL INTO ANALOG AUDIO SIGNAL

Iwao Takiguchi and Motoyuki Yoshizumi, assignors to Sony Corporation
31 October 2000 (Class 381/74); filed in Japan 24 December 1997

This patent has a chutzpah quotient of about 95 percent. A D/A converter and a stereo analog amplifier are both contained in the headphone casing. That is it.—GLA

6,078,669

43.38.Vk AUDIO SPATIAL LOCALIZATION APPARATUS AND METHODS

Robert Crawford Maher, assignor to EuPhonics, Incorporated
20 June 2000 (Class 381/17); filed 14 July 1997

In recent years, techniques for generating virtual sound sources have been stimulated by the unlikely combination of computer games and war room communications. The patent describes a system based on an analysis of moving sound sources in a natural environment rather than head-related transfer functions. ''The system operates successfully for both headphones and speaker playback, and operates properly for all listeners regardless of the physical dimensions of the listener's pinnas, head, and torso.''—GLA

6,154,545

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

Michael Kohut *et al.*, assignors to Sony Corporation; Sony Pictures Entertainment, Incorporated
28 November 2000 (Class 381/23); filed 16 July 1997

The invention is intended to provide more realistic surround sound effects from multichannel cinema sound encoded into two discrete channels. In the recording process, left and right surround signals are shifted 90 degrees. Head related transfer functions are then applied to each of the shifted signals. Additional processing may include phase inversion, level adjustment, and time delay. ''As the encoding process enhances the surround sound signals and further places some of the surround signals onto the front signals, a listener experiences enhanced surround sound effects.''—GLA

6,154,549

43.38.Vk METHOD AND APPARATUS FOR PROVIDING SOUND IN A SPATIAL ENVIRONMENT

Glenn Arnold and Daniel Bates, assignors to Extreme Audio Reality, Incorporated
28 November 2000 (Class 381/104); filed 18 June 1996

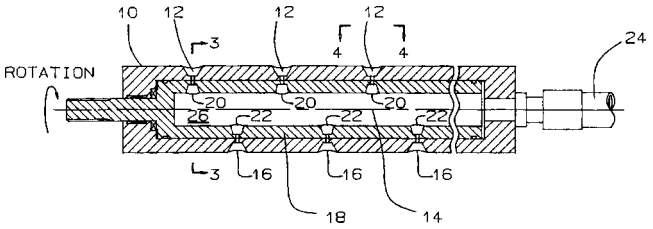
This sound reproduction method is intended to provide controlled sound source localization, including rear images, to one or more listeners in a defined area. Complications are avoided by simply adjusting the relative amplitudes of four loudspeakers located at 90-degree intervals. In other words, quadraphonic panning. Those familiar with the field may wonder if this approach really represents a technical advancement sufficient to justify 44 patent claims.—GLA

6,185,156

43.38.Yn UNDERWATER ACOUSTIC IMPULSE SOURCE

John V. Bouyoucos, assignor to Hydroacoustics-Incorporated
6 February 2001 (Class 367/144); filed 10 June 1999

An acoustic source that can radiate a broad spectrum signal, and whose spectral content can be adjusted to emulate the signal radiated by a surface ship, comprises a multiport air gun that has a stator **10** with a series of ports **12** and **16** in direct communication with the acoustic medium. These ports are alternately valved by a rotor **18** that is coaxial with the stator. The two sets of ports are located diametrically opposite of each other. Pressurized air at about 2000 psi fills the interior bore **26** via inlet **24**. Once during each rotation, as the ports in both the rotor and stator become aligned, pressurized



air escapes, generating acoustic impulses. The speed of the rotor, and hence the spectral content of the radiated signal, is controlled by an electric motor. The amplitude of the radiated signal can be controlled by closing some of the ports by means of a sliding cylindrical piston assembly (not shown) that fits internally within the rotor **18**. The whole assembly is intended to be housed within a hydrodynamically shaped towed vehicle wherein outer face **10** of the source constitutes a portion of the outer face of this vehicle. The electric motor is housed within the vehicle. Electrical power and the pressurized air are conducted to the vehicle via the tow cable.—WT

6,157,196

43.40.Yq METHOD FOR MONITORING OF TAP CHANGERS BY ACOUSTIC ANALYSIS

Tord Bengtsson *et al.*, assignors to ABB Research Limited
5 December 2000 (Class 324/420); filed in Sweden 11 March 1996

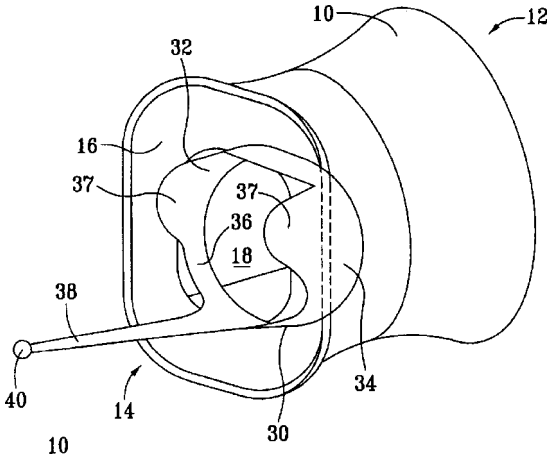
In ac power distribution systems, transformer taps are switched under load to maintain constant voltage. Tap changers suffer from arcing and mechanical wear, yet visual checks are time consuming and necessitate the interruption of service. Sound analysis is a more accurate alternative than temperature or power monitoring. The patent describes a relatively simple method of automatically detecting sonic changes related to the status of a tap changer.—GLA

6,179,085

43.66.Ts RETENTION AND EXTRACTION DEVICE FOR A HEARING AID

Owen D. Brimhall *et al.*, assignors to Sonic Innovations
30 January 2001 (Class 181/135); filed 30 September 1999

A hearing device tip serves as both a retention device and an extraction mechanism. A retention ring is incorporated as part of the tip into which a hearing aid module is inserted. An elongated section on the tip serves as an extraction handle that allows both the hearing device and the tip to be



removed simultaneously from the ear canal. When the hearing aid wearer pulls on the elongated handle, the retention ring cinches around the hearing aid module, preventing it from being separated from the tip.—DAP

6,175,820

43.72.Ar CAPTURE AND APPLICATION OF SENDER VOICE DYNAMICS TO ENHANCE COMMUNICATION IN A TEXT-TO-SPEECH ENVIRONMENT

Timothy Alan Dietz, assignor to International Business Machines Corporation
16 January 2001 (Class 704/235); filed 28 January 1999

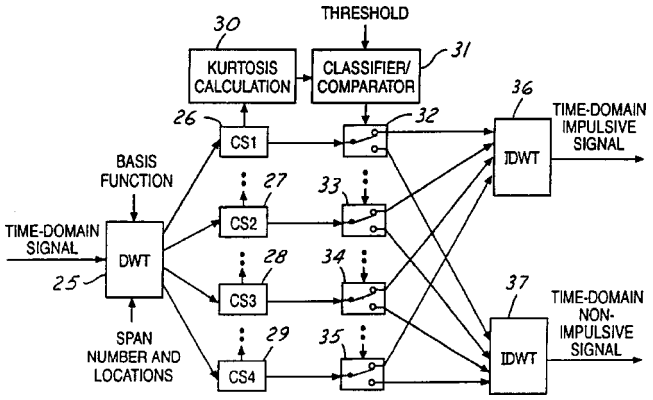
This system first performs recognition of the speech input using a standard software package, such as the IBM ViaVoice system. On a second pass, the speech is further analyzed to extract prosodic information and the text output from the recognition package is automatically annotated to reflect the voice dynamics. The extracted prosodic structure is reduced to a fairly high-level representation, such that the text can be marked up using textual features, such as bolding for loudness, italics for high pitch, and extra spaces inserted for slow speech rates.—DLR

6,182,018

43.72.Ar METHOD AND APPARATUS FOR IDENTIFYING SOUND IN A COMPOSITE SOUND SIGNAL

Vy Tran *et al.*, assignors to Ford Global Technologies, Incorporated
30 January 2001 (Class 702/66); filed 25 August 1998

This ''sound image'' analysis system separates an audio signal into component ''images'' based on the statistics of wavelet coefficient sets. A kind of clustering of the wavelet coefficient sets is done on the basis of kurtosis measurements of waveforms having a strong impulse characteristic.



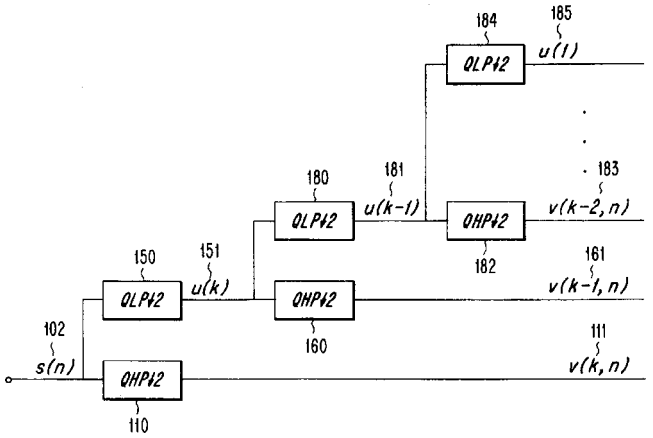
Waveform segments, or components, are assigned to the respective clusters. The separated waveform components are then classified by comparison with a reference database.—DLR

6,182,035

43.72.Ar METHOD AND APPARATUS FOR DETECTING VOICE ACTIVITY

Fisseha Mekuria, assignor to Telefonaktiebolaget LM Ericsson (publ)
30 January 2001 (Class 704/236); filed 26 March 1998

This speech activity detector is based on a fast wavelet transform using quadrature mirror filter pairs. A quadrature mirror filter divides the spectrum equally at one quarter of the sampling rate (half of the Nyquist frequency)



into a low band and a high band. Each half band can be wavelet transformed or further subdivided, producing a tree of frequency band data, with a wavelet transform as the end result of each branch of the tree. The tree structure is defined by an optimization process during the training cycle.—DLR

6,182,036

43.72.Ar METHOD OF EXTRACTING FEATURES IN A VOICE RECOGNITION SYSTEM

Daniel Charles Poppert, assignor to Motorola, Incorporated
30 January 2001 (Class 704/243); filed 23 February 1999

This speech recognition analysis method uses a reduced set of acoustic features in order to reduce memory and computational costs for a small, portable voice processor unit. The inventor has discovered that, because of correlations between the Mel-scale cepstral coefficients (MFCCs), alternate coefficients from the set can be dropped with little effect on the recognition results. Only the odd-numbered MFCCs are retained and even-numbered delta MFCCs are kept.—DLR

6,182,044

43.72.Ar SYSTEM AND METHODS FOR ANALYZING AND CRITIQUING A VOCAL PERFORMANCE

Philip W. Fong and Nelson B. Strother, assignors to International Business Machines Corporation
30 January 2001 (Class 704/270); filed 1 September 1998

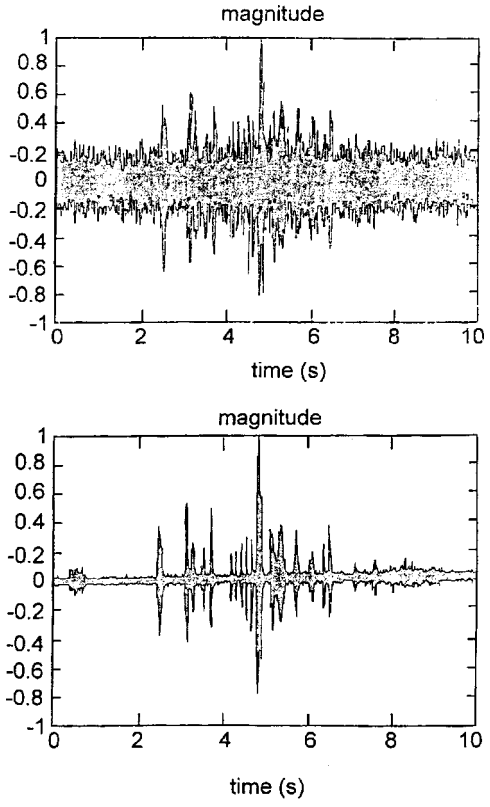
This song analysis system would analyze vocal singing material by extracting pitch, phonetic content, and rhythmic structure from a digitized signal of the singer's voice. The analysis parameters of different songs may be compared, either by the same singer or by different singers, to provide a measure of the difference between the two performances, noting differences in any or all of the extracted qualities.—DLR

6,175,602

43.72.Dv SIGNAL NOISE REDUCTION BY SPECTRAL SUBTRACTION USING LINEAR CONVOLUTION AND CAUSAL FILTERING

Harald Gustafsson *et al.*, assignors to Telefonaktiebolaget LM Ericsson (publ)
16 January 2001 (Class 375/346); filed 27 May 1998

A speech noise reduction system is described that uses the spectral subtraction method of speech enhancement. A novel aspect of the system is the use of low order spectral estimates, which reduces musical tones in the speech output and allows the addition of an interpolation phase, making the



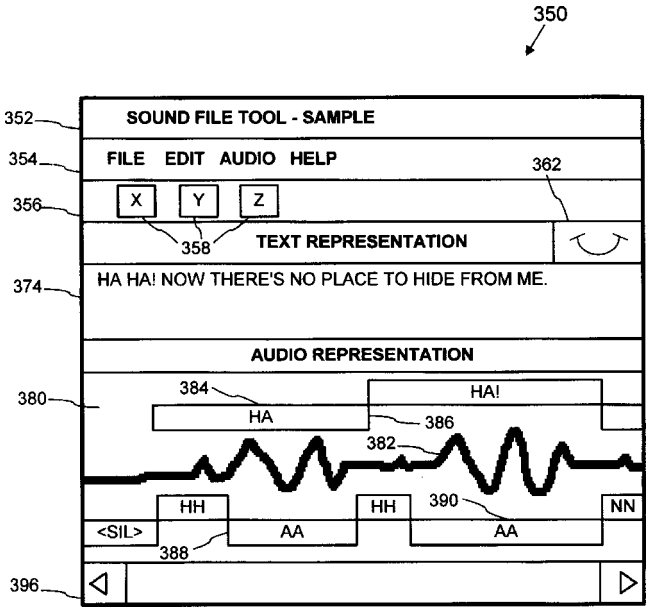
subtraction filter causal and thus preventing waveform discontinuities between frames of the output. The figures show a noisy speech signal as the original and as cleaned up by the method.—DLR

6,181,351

43.72.Ew SYNCHRONIZING THE MOVABLE MOUTHS OF ANIMATED CHARACTERS WITH RECORDED SPEECH

John Wickens Lamb Merrill *et al.*, assignors to Microsoft Corporation
30 January 2001 (Class 345/473); filed 13 April 1998

The system described here allows the synchronization between an animated facial image and spoken dialog. The dialog must be available at least in the form of text and optionally in the form of a recorded speech file. An annotated version of the text is created by adding animation cues, which



generate animated face movements in synch with the speech sounds. The text can also be made to appear in balloons, which change in synch with the sound track. If an animation is prepared with text only, a recorded sound track can be added at any later time.—DLR

6,182,033

43.72.Ew MODULAR APPROACH TO SPEECH ENHANCEMENT WITH AN APPLICATION TO SPEECH CODING

Anthony J. Accardi and Richard Vandervoort Cox, assignors to AT&T Corporation
30 January 2001 (Class 704/223); filed 9 January 1998

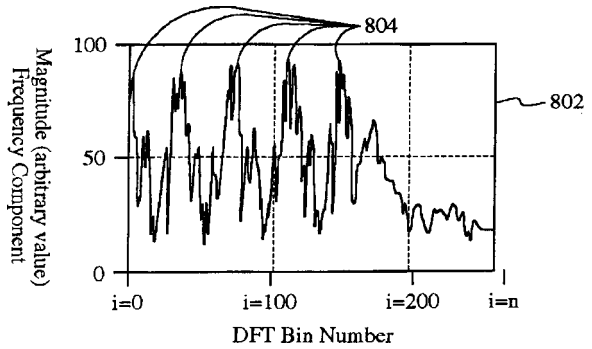
In this modular speech vocoder, the speech signal is first separated into spectral, excitation, and gain components. Each of these components can be enhanced using techniques specific to that signal type. The enhancements described seem to be fairly typical vocoding techniques, such as perceptual filtering of the LPC spectral information.—DLR

6,182,042

43.72.Ew SOUND MODIFICATION EMPLOYING SPECTRAL WARPING TECHNIQUES

Alan Peevers, assignor to Creative Technology Limited
30 January 2001 (Class 704/269); filed 7 July 1998

This speech analysis and resynthesis technique allows the amplitude of individual components of the signal to be selectively modified without changing the magnitude or phase of the remaining components. A FFT spec-



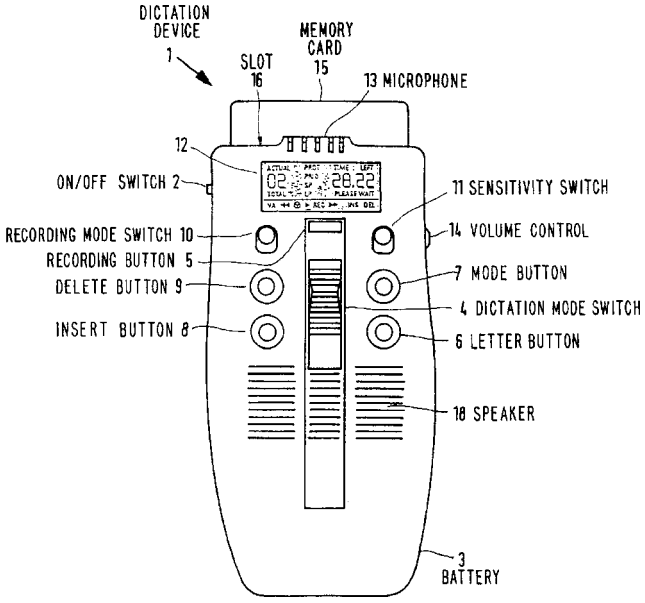
tral envelope of a signal frame is divided into signal components on the basis of spectral peaks. The patent describes a method by which the magnitude and phase of the Fourier coefficients of the selected component are changed, such that the desired signal change is accomplished upon overlap and add recombination of the inverse transforms.—DLR

6,182,043

43.72.Gy DICTATION SYSTEM WHICH COMPRESSES A SPEECH SIGNAL USING A USER-SELECTABLE COMPRESSION RATE

Herbert Böldl, assignor to U.S. Philips Corporation
30 January 2001 (Class 704/270); filed in Austria 12 February 1996

This handheld dictation system would compress the speech signal using various degrees of compression to allow the user to select a trade-off



between signal quality and storage capacity. Intervals of speech would be stored in separate, addressable files on a removable memory card.—DLR

6,173,262

43.72.Ja TEXT-TO-SPEECH SYSTEM WITH AUTOMATICALLY TRAINED PHRASING RULES

Julia Hirschberg, assignor to Lucent Technologies, Incorporated
9 January 2001 (Class 704/260); filed 15 October 1993

This patent describes a method for improving the intonation patterns used when producing phrases or sentences of synthetic speech. A collection of training text is annotated by human linguists to add markers indicating suitable intonation structures. These annotations are then read during text-to-speech synthesis of that text and used to construct statistical representations of the intonation patterns. These statistical representations are then used during the synthesis of new text to generate more natural intonation patterns.—DLR

6,173,263

43.72.Ja METHOD AND SYSTEM FOR PERFORMING CONCATENATIVE SPEECH SYNTHESIS USING HALF-PHONEMES

Alistair Conkie, assignor to AT&T Corporation
9 January 2001 (Class 704/260); filed 31 August 1998

According to this patent, the use of diphone elements provides more intelligible synthetic speech, while the selection of units provides greater naturalness. The described method combines both techniques by collecting large numbers of each possible half-phoneme element from a training speech data corpus. The training speech corpus is then resynthesized and a Viterbi search is used to optimize the selection of half-phoneme units.—DLR

6,175,314

43.72.Ja VOICE ANNUNCIATION OF DATA LINK ATC MESSAGES

George A. Cobley, assignor to Rockwell Collins, Incorporated
16 January 2001 (Class 340/945); filed 25 February 1999

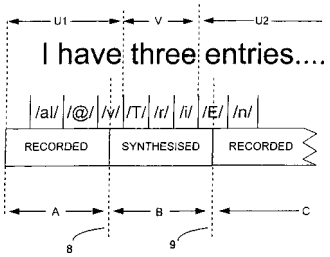
This system for automatically generating ground-to-air messages for use by air traffic controllers is said to meet the provisions of international performance standards RTCA DO-219. Messages are stored in the form of digitized speech waveform samples. Recorded segments allow ground-to-air messages to be produced in various languages, dialects, or speaker-gender, as selected by the cockpit crew.—DLR

6,175,821

43.72.Ja GENERATION OF VOICE MESSAGES

Julian H. Page and Paul Murrin, assignors to British Telecommunications public limited company
16 January 2001 (Class 704/258); filed in European Patent Office 31 July 1997

It is well known that the playback of recorded speech offers the greatest possible naturalness in interactive voice applications, while synthetic speech offers greater flexibility in the choice of words and phrases. This patent describes a system that combines both methods. Recorded speech



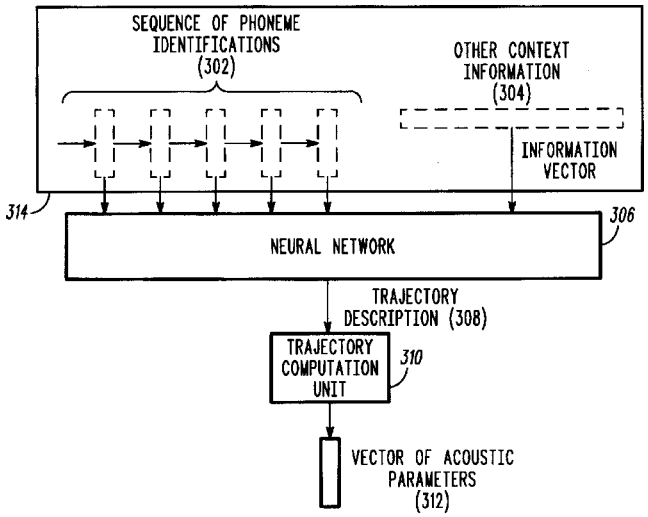
phrases are analyzed and represented in a form compatible with the parameter representation of the synthetic speech. Synthetic words or phrases can then be inserted into a prerecorded phrase, blending the parameters to produce a smooth transition between the speech portions from different sources.—DLR

6,178,402

43.72.Ja METHOD, APPARATUS AND SYSTEM FOR GENERATING ACOUSTIC PARAMETERS IN A TEXT-TO-SPEECH SYSTEM USING A NEURAL NETWORK

Gerald E. Corrigan, assignor to Motorola, Incorporated
23 January 2001 (Class 704/259); filed 29 April 1999

This parametric speech synthesizer uses a neural network to generate acoustic parameter vectors, such as formant frequencies, etc., for a typical formant synthesizer. Prior neural network synthesizers have used the network to generate the entire acoustic parameter trajectories, thus involving a



large amount of network computation. This system simplifies that process by using the network to generate only coefficients for a cubic spline smoothing system. Cubic spline parameters are generated for each half-phoneme, from which the full acoustic parameter trajectories are then directly computed.—DLR

6,182,041

43.72.Ja TEXT-TO-SPEECH-BASED REMINDER SYSTEM

Yizhi Li *et al.*, assignors to Nortel Networks Limited
30 January 2001 (Class 704/260); filed 13 October 1998

This network-based voice reminder system uses a central text-to-speech server to produce reminder messages at specified times for the enrolled clients. Such messages are sent out to the client as a telephone call or an e-mail voice message, according to programmed instructions. The patent addresses server management issues as well as text-to-speech issues.—DLR

6,173,260

43.72.Ne SYSTEM AND METHOD FOR AUTOMATIC CLASSIFICATION OF SPEECH BASED UPON AFFECTIVE CONTENT

Malcolm Slaney, assignor to Interval Research Corporation
9 January 2001 (Class 704/250); filed 29 October 1997

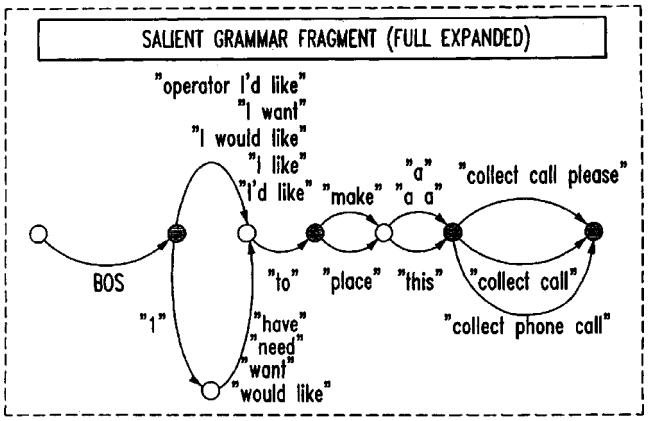
This patent roughly describes a variety of ways to extract emotional content from a speech signal. Feature sets are described only in general acoustical terms, without regard to the classification of specific emotional patterns. Pitch information would be measured at frame rates of 20 to 50 frames per s, while spectral envelope variations are measured at a higher rate. Other possible features are suggested. Utterances are divided into phrases and the features are collected at phrase begin, middle, and end points.—DLR

6,173,261

43.72.Ne GRAMMAR FRAGMENT ACQUISITION USING SYNTACTIC AND SEMANTIC CLUSTERING

Kazuhiro Arai *et al.*, assignors to AT&T Corporation
9 January 2001 (Class 704/257); filed 30 September 1998

The recognition of unconstrained, natural speech presents a difficult problem due to the practically unlimited variety of phrases that can occur. This system extracts grammar fragments from a spoken phrase and provides a method of measuring the distance between phrases using the Kullback–



Leibler measure of distance between probability distributions. In addition to grammar fragments preceding and following a key word, the system also measures semantic distances between various expressions of the key elements of the sentence.—DLR

6,173,266

43.72.Ne SYSTEM AND METHOD FOR DEVELOPING INTERACTIVE SPEECH APPLICATIONS

Matthew T. Marx *et al.*, assignors to SpeechWorks International, Incorporated
9 January 2001 (Class 704/270); filed 6 May 1997

This patent describes a graphical computer interface that allows the rapid construction of speech dialogs for interactive speech applications. Dialogs are constructed from dialog modules, which are represented by icons in the computer model. The icons are manipulated on-screen to specify a call flow graph for the speech interaction. During the graphical process, module parameters may be set to specify attributes, such as prompt time-outs, loop counts, barge-in and beep settings, etc.—DLR.

6,178,401

43.72.Ne METHOD FOR REDUCING SEARCH COMPLEXITY IN A SPEECH RECOGNITION SYSTEM

Martin Franz and Miroslav Novak, assignors to International Business Machines Corporation
23 January 2001 (Class 704/255); filed 28 August 1998

This two-tier speech recognition system performs an acoustic analysis followed by a fast search to generate an initial candidate word list. A grammar analysis then determines a fast match score by comparison with a language model. The acoustic score and the linguistic score are then combined

to determine a final selection of word candidates to be tested with a more thorough, detailed search procedure.—DLR

6,178,403

43.72.Ne DISTRIBUTED VOICE CAPTURE AND RECOGNITION SYSTEM

Michael J. Detlef, assignor to Sharp Laboratories of America, Incorporated
23 January 2001 (Class 704/270); filed 16 December 1998

This hand-held, or ''palm-sized,'' computer unit contains software and hardware capable of recording multiple speech phrases and associating the spoken phrases with any of the several applications that may exist in the device. Phrases may be played back and rerecorded at any time. Later, when the unit is connected to a larger computer system, such as a desktop PC, for the purpose of synchronizing the hand-held unit with the larger system, speech recognition software on the larger system may be used to convert the recorded speech phrases to text.—DLR

6,178,404

43.72.Ne SYSTEM AND METHOD TO FACILITATE SPEECH ENABLED USER INTERFACES BY PROMPTING WITH POSSIBLE TRANSACTION PHRASES

Myra Hambleton and Eric Weeren, assignors to InterVoice Limited Partnership
23 January 2001 (Class 704/275); filed 23 July 1999

Can a telephone interaction menu be improved by adding voice capability? The easy ways to do that, requiring that either exact phrases or menu numbers be spoken, do not help much. This patent provides a better way. Individual spoken words or phrases are used as keys to search the various menu options. The search can be done either at run time or at setup time, generating a keyword grammar. Search parameters are adjustable, allowing a match on anything from a single matching word to requiring an entire correct phrase.—DLR

6,182,038

43.72.Ne CONTEXT-DEPENDENT PHONEME NETWORKS FOR ENCODING SPEECH INFORMATION

Sreeram Balakrishnan and Stephen Austin, assignors to Motorola, Incorporated
30 January 2001 (Class 704/250); filed 1 December 1997

This speech recognizer is arranged for efficient use in a network situation by multiple users, each having unique environmental conditions as well as unique speaker characteristics. The central idea is that trigram phoneme probability tables are separated from the acoustic models used to classify the phonemes and from the language models used to recognize words. This allows a network of trigrams to serve as a single, universal interface between detectors of phoneme sequences as spoken by any speaker and word detectors that may use any language model.—DLR

6,182,039

43.72.Ne METHOD AND APPARATUS USING PROBABILISTIC LANGUAGE MODEL BASED ON CONFUSABLE SETS FOR SPEECH RECOGNITION

Luca Rigazio *et al.*, assignors to Matsushita Electric Industrial Company, Limited
30 January 2001 (Class 704/257); filed 24 March 1998

In this speech recognizer, segment confusability information is combined with syntactic structure constraints to provide better discrimination between likely and unlikely segment sequences. Segments, for this purpose, may be phonemes, words, or in this case, spoken letters as used in spelling names. Valid letter sequences are defined by occurrences in a dictionary of possible names. The spoken letters are divided into 6 classes, such as the ''E'' set, b, c, d, e, g, p, t, v, and z.—DLR

6,182,046

43.72.Ne MANAGING VOICE COMMANDS IN SPEECH APPLICATIONS

Kerry A. Ortega *et al.*, assignors to International Business Machines Corporation
30 January 2001 (Class 704/275); filed 26 March 1998

This patent offers, in some sense, a new way of thinking about the help function of any program which is at least partly controllable by voice commands. The aspect of the help function related to the voice commands is referred to as the ''what can I say?'' (WCIS) function. As an example of some of the thought that has gone into this WCIS system, it is recognized that some voice commands are already cued by visible features of the program. Such commands would not be as prominent in, or omitted altogether from, the WCIS function.—DLR

6,182,037

43.72.Pf SPEAKER RECOGNITION OVER LARGE POPULATION WITH FAST AND DETAILED MATCHES

Stephane Herman Maes, assignor to International Business Machines Corporation
30 January 2001 (Class 704/247); filed 6 May 1997

This speaker recognition system divides the trained speaker population into subgroups. A fast search technique is used to decide which group a candidate speaker best fits into. If ambiguity remains of the candidate's identity, a detailed search then makes the final decision, limited to speakers within the selected group. The detailed search also makes use of speaker-dependent information.—DLR

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.

On the effect of boundaries on radiation resistance of plates

K. Renji
Structures Group, ISRO Satellite Centre, Bangalore 560017, India

(Received 5 April 2000; accepted for publication 30 May 2001)

At high frequencies, radiation resistance of a plate converges to the radiation resistance of the simply supported plate and is not influenced by its boundary conditions. But, at low frequencies the radiation resistance is significantly influenced by the boundary conditions. Up to what frequency the boundaries affect the radiation resistance is experimentally investigated in this study. It is found that this depends on the number of bending waves present in the structure and not merely on the critical frequency. If the plate contains more than two bending waves, the radiation resistance is approximately equal to the radiation resistance of the plate with simply supported boundaries. Also, a plate behaves like an infinite plate at frequencies above the critical frequency only if it contains a few bending waves. [DOI: 10.1121/1.1387997]

PACS numbers: 43.40.At [PJR]

I. INTRODUCTION

Expressions for the radiation resistance of a plate were originally derived by Maidanik.[1] These expressions, with some corrections,[2–4] are used in determining the radiation resistance of a plate with simply supported boundaries. At high frequencies the boundary conditions do not affect the radiation resistance, hence, the radiation resistance of a plate with any type of boundaries is given by that of the plate with simply supported boundaries. At low frequencies the boundary conditions have significant effects on the radiation resistance. Up to what frequency the boundaries affect the radiation resistance is not very clear in the literature. The present study is an experimental investigation of this aspect.

II. RADIATION RESISTANCE

Radiation resistance, denoted by R_{rad}, is a measure of the sound power radiated by the structure, which is related by

$$W = R_{\mathrm{rad}} \langle v^2 \rangle_x , \tag{1}$$

where W is the sound power radiated by the structure when the spatial average of the mean-square value of the velocity of the structure is $\langle v^2 \rangle_x$. The frequency averaged radiation resistance of a plate with simply supported edges kept in a rigid, infinite, and planar baffle and radiating into a reverberant room by one side of the plate can be estimated using the following equations:[1–4] For $f < f_c$ and $ka, kb > 2\pi$

$$R_{\mathrm{rad}} = A\rho_a c\{(\lambda_c \lambda_a / A) 2 (f/f_c) g_1 + (p\lambda_c / A) g_2\}/2,$$

where

$$g_1 = (4/\pi^4)\{(1-2\psi^2)/[\psi(1-\psi^2)^{1/2}]\} \quad \text{for } f/f_c < 0.5,$$

$$g_1 = 0 \quad \text{for } f/f_c \geqslant 0.5,$$

$$g_2 = (1/4\pi^2)\{(1-\psi^2)\ln[(1+\psi)/(1-\psi)] + 2\psi\} \times \{1/(1-\psi^2)^{3/2}\},$$

$$\psi = (f/f_c)^{1/2}.$$

For $f < f_c$ and $ka, kb < 2\pi$

$$R_{\mathrm{rad}} = A\rho_a c (4/\pi^4)(p\lambda_c / A)(f/f_c)^{1/2}/2.$$

For $f = f_c$

$$R_{\mathrm{rad}} = A\rho_a c\{(a/\lambda_c)^{1/2} + (b/\lambda_c)^{1/2}\}/2.$$

For $f > f_c$

$$R_{\mathrm{rad}} = A\rho_a c\{1-(f_c/f)\}^{-1/2}. \tag{2}$$

The plate has dimensions a, b. In Eq. (2) A is the area, p is the perimeter, and f_c is the critical frequency of the panel. The wavelength of sound in air is denoted by λ_a and the wavelength at the critical frequency is denoted by λ_c. The air has a density ρ_a and the speed of sound in air is c. The parameter k is the wave number in air at frequency f.

Maidanik[1] observed that the radiation resistance of a panel with clamped edges is twice that of the panel with simply supported edges. Nikiforov[5] proved that the radiation resistance becomes zero when the edges are free. Equation (2) gives the radiation resistance of plates radiating into a reverberant room by one side of the panel. When both sides of the panel radiate sound, the radiation resistance is double the value obtained using Eq. (2).[4,6] In many situations one has to find the radiation resistance of a plate which is not kept in a planar baffle, as in the case of a wall of a rectan-

gular box. In such cases Lyon[7] obtained the radiation resistance using Eq. (2) multiplied by a factor of 0.5 for an unbaffled plate.[4,7] For unbaffled plates, apart from the above discussed factor of 0.5, yet another phenomenon called ''short circuiting'' affects the radiation resistance. This is due to the flow of air from one side of the plate to the other without the air getting compressed, consequently, it reduces the radiation resistance and the effect is more obvious at low frequencies. Oppenheimer and Dubowski[8] derived expressions for this effect. Radiation resistance of any type of plate construction can be estimated using Eq. (2) with factors for boundary conditions, effect of neighboring structure, and effect of inertial flows.

One important question remains, which is up to what frequency the boundary conditions affect the radiation resistance. In other words, up to what frequency the factor for boundary conditions need be applied. Maidanik[9] and Fahy[10] applied the factor 2 for the radiation resistance of clamped plates for frequencies up to $0.5f_c$, which was progressively reduced to unity at critical frequency. The boundaries do not affect the radiation resistance beyond critical frequency.[9,10] The critical frequency of a panel does not depend on its dimensions. Hence, it is possible to have panels having frequencies of the first few bending modes larger than their critical frequencies. If the effects of boundaries on the radiation resistance of plates are applied only up to their critical frequencies, it implies that it is possible that the boundaries do not affect the radiation resistance of even the first few modes. Hence, the frequency up to which the boundaries affect the radiation resistance should be dependent on the number of bending waves present in the structure rather than the critical frequency. Therefore, it is logical to assume that if the plate contains a few bending waves, the effect of boundary conditions may not be significant. In other words, the radiation resistance of a plate is very close to that of the simply supported plate at frequencies where $\lambda_b \ll a,b$, where λ_b is the wavelength of the bending waves in the plate. This behavior is experimentally verified in this study.

III. EXPERIMENTAL RESULTS

Radiation resistance of a typical plate is determined experimentally. The structure considered is a honeycomb sandwich panel with face sheets made of aluminum alloy. The panel has dimensions 1.3×1.1 m. The face sheet has a thickness of 0.19 mm and the thickness of the core is 25.4 mm. The measured mass of the panel is 4.3 kg. The mass of the panel without the concentrated masses is 2.75 kg. The critical frequency of the panel is estimated to be 382 Hz when the speed of sound in air is 346 m/s.

Radiation resistance is determined experimentally by exciting the panel using shaker systems and measuring the vibration responses of the panel, as well as the sound-pressure level in the room. From the above-measured values, the radiation resistance is given by[4]

$$R_{\mathrm{rad}}=\langle p^2\rangle_x\{s\alpha/(1-\alpha)\}/\{4\rho_a c\langle v^2\rangle_x-(nc^2/\pi f^2\rho A)\langle p^2\rangle_x\}, \tag{3}$$

where $\langle p^2\rangle_x$ is the spatial average of the mean-square value of the sound in the room and n is the modal density of the plate. The room has a surface area of s and its average absorption coefficient is α. The parameter α includes the absorption due to walls of the room as well as that due to air. The panel is hung in the reverberation chamber. The four edges of the panel have free boundary conditions and no baffle is used. Three driving-point positions are used and the acceleration levels are measured at 11 randomly selected locations. The acoustic field is measured at three locations. In the present experiment, the radiation resistance values are obtained up to 2500 Hz in 1/3-octave bands. Since the number of modes in 1/3-octave bands below 315 Hz is less than one, the radiation resistance values below 315 Hz, 1/3-octave bands are not given here. The measured radiation resistance values are given in Fig. 1.

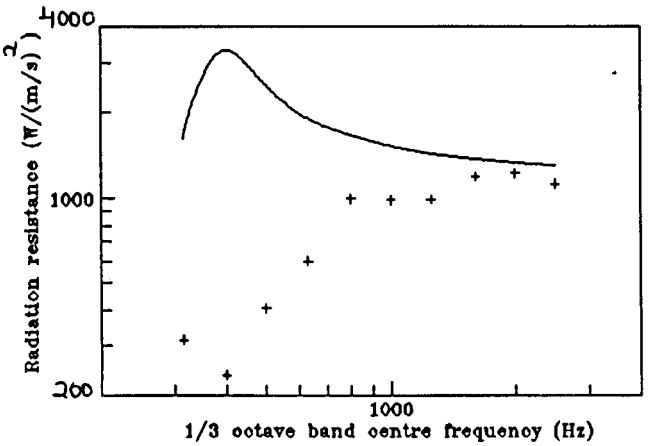


FIG. 1. Radiation resistance of a panel: ———, theory: +, experiment.

The radiation resistance of the panel is theoretically estimated using Eq. (2) and shown in Fig. 1. The boundaries are assumed to be simply supported. Since the plate is unbaffled, a factor of 0.5 is used as discussed earlier. For the discussion of the results, it is important to obtain the wavelength of the bending waves in the panel. The wavelength of the bending wave is estimated to be 0.92 m at the critical frequency, hence, there is approximately only one bending wave present in the structure.

Let us now compare the experimentally determined radiation resistance with the estimated values. It is to be noted that for estimating the radiation resistance, the boundaries are assumed to be simply supported. The results show that the experimentally obtained radiation resistance values are very low compared to the theoretical estimates. The experimental results match with the theoretical results only at frequencies above 1250 Hz, 1/3-octave band. One can observe that the radiation resistance is influenced by the boundary conditions up to 1250 Hz, 1/3-octave band, even beyond the critical frequency. Many researchers, like Maidanik,[9] Fahy,[10] Lyon,[11] and Clarkson and Brown[12] followed the assumption that the radiation resistance is independent of the boundary conditions for frequencies above the critical frequency. It is now clear that this is not true. It is important to note that at the critical frequency there is only one bending wave present in the panel, whereas beyond the 1250-Hz octave band, where the effects of boundary conditions are not seen, there are more than two bending waves present. The frequency up

to which the boundary conditions can be considered to be affecting the radiation resistance depends mainly on the number of bending waves present in the panel and not merely on the critical frequency.

It can be argued that the expressions used for the radiation resistance are valid only when the modal densities of the structure are high. But, this is true for frequencies up to the critical frequency. For frequencies above the critical frequency, the expression for the radiation resistance is not derived based on this assumption. It is based on the fact that a plate behaves like an infinite plate for frequencies above the critical frequency. In the present study the results are discussed for frequencies above the critical frequency. Hence, this limitation does not arise in the present case. But, it is important to note from the experimental results that it is not correct that a plate always behaves like an infinite plate for frequencies above the critical frequency. At frequencies above the critical frequency a plate radiates like an infinite plate only if it contains a few bending waves.

IV. CONCLUSIONS

The frequency up to which the boundaries affect the radiation resistance depends on the number of bending waves present in the structure and not its critical frequency alone. The experimental results show that if there are more than two bending waves present in the plate, radiation resistance is independent of its boundary condition. A plate radiates like an infinite plate at frequencies above the critical frequency only if a few bending waves are present in the structure.

ACKNOWLEDGMENT

The author wishes to thank Dr. P. S. Nair, Group Director, Structures, ISRO Satellite Centre for his valuable suggestions.

[1] G. Maidanik, ''Response of Ribbed Panels to Reverberant Acoustic Fields,'' J. Acoust. Soc. Am. **34**, 809–826 (1962).

[2] M. J. Crocker and A. J. Price, ''Sound Transmission using Statistical Energy Analysis,'' J. Sound Vib. **9**, 469–486 (1969).

[3] G. Maidanik, Erratum: ''Response of Ribbed Panels to Reverberant Acoustic Fields,'' J. Acoust. Soc. Am. **57**, 1552 (1975).

[4] K. Renji, P. S. Nair, and S. Narayanan, ''On Acoustic Radiation Resistance of Plates,'' J. Sound Vib. **212**, 583–598 (1998).

[5] A. S. Nikiforov, ''Radiation from a Plate of Infinite Dimensions with Arbitrary Boundary Conditions,'' Sov. Phys. Acoust. **10**, 178–182 (1964).

[6] M. P. Norton, *Fundamentals of Noise and Vibration Analysis for Engineers* (Cambridge University Press, Cambridge, England, 1989), p. 415.

[7] R. H. Lyon, ''Noise Reduction of Rectangular Enclosures with one Flexible Wall,'' J. Acoust. Soc. Am. **35**, 1791–1797 (1963).

[8] C. H. Oppenheimer and S. Dubowsky, ''A Radiation Efficiency for Unbaffled Plates with Experimental Validation,'' J. Sound Vib. **199**, 473–489 (1997).

[9] G. Maidanik, ''Radiation Efficiency of Panels,'' J. Acoust. Soc. Am. **35**, 115 (1963).

[10] F. J. Fahy, ''Vibration of Containing Structures by Sound in the Contained Fluid,'' J. Sound Vib. **10**, 490–512 (1969).

[11] R. H. Lyon, *Statistical Energy Analysis of Dynamical Systems: Theory and Applications* (MIT Press, Cambridge, MA, 1975), p. 300.

[12] B. L. Clarkson and K. T. Brown, ''Acoustic Radiation Damping,'' ASME J. Vib., Acoust., Stress, Reliab. Des. **107**, 357–360 (1985).

The effects of pressure loading on the dynamics of a curved piezostructure

James K. Henry[a] and Robert L. Clark[b]
Department of Mechanical Engineering and Materials Science, Duke University, Box 90302, Durham, North Carolina 27708-0302

(Received 8 May 2000; accepted for publication 22 May 2001)

In an article previously published in this journal by the authors, a state-space model of a curved piezostructure was developed and used to investigate the effects of curvature on panel dynamics. Panel curvature was shown to increase control system bandwidth and to affect transducer coupling. In the present work, the model is extended to include the effects of internal pressure loading on the dynamics of a curved piezostructure. The impact of internal pressure loading on the dynamics of a curved piezostructure, and subsequently on active structural acoustic control (ASAC) of sound transmission through it, are reviewed. © *2001 Acoustical Society of America.*
[DOI: 10.1121/1.1387998]

PACS numbers: 43.40.Vn [PJR]

I. INTRODUCTION

Model-based control system design depends upon model accuracy to ensure realistic prediction of closed-loop performance. Recent research on active structural acoustic control (ASAC) of curved panels for reduction of interior cabin noise in aircraft has sought to include complicating effects in panel dynamics in an effort to more closely simulate the service conditions of actual aircraft panels. As part of this research, curvature was shown to significantly affect control system bandwidth and transducer coupling.[1] The present work will focus upon the effects of pressure loading on the dynamics of a curved panel and the implications of pressure loading on ASAC design strategies.

A Hamiltonian energy method will be used to develop a state-space model of a tensioned curved panel with attached piezoelectric actuators. The effects of the internal pressure loading on the dynamics of the curved piezostructure will be reviewed briefly. Finally, the impact of the internal pressure on ASAC design strategies will be discussed.

II. MODEL DEVELOPMENT

The internal pressure model is developed through the extension of an energy method used to model structures with attached piezoelectric actuators first described by Hagood *et al.*,[2] and later used by the authors to model a curved piezostructure.[1] In order to model the effects of internal pressure, a curved panel is assumed to be a portion of a complete cylinder with hemispherical endcaps. The cylinder is subjected to internal pressure loading. The panel is assumed to be far from the endcaps of the cylinder, allowing the internal pressure loads to be modeled as in-plane tensile forces. The effects of pressure loading on the panel dynamics are investigated by subjecting a simply supported curved panel with attached piezoelectric actuators to in-plane tensile forces corresponding to a given internal pressure load.

A. Internal pressure

A cylinder with endcaps is subjected to a static pressure load, p_s. At a sufficient distance from the endcaps, membrane theory defines the in-plane tensile forces that a cylindrical portion of the cylinder with radius R encounters

$$N_{xx}=p_sR/2, \quad N_{\phi\phi}=p_sR, \quad N_{x\phi}=0, \tag{1}$$

where p_s is determined from the pressure differential between the interior cabin pressure and the exterior atmospheric pressure

$$p=P_{\text{int}}-P_{\text{ext}}. \tag{2}$$

The in-plane tensile forces are applied to the curved piezostructure to simulate static pressure loading.

B. Energy method

A curved panel with shear-diaphragm boundary conditions is subjected to the combined action of a transverse load from a piezoelectric actuator and in-plane tensile forces corresponding to a given static pressure load as shown in Fig. 1. The contribution of the tensile load and the transverse load, as well as the interaction between the two, is modeled.

The generalized energy balance from Hamilton's principle for electromechanical systems[2–4] is

$$\int_{t_1}^{t_2}[\delta(T-U+W_e-W_m)+\delta W]dt=0. \tag{3}$$

The additional energy contribution of the tensile forces and the interaction between the tensile loads and the transverse load only affect the potential energy of the system, U. The new expression for the potential energy will be built term by term.

Initially, the curved panel is subjected to the tensile load. The first term of the potential energy expression, U_{pr}, results from the stretching of the midplane of the curved panel

$$U_{pr}=\frac{1}{2}\int_{S_s}\{\boldsymbol{\epsilon}\}^T\{\mathbf{n_{pr}}\}dS_s, \tag{4}$$

[a]Graduate Research Assistant, Student Member ASA.
[b]Professor, Member ASA.

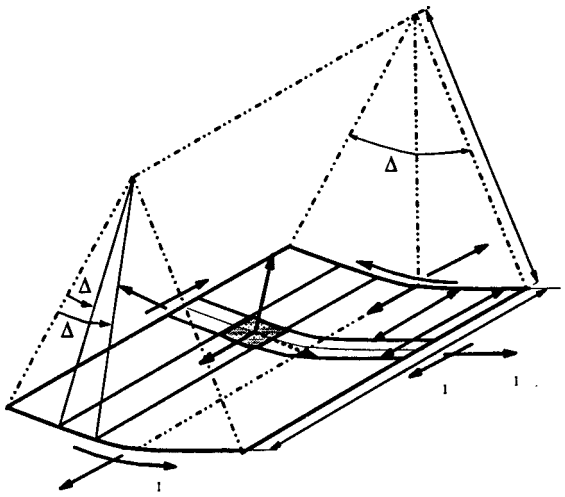


FIG. 1. Coordinate system of curved panel with attached piezoelectric transducer (shaded) subjected to tensile loads.

where S_s is the surface of the structure, $\mathbf{n_{pr}}$ is the vector of tensile stress resultants fron Eq. (1)

$$\mathbf{n_{pr}}=\{N_{xx} \quad N_{\phi\phi} \quad N_{x\phi}\}^T, \tag{5}$$

and ϵ is the vector of material strains

$$\{\epsilon\}=\{\epsilon_{xx} \quad \epsilon_{\phi\phi} \quad \epsilon_{x\phi}\}^T. \tag{6}$$

Since the values of the tensile stress resultants are known, Hooke's law can be used to obtain an expression for the strain vector in terms of the tensile stress resultants

$$\epsilon=\frac{1}{h_s}\mathbf{C_s}^{-1}\mathbf{n_{pr}}, \tag{7}$$

where $\mathbf{C_s}$ is the structural material stiffness matrix

$$\mathbf{C_s}=\begin{bmatrix} \frac{E_s}{1-\nu_s^2} & \frac{E_s\nu_s}{1-\nu_s^2} & 0 \\ \frac{E_s\nu_s}{1-\nu_s^2} & \frac{E_s}{1-\nu_s^2} & 0 \\ 0 & 0 & \frac{E_s}{2(1+\nu_s)} \end{bmatrix}. \tag{8}$$

Expanding, the first term of the potential energy of the system becomes

$$U_{pr}=\frac{1}{2h_sE_s}\int_0^{\alpha}\int_0^{L_x}[N_{xx}^2+N_{\phi\phi}^2-2\nu N_{xx}N_{\phi\phi} +2(1+\nu)N_{x\phi}^2]dx\,R\,d\phi. \tag{9}$$

Note that the midplane stretching potential energy contribution, U_{pr}, is not a function of the displacements of the curved panel.

It is then assumed that while the panel is tensioned, the piezoelectric transducers are attached. A transverse load is applied to the structure by the attached transducers. The potential energy added by the transverse load increases the potential energy of the system by U_t, where

$$U_t=\frac{1}{2}\int_{V_s}\epsilon^T(\mathbf{x})\sigma(\mathbf{x})dV_s + \sum_{n_p=1}^{N_p}\frac{1}{2}\int_{V_{pn_p}}\epsilon^T(\mathbf{x})\sigma(\mathbf{x})dV_{pn_p}, \tag{10}$$

and the vector of material stresses, σ, is

$$\{\sigma\}=\{\sigma_{xx} \quad \sigma_{\phi\phi} \quad \sigma_{x\phi}\}^T. \tag{11}$$

As detailed in Ref. 1, Hooke's law can be used to express the stress vector in terms of the strain vector and material properties. Further, the strain vector can be described in terms of a linear differential operator and the midplane displacements of the curved piezostructure. Thus, the entire potential energy of bending can be written as a function of the midplane displacements

$$U_t=\frac{1}{2}\int_{V_s}\{\mathbf{L_{u_m}u_m}\}^T\mathbf{C_s}\{\mathbf{L_{u_m}u_m}\}dV_s + \sum_{n_p=1}^{N_p}\frac{1}{2}\int_{V_{pn_p}}\{\mathbf{R_sL_{u_m}u_m}\}^T\mathbf{C}^E\{\mathbf{R_sL_{u_m}u_m}\}dV_{pn_p}. \tag{12}$$

The transverse load also produces additional strain in the midplane of the panel: strain which is negligible in the absence of the tensile load. However, the small midplane strains in combination with the tensile loads contribute terms to the potential energy that are of the same order as the potential energy added by the transverse load, U_t. The additional strains in the midplane of the panel induced by bending are found using the strain–displacement relations for large displacements in a curved panel. The large displacement strain–displacement relations for a curved panel with the coordinate system shown in Fig. 1 according to Leissa[5] are

$$\{\epsilon'\}=\begin{Bmatrix}\epsilon'_{xx}\\ \epsilon'_{\phi\phi}\\ \epsilon'_{x\phi}\end{Bmatrix}=\begin{Bmatrix} \frac{\partial u}{\partial x}+\frac{1}{2}\left(\frac{\partial w}{\partial x}\right)^2 \\ \frac{1}{R}\frac{\partial v}{\partial\phi}+\frac{w}{R}+\frac{1}{2R^2}\left(\frac{\partial w}{\partial\phi}\right)^2 \\ \frac{1}{R}\frac{\partial u}{\partial\phi}+\frac{\partial v}{\partial x}+\frac{1}{R}\frac{\partial w}{\partial x}\frac{\partial w}{\partial\phi} \end{Bmatrix}. \tag{13}$$

The potential energy contributed by the combination of the small midplane strains and the tensile loads is

$$U_c=\int_{S_s}\{\epsilon'\}^T\{\mathbf{n_{pr}}\}dS_s, \tag{14}$$

where S_s is the surface of the curved panel structure. Substituting Eqs. (5) and (13) into Eq. (14) and expanding,

$$U_c=U_{c_1}+U_{c_2}, \tag{15}$$

where

$$U_{c_1}=\int_0^{\alpha}\int_0^{L_x}\left[N_{xx}\left(\frac{\partial u}{\partial x}\right)+N_{\phi\phi}\left(\frac{1}{R}\frac{\partial v}{\partial \phi}+\frac{w}{R}\right)+N_{x\phi}\left(\frac{1}{R}\frac{\partial u}{\partial \phi}+\frac{\partial v}{\partial x}\right)\right]dx\,R\,d\phi, \tag{16}$$

and

$$U_{c_2}=\frac{1}{2}\int_0^{\alpha}\int_0^{L_x}\left[N_{xx}\left(\frac{\partial w}{\partial x}\right)^2+N_{\phi\phi}\left(\frac{1}{R}\frac{\partial w}{\partial \phi}\right)^2+2N_{x\phi}\left(\frac{1}{R}\frac{\partial w}{\partial x}\frac{\partial w}{\partial \phi}\right)\right]dx\,R\,d\phi. \tag{17}$$

Equation (16) can be rewritten in matrix form

$$U_{c_1}=\int_0^{\alpha}\int_0^{L_x}[\mathbf{L_u}\mathbf{u_m}]^T\mathbf{n_{pr}}\,dx\,R\,d\phi, \tag{18}$$

where the linear differential operator, $[\mathbf{L_u}]$, in the midplane of the curved panel is

$$\{\mathbf{L_u}\}=\begin{bmatrix}\frac{\partial}{\partial x} & 0 & 0\\ 0 & \frac{1}{R}\frac{\partial}{\partial \phi} & \frac{1}{R}\\ \frac{1}{R}\frac{\partial}{\partial \phi} & \frac{\partial}{\partial x} & 0\end{bmatrix}. \tag{19}$$

Equation (17) can also be expressed in matrix form

$$U_{c_2}=\frac{1}{2}\int_0^{\alpha}\int_0^{L_x}[\mathbf{L_{p_1}}\mathbf{u_m}]^T\mathbf{N}[\mathbf{L_{p_2}}\mathbf{u_m}]dx\,R\,d\phi, \tag{20}$$

where the two linear differential operators, $\mathbf{L_{p_1}}$ and $\mathbf{L_{p_2}}$, are

$$\mathbf{L_{p_1}}=\begin{bmatrix}0 & 0 & \frac{\partial}{\partial x}\\ 0 & 0 & \frac{1}{R}\frac{\partial}{\partial \phi}\\ 0 & 0 & \frac{1}{R}\frac{\partial}{\partial x}\end{bmatrix}, \tag{21}$$

and

$$\mathbf{L_{p_2}}=\begin{bmatrix}0 & 0 & \frac{\partial}{\partial x}\\ 0 & 0 & \frac{1}{R}\frac{\partial}{\partial \phi}\\ 0 & 0 & \frac{1}{R}\frac{\partial}{\partial x}\end{bmatrix}, \tag{22}$$

The tensile stress matrix, $\mathbf{N}$, is the diagonalization of the tensile stress vector, $\mathbf{n_{pr}}$, such that

$$\mathbf{N}=\begin{bmatrix}N_{xx} & 0 & 0\\ 0 & N_{\phi\phi} & 0\\ 0 & 0 & N_{x\phi}\end{bmatrix}. \tag{23}$$

Expressing the potential energy terms in matrix form simplifies the generation of the equations of motion of the system. The expression for total potential energy is

$$U=U_t+U_{pr}+U_{c_1}+U_{c_2}. \tag{24}$$

Following the Rayleigh–Ritz procedure, the displacements of the curved panel are expressed in terms of generalized coordinates and mode shape functions

$$\mathbf{u_m}(x,\phi,t)=\sum_{k=1}^{K}\boldsymbol{\Psi}^{\mathbf{d}}k(x,\phi)\mathbf{q}_k(t). \tag{25}$$

The equations of motion of the system are found by substituting for each energy term in the generalized Hamilton's principle defined in Eq. (3). Since only the potential energy term has been affected by this development, it is not necessary to discuss the variations of the other energy terms. Substituting Eq. (25) for the displacement, the variations of each term of the total potential energy are taken starting with the potential energy contributed by the transverse load

$$\delta U_t=[\mathbf{K_s}+\mathbf{K_p}]\{\mathbf{q}\}. \tag{26}$$

The structural stiffness matrix, $\mathbf{K_s}$, and piezoelectric actuator stiffness matrix, $\mathbf{K_p}$, were derived in Ref. 1. Continuing, the variation of the midplane stretching term is zero

$$\delta U_{pr}=[\mathbf{0}], \tag{27}$$

since it is not a function of the generalized coordinates. The variation of U_{c_1} yields a generalized force term

$$\delta U_{c_1}=[\mathbf{P_f}]\{\mathbf{f}\}, \tag{28}$$

where $\mathbf{P_f}$ is given by

$$\mathbf{P_f}=\int_0^{\alpha}\int_0^{L_x}[\mathbf{L_u}\boldsymbol{\Psi}^{\mathbf{d}}]^T\mathbf{n_{pr}}\,dx\,R\,d\phi. \tag{29}$$

When calculated for each mode in the model, the resulting generalized force expression, $\mathbf{P_f}$, is a $3K\times 1$ vector, where K is the number of assumed modes. Finally, the variation of U_{c_2} results in a stiffness contribution

$$\delta U_{c_2}=[\mathbf{K_{tf}}]\{\mathbf{q}\}, \tag{30}$$

where the tensile force stiffness is

$$\mathbf{K_{tf}}=\frac{1}{2}\int_0^{\alpha}\int_0^{L_x}[\mathbf{L_{p_1}}\boldsymbol{\Psi}^{\mathbf{d}}]^T\mathbf{N}[\mathbf{L_{p_2}}\boldsymbol{\Psi}^{\mathbf{d}}]dx\,R\,d\phi. \tag{31}$$

After evaluation, the stiffness contribution of the static pressure load is a $3K\times 3K$ matrix. Note from the above expressions that the added stiffness only affects the out-of-plane modes.

Compiling the terms, the equations of motion of the tensioned curved piezostructure can be written

$$[\mathbf{M_s}+\mathbf{M_p}]\{\ddot{\mathbf{q}}\}+[\mathbf{K_s}+\mathbf{K_p}+\mathbf{K_{tf}}]\{\mathbf{q}\}=[\Theta]\{\mathbf{v}\}+[\mathbf{B_f}+\mathbf{P_f}]\{\mathbf{f}\}. \tag{32}$$

The equations of motion can be cast in state-variable form in the same manner as that shown in Ref. 1.

III. SUMMARY OF RESULTS AND IMPLICATION FOR CONTROL

The effects of pressure loading on the dynamics of shells are well known. As in the case of curvature, pressure loading increases the effective stiffness of the curved panel, resulting in higher natural frequencies, thus affecting the bandwidth of control systems used to reduce sound radiation. In the case of a curved panel as a portion of an internally pressurized cylindrical shell, the pressure load in the circumferential direction is $2\times$ the load in the axial direction. Therefore, pressure loading affects circumferentially oriented modes more significantly than axially oriented modes, causing modal reordering similar to that discussed in Ref. 1.

However, whereas curvature has a constant effect, pressure loading dynamically changes during flight conditions. As the pressure load varies, the modal order and bandwidth change. The changes in the system dynamics make model-based control difficult, since the control system must adapt to changes in the model on which it is based. As a result, techniques such as gain scheduling or adaptive control must be considered.

In active structural acoustic control, the physical installation of piezoelectric actuators is also impacted by the pressure loading. Since piezoelectric actuators have low tensile strength, care must be taken to ensure that the actuators do not encounter the tensile loads caused by pressurization. This can be accomplished by either subjecting the panel to a tensile force or compressing the actuator prior to actuator installation. As a result, the piezoelectric actuator will be under compression when the structure is relaxed and catastrophic failure of the actuator can be avoided.

ACKNOWLEDGMENT

This work was supported by NASA Langley Research Center under Grant No. NCC 1 250.

[1] J. K. Henry and R. L. Clark, "A curved piezostructure model: Implications on active structural acoustic control," J. Acoust. Soc. Am. **106**, 1400–1407 (1999).

[2] N. W. Hagood, W. H. Chung, and A. von Flotow, "Modeling of piezoelectric actuator dynamics for active structural control," J. Intell. Mater. Syst. Struct. **1**(3), 327–354 (1990).

[3] S. A. Crandall, D. C. Karnopp, E. F. Kurtz, and D. C. Pridmore-Brown, *Mechanical and Electromechanical Systems* (Krieger, Malabar, FL, 1968).

[4] R. L. Clark, G. P. Gibbs, and W. R. Saunders, *Adaptive Structures, Dynamics and Control* (Wiley, New York, 1998).

[5] Arthur Leissa, *Vibration of Shells* (Acoustical Society of America, Woodbury, NY, 1993).

Comment on "Reducing seat dip attenuation" [J. Acoust. Soc. Am. 108, 2211–2218 (2000)]

David Lloyd Klepper
P.O. Box 24379, Mt. Scopus, Jerusalem 91240, Israel

(Received 2 February 2001; revised 1 May 2001; accepted 12 June 2001)

A recent article has proposed a novel approach to reducing the ''seat dip attenuation'' commonly observed in major auditoria. The proposed solution might have significant impact on other attributes of auditorium acoustics. An approach to evaluate such effects is proposed. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1392381]

PACS numbers: 43.55.Br, 43.55.Ka [JDQ]

My congratulations to authors W. J. Davies and T. J. Cox for their excellent research and their well-written paper. One serious music acoustics problem now appears on its way to solution due to their fine work.

Some questions remain unanswered:

(1) How are seat and audience absorption coefficients affected by the well under the audience floor?
(2) In reverberation time calculations, should one include the well volume in the hall volume? Does the answer to this question depend on the frequency band of interest?
(3) Are there frequency bands where the seats and audience should be treated as objects in the midst of a volume rather than on a boundary surface?

An experiment is suggested to answer these questions: An acoustics-architecture team responsible for designing a new large worship space will optimize music acoustics by incorporating the well principle in the main-floor congregational seating area. This should be a worship space where: (a) traditional music is important; (b) all finish surfaces are hard, including seating; and (c) the volume is large enough to insure a reverberation time in excess of 3 s when occupied, rising to 4.5 s or more when empty, with an occupied bass ratio of 1.4 or higher. Speech intelligibility would be handled by state-of-the-art electronic reinforcement, as has already been done in other similar situations. Then, various experiments would be conducted with and without pew or seat cushions added, and with and without heavy wood panels over the floor grills to by-pass the effect of the wells, before and after seats are installed. The data should provide the answers to the above three questions important for concert hall design. Tests with and without the heavy wood covers could determine the effects on congregational singing. I hope some reader will have the interest and support to perform this experiment.

Response to "Comment on 'Reducing seat dip attenuation'" [J. Acoust. Soc. Am. 110, 1260 (2001)]

W. J. Davies and T. J. Cox
School of Acoustics and Electronic Engineering, Salford University, Salford M5 4WT, United Kingdom

(Received 1 May 2001; revised 9 May 2001; accepted 12 June 2001)

This letter responds to Klepper's comments [J. Acoust. Soc. Am. **110**, 1260 (2001)] on the subject paper, which is concerned with ameliorating seat dip attenuation in auditoria by introducing a pit under the seats. Klepper asks what the effect of the pit will be on seat absorption and reverberation times. A little evidence is presented to support the idea that low-frequency absorption in an auditorium will increase with a pit. It is further speculated that reverberation times could be predicted by using a coupled space model. Klepper's suggestion of an experiment to answer his questions is supported. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1392382]

PACS numbers: 43.55.Br, 43.55.Fw, 43.55.Ka, 43.20.Fn [JDQ]

I. INTRODUCTION

In the subject paper[1] Davies and Cox investigated early scattered sound from seats as a cause of seat dip attenuation in auditoria. To reduce the attenuation they propose a scheme for modifying the scattered sound by introducing a pit or well under the seats. In his comments on this paper, Klepper asks three interesting questions about the practical effects of such a pit. The three questions are concerned with the effect that the pit might have on audience absorption and the prediction of reverberation time. One of us has some experience in measuring seating absorption,[2] and we offer some tentative responses to Klepper's questions in what follows.

II. EFFECT OF THE PIT ON SEATING ABSORPTION

The pit works by altering the phase of some of the scattered sound from the seats and floor, so that destructive interference is reduced at the listener's head. It is not intended to act as a low-frequency absorber. It seems quite possible, however, that the low-frequency absorption in the auditorium will be increased. If there is a grill over the pit (to support the seats) then the assembly may act as a Helmholtz resonator and this is likely to increase the random incidence absorption at low frequencies. In another scheme for reducing seat dip attenuation, due to Ando *et al.*,[3] devices are introduced into the floor which are specifically intended to be Helmholtz absorbers. One way of achieving this would be to use a ventilation system with small outlet boxes set into the floor between the seats. This has been tried in the laboratory and it was found[4] that the outlet boxes did increase the random-incidence absorption of unoccupied padded seating, as Fig. 1 shows. The resonant frequency of each box was calculated to be 249 Hz. (This measurement was performed in a chamber which was not very diffuse, so the data should be regarded as approximate, especially at low frequencies.)

III. EFFECT OF THE PIT ON RT PREDICTION

Klepper's other two questions relate to reverberation time (RT) prediction. We have no experimental evidence to cite for these, but it does seem likely that the pit is going to play a part in the reverberant sound field, at least at low frequencies. This might be modeled simply by treating the pit and the auditorium as coupled spaces, though the coupling coefficient will probably depend on frequency. In this model, the seats would not be treated as objects in the midst of a volume. Instead, they would lie on the boundary of each space and we would need absorption coefficients for sound incident on the underside of the seats to calculate the reverberant field in the pit. (One possible drawback here is that this concept would seem to further complicate the current debate on auditory spaciousness in auditoria—the listener is now surrounded by sound from below as well!) In any case,it would seem sensible to measure the absorption of some seats with a pit in a reverberation chamber before the hall design is finalized.

IV. CONCLUSION

None of the preceding speculation really provides an adequate answer to Klepper's questions, and we would be very pleased if an experiment were conducted (either at full

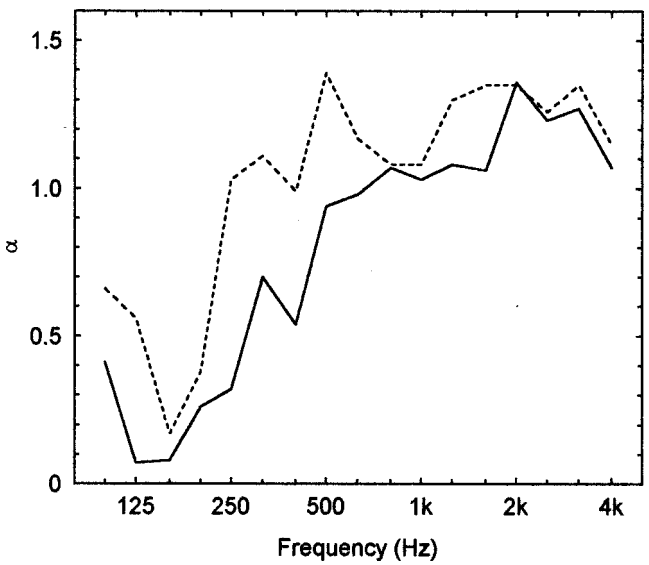


FIG. 1. Absorption coefficient of unoccupied well-upholstered seats on wooden floor. Three rows of six seats were placed in the corner of the chamber with 1-m row spacing and the front and side of the array left exposed. --- Floor ducts exposed; ——— floor ducts covered.

scale or in a scale model) to investigate further. We would welcome any opportunity to be involved with such an experiment.

ACKNOWLEDGMENT

The authors are grateful to R. J. Orlowski of Arup Acoustics for originally suggesting the idea of using ventilation outlets as resonant floor absorbers.

[1] W. J. Davies and T. J. Cox, ''Reducing seat dip attenuation,'' J. Acoust. Soc. Am. **108**, 2211–2218 (2000).

[2] W. J. Davies, R. J. Orlowski, and Y. W. Lam, ''Measuring auditorium seat absorption,'' J. Acoust. Soc. Am. **96**, 879–888 (1994).

[3] Y. Ando, M. Takaishi, and K. Tada, ''Calculations of the sound transmission over theater seats and methods for its improvement in the low-frequency range,'' J. Acoust. Soc. Am. **72**, 443–448 (1982).

[4] W. J. Davies, ''The Effects of Seating on the Acoustics of Auditoria,'' Ph.D. thesis, Salford University (1992).

Improving object classification with biomimetic sonar

Roman Kuc
Department of Electrical Engineering, Yale University, New Haven, Connecticut 06520-8284

(Received 25 October 1999; revised 18 February 2001; accepted 22 June 2001)

This paper presents improvements to a previously described biomimetic system to make it more robust in terms of localizing objects and increasing the types of objects that can be classified. Time-of-flight determination employs a large threshold and searches the echo envelope backward in time to find a smaller threshold crossing point, rather than using a single small threshold. This change eliminates small nuisance echoes, which can cause instabilities in object positioning. Second, the echo envelope is also searched forward in time to find the threshold crossing point, thus defining a window containing significant echo data. This changes the analysis from a fixed-duration data window to one that is variable and allows echo duration to be an object feature. Third, separate feature vectors were extracted from right and left receivers rather than employing a concatenated feature vector. The new system can then classify objects that are not isolated. The major significance of these changes is that it allows a greater variety of objects to be classified without degrading the performance of the original system. Experimental results compare original and new methods. Since biological sonars encounter similar problems, it is possible that they exploit principles similar to those described in this paper. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1395595]

PACS numbers: 43.60.Lq, 43.80.Ev, 43.60.Qv [WA]

Our previously described system[1] employs techniques observed in biological sonars[2,3] to classify objects from their echo waveforms. The key feature is that the system is adaptive in that it changes its configuration and location using the information in the detected echoes, allowing the object to be observed at a repeatable and favorable location.

Our sonar consists of three electrostatic transducers that can act as either transmitters or receivers.[4] Each transducer has a circular aperture with diameter $a=3.75$ cm and is resonant at 60 kHz ($\lambda=0.6$ cm in air), thus producing a 10° beam (full width at half-maximum amplitude levels). The center transmitter is flanked by two receivers separated by $D=4.5$ cm, which rotate to focus in on a object. The sonar mounts on a small robot arm to allow three translational and two rotational degrees of freedom. The sonar moves in a horizontal plane 11 cm above a smooth working surface and looks downward at a 45° angle. The system scans the environment by transmitting interrogation pulses every 50 ms and processing the echoes while performing a sector scan. Each receiver connects to a channel of an analog-to-digital converter (Gage CS1012) sampling at 2 MHz with 12-bit resolution. Utilizing echo information, the sonar moves to position an object along its axis at 15-cm range. This range is in the far field of each transducer ($r>a^2/\lambda$), thus causing the transmitter to insonify the object with plane waves and the receivers to detect plane waves. Locating the object in the sonar system near field ($r<D^2/\lambda$) provides a focusing capability and allows each receiver to obtain a different view of the object.

This paper describes three changes made to the original system, which affect time-of-flight (TOF) determination, echo analysis interval, and database structure. These changes increase the variety of objects that can be classified, specifically, those larger than the beamwidth and those that are not isolated. The original system had difficulties classifying these object types. Echoes from large extended objects did not allow stable placement and focus because small echoes from structures at the beam edges led to unstable positioning. When other objects are present in the beam, the resulting echoes produce feature vectors not represented in the database.

Original and new TOF methods are compared in Fig. 1. Originally, TOF T_1 was a three-step process operating on the echo waveform: T_a is the time the echo waveform first exceeds threshold τ_a, set equal to five times the standard deviation (s.d.) of the ambient zero-mean noise level. The time the waveform next crosses zero defines T_x. T_1 then equals T_x minus one period (17 μs or 34 samples). The main advantage of this method is its independence of the waveform amplitude, so long as the first peak exceeds the threshold. With this in mind, the emitter is driven so that the echo waveform from a small sphere near the focal region, where accurate positioning is important, exhibits its maximum value at the first positive peak. Hence, objects producing a strong initial echo were accurately and repeatably positioned. However, the original method exhibited two problems. First, small initial echoes, such as from a small cube that initially scatters from an edge, often exhibit maximum amplitudes around τ_a during the positioning process. When small echoes form the beginning of a composite echo, random TOF jumps on the order of the period can occur as the threshold was exceeded by adjacent peaks. Second, strong reflectors lying on the beam edge produce small echoes, which can affect system movement if they exceed τ_a. When these small echoes fall below τ_a after the movement, other echoes redirect the sonar back to the previous position, resulting in a limit cycle that fails to yield a stable object position.[5]

The new TOF method solves both problems. It is motivated by the appealing idea that the TOF should be based on strong rather than weak echoes, when strong echoes are

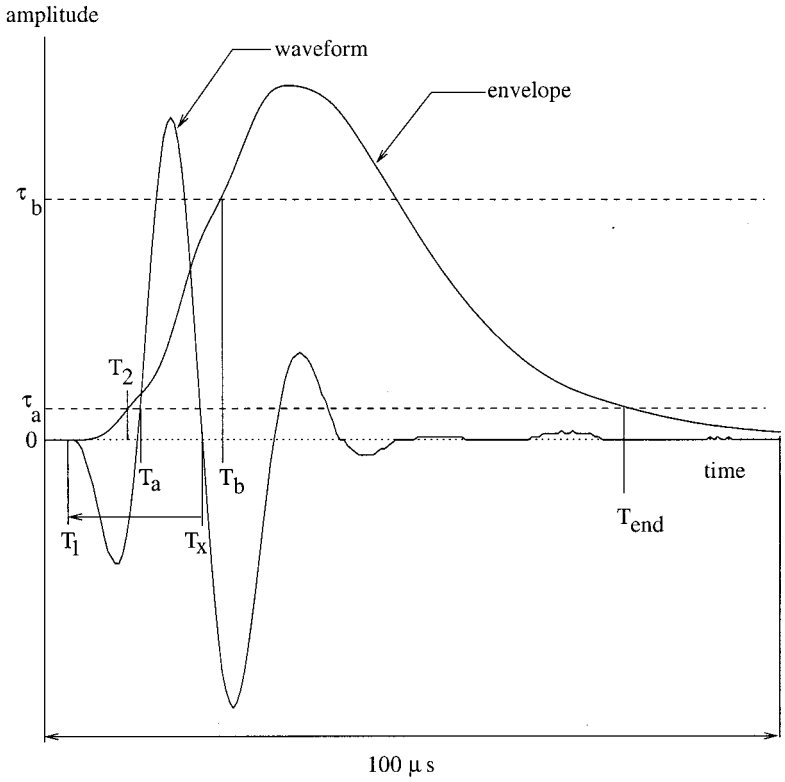


FIG. 1. Determining time of flight from the echo waveform and envelope.

present. The echo waveform is full-wave rectified and low-pass filtered to approximate its envelope. The envelope exceeds a large threshold τ_b, with $\tau_b > \tau_a$, at time T_b. The envelope is then searched backwards in time to find when it first falls below τ_a, this time then defining TOF T_2. In this case, τ_a equals the mean plus 5 times the s.d. of the ambient noise level in the envelope processor output. The envelope was normalized to have the same value of τ_a for both systems. Envelope smoothness, governed by the low-pass filter bandwidth, eliminates period-sized TOF jumps. Our low-pass filter was a second-order autoregressive filter with both poles located at $r = 0.94$.[6] While T_2 is slightly different than T_1, resulting in a slightly different object placement, both methods produce the desired stable and repeatable object placement. However, T_2 accomplishes this over a greater variety of objects because it ignores initial small-amplitude echoes. T_2 has an echo-amplitude dependent bias because it occurs after the true TOF. We can reduce this bias by fitting a parabola to the envelope around T_2 and using the parabola's minimum location as the TOF value.[7] We did not include this additional processing in our system because T_2 provided sufficient object positioning repeatability.

Object identity information is contained in a feature vector extracted from the echo envelope observed at each receiver. Originally, the processed echo waveform contained a fixed 512 samples. This data window size was sufficient to include all the echo components produced by the largest object in our original object set, a 2.54-cm-diameter O-ring. To increase the variety of objects our new system could examine, a variable data window replaced the fixed data window. The variable data window starts one period prior to T_2, to include low-level signals at the start of the echo, and ends when the echo envelope falls into the noise. After T_2 determination, the envelope is searched forward in time starting at time T_b to determine when it falls below τ_a, denoted T_{end} in Fig. 1. The analysis window ends after T_{end} at a point that makes the total window duration a multiple of 16, for feature extraction described below.

Originally, the feature vector was composed of elements equal to the average envelope energies in a set of contiguous segments, each containing 32 echo waveform samples. This resulted in 16 elements from each receiver, which were concatenated to form a 32-element feature vector. In the new approach, the segment was reduced to 16 samples (hence, the multiple of 16 in the data window). This smaller segment forms a better representation of the envelope, but yields larger 64-element feature vectors for the 512-point analysis window. The variable data window limits database growth, since the smallest objects in our set exhibit windows containing approximately 192 samples. The new method can also classify larger objects, or those producing more than 512 samples. This extended our experimental capabilities from washers and coins to plastic models of frogs and lizards. The single 32-element vector in the database was also replaced by separate vectors from the left and right receiver.

Figure 2 compares the new and old systems examining a single #8 machine washer. The same data durations (512 samples) are compared, with zeros padded to the new system data. The feature vectors are displayed as bar graphs above their respective waveforms. The feature vector in the new system is a better representation of the envelope and exhibits two clear peaks. Applying threshold detection in both backward and forward directions extracts 240 samples from the left receiver, forming a 15-element left feature vector, and 208 from the right receiver, forming a 13-element right feature vector. Waveform samples that are essentially noise are

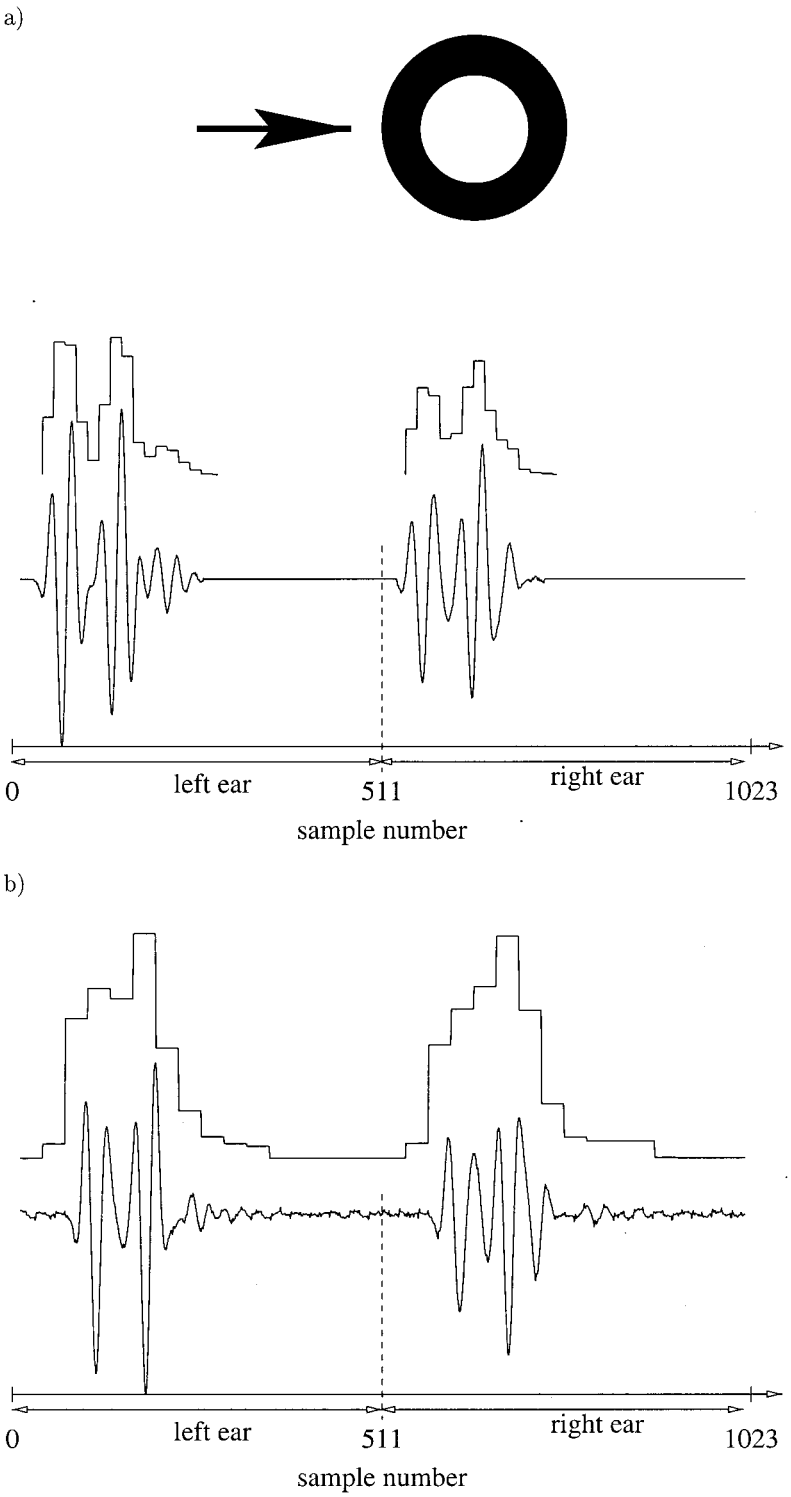


FIG. 2. Echoes from #8 machine washer. (a) New method. (b) Original method. Arrow shows top view of propagation direction of incident pulse.

eliminated. The original system employed a constant 512-point data window with larger segments that smooth out envelope features and included data containing mostly noise.

Figure 3 shows the case when two washers are encountered. Our system was trained using only single washers and only corresponding feature vectors are represented in the database. An observed feature vector is matched to a database entry by computing the sum of the squared differences between the observed vector and each database entry and taking the entry producing the minimum value. Our original system would not have classified this case because the observed 32-element feature vector is significantly different from any database entry. The sum of squares exceeding a large classification threshold prevents the closest match in the database from being chosen as the identity. In this case, rather than a guess, the system response is *I don't recognize the object*. The new system compares the observed vectors with only the number of elements in the database entry. In

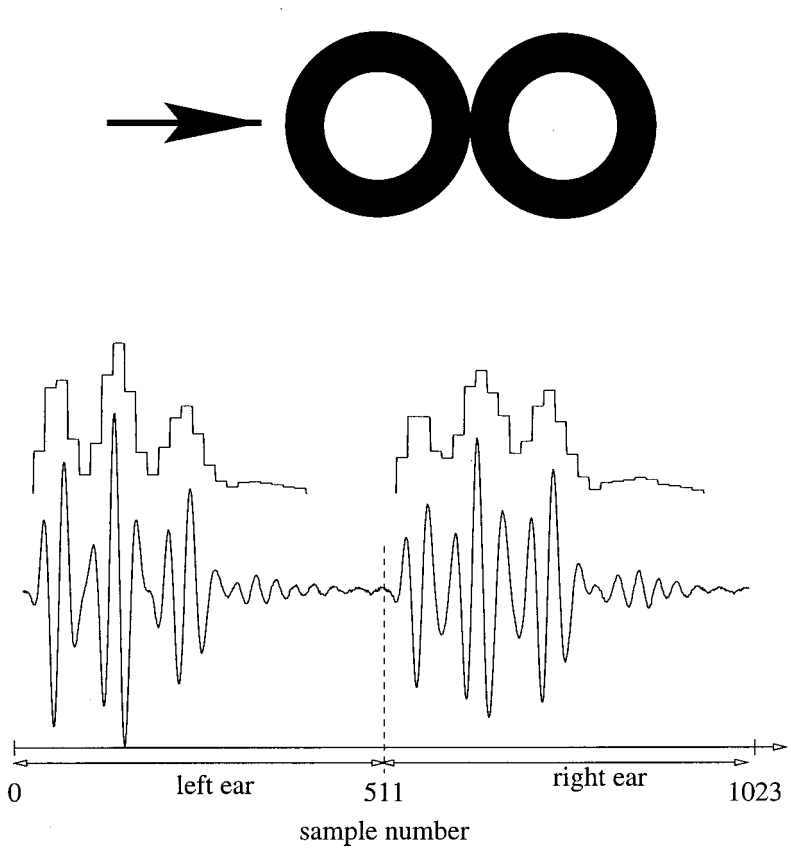


FIG. 3. Echoes from a pair of #8 machine washers, with both washer centers lying on the sensor axis.

this case, the new system recognizes the closer washer because the leading section of the observed vector forms a close match to the washer entry, shown in Fig. 2(a). The additional components in the observed feature vector indicate that there are additional objects to the washer. This information may direct the sonar to view the objects from another aspect angle.

Other experiments located the farther washer off to one side of the closer washer. The echoes from the farther washer then tended to be stronger at the ipsilateral receiver than at the other. As long as one receiver formed a feature vector that was sufficiently close to a washer database entry the washer was classified correctly. Classification failed when the other washer produced echoes that influenced the TOF values, which affect the positioning of the object with respect to the sonar. Since the echo waveform is strongly dependent on object location, it is not unexpected that correct classification failed. For the same reason, objects lying closer than the washer caused classification failure because their echoes determined the sonar location. For correct classification in the presence of multiple objects, the object represented in the database must determine the TOFs. Otherwise, the object location in the sonar beam will be different than that used in generating the database and the observed vectors will not find matches.

This paper described improvements and limitations to an adaptive sonar system mounted on a robot arm that classifies objects from echo waveforms. The time of flight is determined in a more robust manner by detecting a large echo and searching backward in time to find the threshold crossing point. The data window size is made variable so that only significant echo waveform data are processed. Separate feature vectors extracted from the echoes detected at the right and left receivers allow more robust object classification when only partial views are available.

ACKNOWLEDGMENTS

The research described in this paper was supported by the National Science Foundation under Grant No. IRI-9504079 and the U.S. Civilian Research and Development Foundation Grant No. CRDF-UE 2-300.

[1] R. Kuc, "Biomimetic sonar system recognizes objects using binaural information," J. Acoust. Soc. Am. **102**, 689–696 (1997).
[2] *Animal Sonar: Processes and Performance*, edited by P. E. Nachtigall and P. W. B. Moore (Plenum, New York, 1988), Chaps, 5, 6, pp. 655–835; Proceedings of NATO Advanced Study Institute on Animal Sonar Systems, 10–19 September 1986, Helsingor, Denmark.
[3] W. W. L. Au. *The Sonar of Dolphins* (Springer, New York, 1993).
[4] C. Biber, S. Ellin, E. Sheck, and J. Stempeck, "The Polaroid ultrasonic ranging system," in Proceedings of the 67th Audio Engineering Society Convention, 1990. Reprinted in Polaroid Ultrasonic Ranging System Handbook.
[5] R. Kuc, "Three-dimensional docking using qualitative sonar," Rob. Auton. Syst. **11**, 213–219 (1993).
[6] R. Kuc, *Introduction to Digital Signal Processing* (McGraw-Hill, New York, 1988).
[7] B. Barshan and R. Kuc, "Active sonar for obstacle localization using envelope shape information," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Toronto, Canada, 13–17 May 1991, pp. 1277–1280.

Masked threshold difference due to masker harmonicity: Uniform modulation rates of auditory filter outputs versus periodicity in waveform fine structure

William C. Treurniet[a] and Darcy R. Boucher
Communications Research Centre, Box 11490, Station H, Ottawa, Ontario K2H 8S2, Canada

(Received 21 March 2001; accepted for publication 11 June 2001)

In a previous paper [Treurniet and Boucher, J. Acoust. Soc. Am. **109**, 306–320 (2001)], a model was presented which successfully accounted for the lower threshold obtained using a harmonic masker instead of a similar inharmonic masker. For the harmonic masker, the frequencies of partials were equally spaced, so the envelope modulation rates of auditory filter outputs were invariant across filters. However, for the inharmonic masker, the intervals between adjacent partials were not equal and this resulted in a decreased uniformity of modulation rates across filters. The model proposed that the lower uniformity interferes with detection of a probe-induced change in the modulation rates, thus accounting for the masked threshold difference. This paper shows that an inharmonic masker yields thresholds similar to a harmonic masker provided that the modulation rates are uniform across the auditory filters involved in processing the masker and probe. Thus, the lowered threshold associated with a harmonic masker appears to arise from uniform modulation rates across auditory filters, and does not require periodicity in the fine structure of the waveform produced by partial frequencies that are integer multiples of a fundamental. [DOI: 10.1121/1.1391245]

PACS numbers: 43.66.Dc, 43.66.Ba [SPB]

I. INTRODUCTION

Treurniet and Boucher (2001) recently demonstrated that a noise probe signal was less effectively masked by a harmonic sound than by a similar inharmonic sound. This observation occurred only when auditory filters could not resolve the partials. The threshold difference increased monotonically with the inharmonicity of the masker, but did not depend on the level of the masker. Further, an across-channel mechanism was implicated when a harmonic masker flanked by harmonic bands yielded a lower threshold than a harmonic masker flanked by inharmonic bands. The effect did not occur in a forward-masking experiment where the signal immediately followed the masker.

A model developed to explain the observed differences recognizes that the envelope of the output of an auditory filter is modulated when the filter passes two or more sinusoids, and that the modulation spectrum is determined by the differences among filter-input frequencies. For a harmonic masker, the frequency differences between adjacent partials are identical, and all auditory filters have the same dominant modulation rate. For an inharmonic masker, however, the frequency differences are not the same and the envelope modulation rate is less uniform across filters. The model proposes that a lower uniformity interferes with detection of a probe-induced change in the modulation rates, thus accounting for the masked threshold difference. When the model treated variable filter modulation rates as ''noise'' in addition to the energy ratio of the probe and masker, it more accurately predicted observed thresholds.

In the experiments described by Treurniet and Boucher (2001), harmonic maskers were created by selecting sets of partials from a harmonic series with an 88-Hz fundamental and 45 consecutive partials. Inharmonic maskers differed in that the partial frequencies were perturbed to nearby values that were not integer multiples of the fundamental frequency. For these maskers, therefore, the degree of uniformity of the filter-output modulation rates was confounded with whether or not all partials were integer multiples of a fundamental. That is, for the harmonic masker, the modulations were highly uniform and all partials were also integer multiples of a fundamental; for the inharmonic masker, the modulations were not uniform and all partials were also not integer multiples of a fundamental. The lower masked threshold with uniform spacing of the partials might have resulted from the uniform modulation rates as proposed in the model, or from a probe-induced change in the sound arising from periodicity in the waveform's fine structure. The present experiment was designed to distinguish between these two factors (*modulation uniformity versus partial interval uniformity*) as predictors of the masked threshold difference.

In the model proposed by Treurniet and Boucher (2001), the threshold is affected by modulation rate *variability*. Modulation in the experiments occurred since all intervals between adjacent partials were smaller than the bandwidths of the auditory filters. Such modulation would disappear if the distance between adjacent partial frequencies were increased to exceed the filter bandwidths. However, it would be reintroduced in some auditory filters by the addition of new partials at nearby frequencies. That is, when the center frequencies of any two auditory filters are sufficiently far apart, the modulation rate of each filter output is influenced only by partials situated within the bandwidth of the filter. Therefore, modulation rates across filters may be made uni-

[a]Now at VoiceIQ, 175 Columbia St., Waterloo, ON N2L 5Z5, Canada.

TABLE I. Definitions of maskers and probe.

Partial no.	Phase (deg)	Masker partial frequencies (Hz)				Probe Fc (40-Hz BW)
		Harmonic	Hybrid 1	Hybrid 2	Inharmonic	
1	286	880	889	863	863	920
2	87	968	977	951	977	
3	164	1320	1311	1327	1327	1363
4	155	1408	1399	1415	1399	
5	110	1760	1743	1777	1777	1804
6	351	1848	1831	1865	1831	
7	143	2200	2209	2221	2221	2255
8	277	2288	2297	2309	2297	
9	37	2640	2619	2657	2657	2682
10	213	2728	2707	2745	2707	

form by stimulating each filter with a pair of partials, where the frequency difference between the two partials of each pair is the same. The separations between adjacent pairs need not be integer multiples of a fundamental frequency. A complex masker may be constructed by summing the partials at all filter positions. Since the uniform modulation of the filters is a property of harmonic maskers, and the nonuniform spacing of the partials is a property of inharmonic maskers, such a complex is called *a hybrid* masker in this paper.

In the following experiment, listeners were presented with harmonic, inharmonic, and hybrid maskers. The model proposed by Treurniet and Boucher (2001) predicts that the inharmonic masker will yield a significantly higher threshold than the harmonic masker. Further, since the hybrid masker gives rise to uniform filter-output modulations, it should yield the same threshold as the harmonic masker. Such results would argue against the contribution of other effects arising from the uniform spacing of partials in the harmonic masker.

II. METHOD

The properties of a harmonic and an inharmonic masker, as well as two hybrid maskers, are presented in Table I. Each masker consisted of five pairs of equal-amplitude partials, each partial having a random starting phase, thus ensuring that modulations were not coherent across auditory channels. The lack of coherence means that any effect of the harmonic masker on masked threshold cannot be due to comodulation masking release (Hall, 1987). For the harmonic masker, the members of each pair were separated by 88 Hz, while each pair was separated from the adjacent pair by 352 Hz (i.e., 4×88 Hz). Since the latter exceeds the equivalent rectangular bandwidth (ERB) of about 320 Hz for the highest masker frequencies in Table I (e.g., Moore, 1997), each affected filter should process primarily a single pair of partials. The partials of the inharmonic masker were similar, but the frequencies were perturbed to nearby prime numbers. This ensured that the interval between members of successive pairs was never the same. In this masker, each pair was separated from the adjacent pair by an average of 370 Hz.

Two hybrid maskers were constructed based on the partial frequencies of the inharmonic masker. In the first case, the frequency of the lower partial of each pair was given the same value as the corresponding partial in the inharmonic masker. The upper partial was then assigned a frequency that was 88 Hz higher than that of the lower partial. In the second case, the upper partial of the pair was assigned the same frequency as the corresponding partial in the inharmonic masker, and the lower partial was given a frequency that was 88 Hz lower than that of the upper partial. For the two hybrid maskers, each pair of partials was separated from the adjacent pair by an average of 344 and 360 Hz, respectively.

A noise probe consisting of a single band extending over the frequency range of the ten masker partials would be too easily detected in the filters not receiving any masker energy. To prevent this, a probe was constructed by filtering wideband Gaussian noise to yield five 40-Hz passbands located in the regions of the partial pairs in the maskers. Each passband fell inside the corresponding partial-pair frequency range common to all four maskers. The five bands are also listed in Table I. A 2-s interval of noise was generated, and the probe was extracted from this longer interval after selection of a random starting point in the sequence.

A. Apparatus

Audio was generated with 16-bit linear resolution at a sampling rate of 44.1 kHz, and was presented via a Digital Audio Labs CardD digital sound card and an external Spectral Synthesis model 2218 D/A converter. The signal from the D/A converter was fed to Stax SRM-1 headphones located in an audiometric chamber. All sounds were presented diotically. A listener interacted with a computer screen through a window in the chamber using a mouse control. Software to control the experiment was developed in-house.

B. Experimental procedure

An adaptive three-alternative forced-choice psychophysical procedure was used to measure the masked thresholds. On a given trial, the masker was presented three times. One of the presentations was chosen randomly as a target interval, and the probe was added to the masker during this interval. The three audio intervals were presented sequentially in synchrony with a visual cue on the computer screen. The listener's task was to indicate which was the target interval by selecting a corresponding button on the screen with the mouse. The level of the probe was raised by a fixed

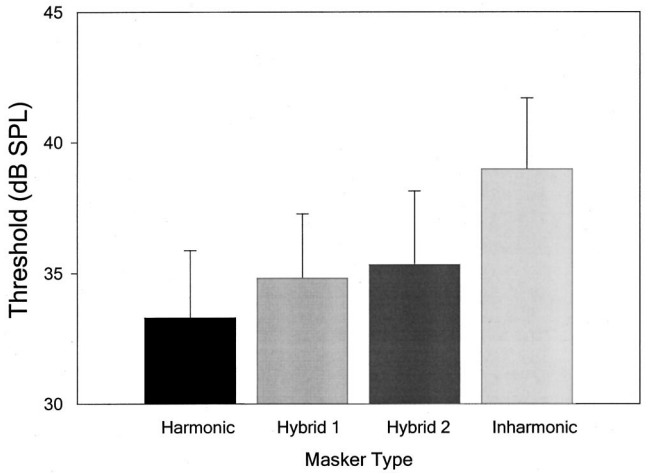


FIG. 1. Average masked thresholds for the harmonic, inharmonic, and hybrid maskers. Maskers were presented at an overall level of 55 dB SPL. The vertical bars indicate 1 standard deviation of the threshold estimates across listeners.

amount if the response was incorrect, and it was lowered by the same amount if the response was correct on the current trial as well as on the two preceding trials. Each session consisted of a total of 12 reversals in direction. The change in signal level was 4 dB before the first two reversals and 1 dB subsequently. The measured threshold level was the mean of the levels at each of the last ten reversals. In theory, this procedure converges to the signal level on the psychometric function where 79.4 percent of signals are detected (Levitt, 1971).

The masker and probe signals were each 400 ms in duration, including 50-ms raised-cosine onset and offset ramps, and were presented simultaneously. The interpresentation interval within the set of three presentations was 300 ms, and the time between successive sets was 1200 ms unless the listener failed to respond within that time. When the response was delayed, the next iteration began immediately after the response was given. The overall level of the masker was 55 dB SPL, and the initial level of the probe was 5 dB higher.

Of the six listeners who participated in the experiment, three had taken part in similar experiments. All listeners were males between the ages of 20 and 55 with no history of hearing disorders. Conditions were presented in consecutive blocks of trials in order to minimize the effect of stimulus uncertainty, and the order of conditions was randomized for each listener. Each listener's final threshold was the average of three threshold determinations for each condition.

III. RESULTS

The trends in the data were consistent across listeners, so only average masked threshold are presented. Figure 1 shows the average thresholds for the harmonic, hybrid, and inharmonic maskers. An analysis of variance showed that the within-subject factor of masker type was significant [$F(3,15)=92.913$, $p<0.0001$]. Multiple comparisons of means with the Scheffé test (e.g., Hays, 1963) showed that all of the differences between means were significant ($p<0.01$) with the exception of the difference between the two hybrid maskers. The latter difference of 0.5 dB was not significant even at the reduced level of $p<0.05$. The differences between the harmonic masker and the hybrid maskers were 1.5 and 2.0 dB, while the inharmonic masker differed from the hybrid maskers by 3.7 and 4.2 dB.

IV. DISCUSSION

As predicted by the model (Treurniet and Boucher, 2001), a significant difference was found between the inharmonic and the hybrid maskers. However, the prediction of no difference between the harmonic and the hybrid maskers was not confirmed. The latter difference, although small, was significant and requires an explanation.

The significant difference between the inharmonic and hybrid maskers is consistent with the *modulation uniformity* hypothesis described by the model. That is, only the hybrid masker had uniform filter-output modulations, while both masker types shared the lack of uniformity in the spacing of the partials. Therefore, the difference in uniformity of filter-output modulations can explain the result.

The difference between the harmonic masker and the hybrid maskers was not predicted and appears, at first glance, to support the alternative *partial interval uniformity* hypothesis. The maskers shared the characteristic of uniform filter-output modulations and differed in the regularity of the spacing between adjacent partials. However, the uniformity of filter-output modulations for the hybrid maskers depends on the assumption that the bandwidth of an auditory filter is restricted to the region specified by the ERB calculation. A relatively small but consistent difference could be taken as evidence of energy leakage between filters, a possibility suggested by one form of the filter model initially proposed by Patterson *et al.* (1982). Such leakage could make the filter-output modulations from the hybrid maskers less uniform than those from the harmonic masker, resulting in the slightly elevated masked threshold observed in the data. According to this interpretation, the difference between the harmonic and hybrid maskers is also consistent with the *modulation uniformity* hypothesis.

V. CONCLUSION

The results showed that the average hybrid masker threshold was about 4 dB lower than the inharmonic masker threshold. Since both maskers shared the nonuniform spacing between masker partials but differed in the uniformity of auditory filter-output modulations, the latter is a more likely basis for explaining this difference. The results also showed that the average hybrid masker threshold was about 1.75 dB higher than the harmonic masker threshold. This may be evidence that leakage occurs between filters with center frequencies separated by more than 1 ERB. Such leakage could slightly decrease the uniformity of auditory filter-output modulations for the hybrid maskers, and thus explain the observed small increment in threshold.

In general, the filter envelope-modulation-rate variability hypothesis proposed by Treurniet and Boucher (2001) is able to explain all the masked threshold differences observed

in the experiment, while other interpretations based on periodicity of waveform fine structure arising from uniform spacing of masker partials received less support.

Hall, J. W. (**1987**). ''Experiments on comodulation masking release,'' in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ).

Hays, W. L. (**1963**). *Statistics for Psychologists* (Holt, Rhinehart and Winston, New York).

Levitt, H. (1971). ''Transformed up–down methods in psychophysics,'' J. Acoust. Soc. Am. **49**, 467–477.

Moore, B. C. (**1997**). *An Introduction to the Psychology of Hearing* (Academic, New York).

Treurniet, W. C., and Boucher, D. R. (**2001**). ''A masking level difference due to harmonicity,'' J. Acoust. Soc. Am. **109**, 306–320.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). ''The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold,'' J. Acoust. Soc. Am. **72**, 1788–1803.

On scattering from a bubble located near a flat air–water interface: Laboratory measurements and modeling

George Kapodistrias[a)]
Applied Physics Laboratory and Department of Mechanical Engineering, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105

Peter H. Dahl
Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105

(Received 2 February 2001; revised 31 May 2001; accepted 8 June 2001)

Scattering by a single bubble near a flat air–water interface is investigated theoretically and experimentally. A ray-acoustic interpretation is used to describe the four scattering paths, from source to bubble to receiver, that determine the response of the bubble. Multiple scattering effects are accounted for using a closed-form solution derived from the multiple scattering series. Experiments are performed by placing a bubble with radius $a \approx 425\,\mu$m on a fine nylon thread, which is approximately 100 μm in diameter and practically transparent to sound, at a distance d from the interface. The primary variable is d and it ranges from $1a$ to $100a$. The bubble is excited by tone bursts with a center frequency of 120 kHz, with the transducers arranged in both bistatic and monostatic configurations. Theory and experiment are in good agreement, verifying the dominant effect of the four paths in the response of the bubble, with multiple scattering playing a role for $kd<1$, where k is the wave number of the medium. In the long-range limit our simulations agree with those of Ye and Feuillade [J. Acoust. Soc. Am. **102**, 798–805 (1997)] including the shifting of the bubble's resonant frequency. The dependence of scattering on transducer arrangement, range to bubble, grazing angle, and phase relation among the four paths, *vis-à-vis* monostatic and bistatic scattering, is discussed. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1390335]

PACS numbers: 43.20.Fn, 43.30.Cq, 43.30.Gv, 43.30.Zk [DLB]

I. INTRODUCTION

Bubbles in the vicinity of the air–sea interface can easily dominate the return scatter of sound from the sea surface,[1] owing to their scattering efficiency compared to Bragg scattering. One reason is resonance scattering by individual bubbles, which predominates in the 10–100 kHz frequency range. Another reason, which is the primary subject of this work, is that a near-surface bubble can scatter sound via four distinct paths: the direct one, plus three paths created by the interaction with the interface (see Fig. 1).

The basic concept of scattering via these four paths was put forth by Crowther,[2] who incorporated them into a model for backscattering from a subsurface bubble layer. Sarkar and Prosperetti[3] showed that Foldy's effective medium theory,[4] a theory that incorporates multiple scattering effects, can be reconciled with Crowther's approach by applying the WKB approximation to Foldy's theory. Sarkar and Prosperetti demonstrated clearly how the four paths emerge as a useful model for describing the process associated with backscattering from subsurface bubbles. More important, they recognized that for a rough surface and a monostatic configuration (i.e., a source and receiver that are co-located) two of the four paths add in phase, since these two paths are exactly the same and thus are affected in the exact same way by the interaction with the surface. McDaniel[5] also discusses the four-path formalism in the context of backscattering from near-surface bubble layers that are thick compared to the acoustic wavelength. Dahl *et al.*[1] implemented an expression for backscattering via the four paths into a model that can be used for inverting field measurements. Specifically, they estimated the depth-integrated extinction cross section per unit volume, due to both backscattering and absorption by near-surface bubbles, from measurements of backscattering from the sea surface at variable wind speed. Their results were consistent with estimations of near-surface bubble concentrations derived using alternative methods.

The manner by which the four paths add up, with respect to the distance, d, from the interface, results in an oscillatory scattered pressure field, whose amplitude depends on the surface reflection coefficient, R, which, in turn, depends on the degree of surface roughness *vis-à-vis* the acoustic wavelength. (We discuss the rough surface case, along with experimental data, in a follow-up paper.) For scattering from a bubble beneath a perfectly flat, pressure release surface, R is -1, a value we use in our model. Later in this paper we shall demonstrate that the apparent backscattering cross section of a single bubble near such a surface can increase by a factor of 16 (i.e., a gain of 12 dB over the target strength of a single bubble in the free field). Furthermore, as the bubble gets closer to the interface, effects due to multiple scattering between the bubble and the interface are induced.

Although the theory concerning scattering from bubbles near the sea surface based on four paths is well established, explicit experimental verification is lacking. In this paper, we perform both a theoretical and an experimental study of scattering from a single bubble close to a flat, pressure release interface, addressing both monostatic and bistatic conditions.

[a)]Now at: ATL Ultrasound, Philips Medical Systems, 22100 Bothell Everett Highway, Bothell, WA 98021; electronic mail: george.kapodistrias@philips.com

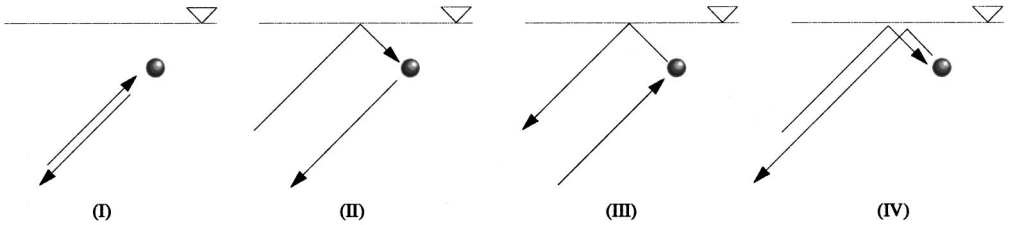


FIG. 1. The four paths that determine scattering from subsurface bubbles.

In Sec. II we develop our model used for comparison with our laboratory measurements. The model is based on ray synthesis of the aforementioned four paths, while higher-order scattering terms, associated with multiple scattering between the bubble and the surface, are accounted for using a simplified expression derived from the multiple scattering series. A number of numerical examples are given, and a detailed discussion is presented to elucidate the effect of multiple scattering as a function of frequency and depth of the bubble. We also compare our model with an alternative one developed by Ye and Feuillade.[6]

The experimental method is described in Sec. III, while theoretical and experimental results are compared and interpreted in Sec. IV. Here, our model is compared with monostatic and bistatic scattering measurements from a bubble of radius $a \approx 425\,\mu$m, located at a variable distance d from the interface. A summary of the work is given in Sec. V.

II. THE RESPONSE OF A SINGLE BUBBLE CLOSE TO A FLAT AIR–WATER INTERFACE

For monostatic sound scattered by a single air bubble close to a flat air–water interface, we are aware of a solution by Ye and Feuillade[6] that is valid for the low ka range and for a source/receiver located at a distance, r, from the bubble that is large compared to the distance, d, of the bubble from the interface (that is, when $d/r \ll 1$). Therefore the paths between the bubble and source/receiver shown in Fig. 1 can be approximated as parallel; henceforth we refer to this condition as the ''long-range limit.'' In their model, the incident field was described in terms of a direct wave from the source, its reflection from the surface, plus waves that underwent multiple scattering between the bubble and the interface, while the scattered field included contributions from the bubble and its image. Ye and Feuillade also derived equations for the change of the damping and the resonance frequency of the bubble. Their calculations showing an increase in the resonance frequency of the bubble, due to its proximity to a free surface, are in excellent agreement with theoretical results published by Strasberg[7] and with measurements performed by Lauer (see pp. 117–122 in Ref. 8).

Another approach to this same problem is that by Gaunaurd and Huang,[9] who use spherical harmonics in their model. However, their use of a real valued wave number for the gas of the bubble accounted only for radiation damping, in which case the resonance peak of the bubble becomes exceedingly high. This feature carried over in their calculation of the frequency response curves, as commented on by Strasberg.[10] In addition, their calculations for the frequency response of an 100 μm bubble at a distance $d=5$ cm from the surface are at variance with the calculations published for the same case by Ye and Feuillade.

In this paper the theoretical analysis is performed for an arbitrary geometry [shown in Fig. 2(I)]. The positions of the source and the receiver are defined by the Cartesian coordinates $[X_S, Z_S]$ and $[X_R, Z_R]$, respectively, while the bubble is located at $[X_B, Z_B]$ (note that $|Z_B| = d$). In addition, the bubble is assumed to be in the far field and on the axis of the combined beam of the source/receiver system. The speed of sound in the water is constant ($c = 1490$ m/s), and the losses due to attenuation are negligible.

Using the notation of Fig. 2, the pressure at the receiver due to path (I) becomes (dropping the time dependence $e^{-i\omega t}$)

$$p_{\mathrm{I}} = \frac{e^{ikr_{SB}}}{r_{SB}} f_B \frac{e^{ikr_{BR}}}{r_{BR}}, \tag{1}$$

where the ranges r_{SB} and r_{BR} are identified in Fig. 2(I), k is the acoustic wave number for the water medium, and f_B is the complex scattering function of the bubble. Here, and similarly with paths (II)–(IV), we take the incident pressure to be of unit amplitude, i.e., equal to $e^{ikr_{SB}}/r_{SB}$ for path (I).

For this work we shall use an f_B valid for $ka \ll 1$, which (consistent with the $e^{-i\omega t}$ dependence) equals

$$f_B = \frac{a}{(f_{\mathrm{res}}/f)^2 - 1 - i\delta}. \tag{2}$$

In Eq. (2), f is the frequency of the insonifying wave, f_{res} is the resonance frequency of the bubble, and δ is the total damping coefficient (see Ref. 11), which is the sum of the radiation, δ_{ka}, thermal, δ_{th}, and viscous, δ_v, damping components. For higher values of ka, an f_B based on the general solution, derived by Anderson,[12] should be used. Comments on the differences between the two scattering functions have been provided by the authors in Sec. II of Ref. 13.

For the forward surface reflection [i.e., the incident wave in path (II)], we replace the source with its mirror image, as shown in Fig. 2(II), and take R to be ideally -1 to account for the presence of the interface. Thus, the pressure at the receiver due to path (II) is

$$p_{\mathrm{II}} = R \frac{e^{ik(r_{SI}+r_{IB})}}{r_{SI}+r_{IB}} f_B \frac{e^{ikr_{BR}}}{r_{BR}}. \tag{3}$$

Following the same procedure, path (III) [see Fig. 2(III)] is written as

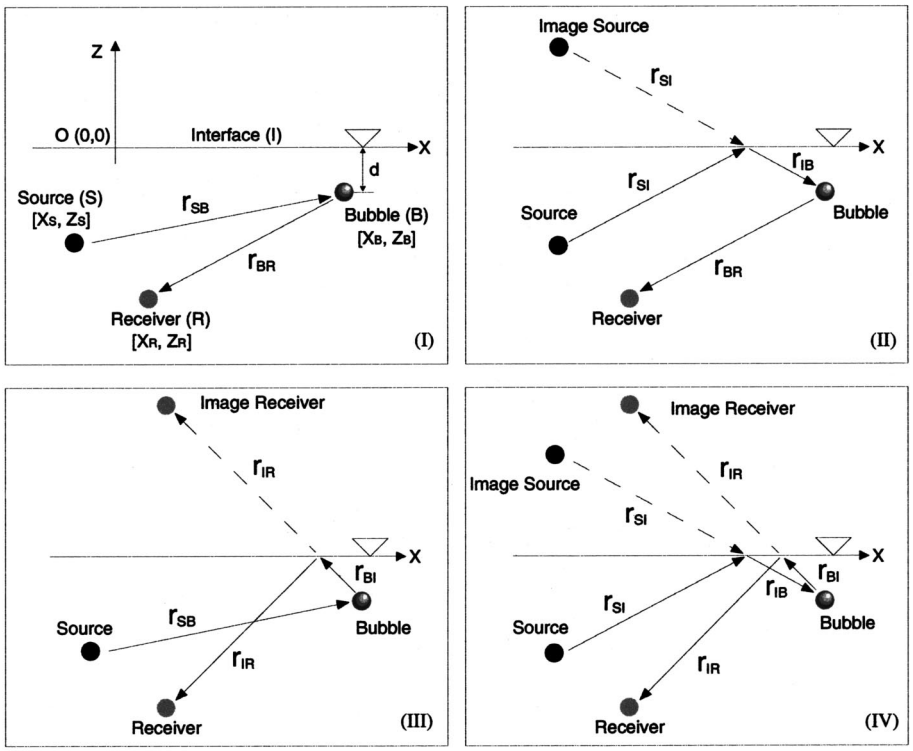


FIG. 2. Diagram showing the four paths, (I)–(IV), between source, bubble, and receiver used to describe the problem.

$$p_{\mathrm{III}} = R\frac{e^{ikr_{SB}}}{r_{SB}} f_B \frac{e^{ik(r_{BI}+r_{IR})}}{r_{BI}+r_{IR}}, \tag{4}$$

while path (IV) [see Fig. 2(IV)] becomes

$$p_{\mathrm{IV}} = R^2\frac{e^{ik(r_{SI}+r_{IB})}}{r_{SI}+r_{IB}} f_B \frac{e^{ik(r_{BI}+r_{IR})}}{r_{BI}+r_{IR}}. \tag{5}$$

The total scattered pressure at the receiver due to the aforementioned four paths equals their coherent sum, which we identify as

$$p_T^0 = p_{\mathrm{I}} + p_{\mathrm{II}} + p_{\mathrm{III}} + p_{\mathrm{IV}}. \tag{6}$$

Later in this section we show that for many practical applications the response of a single bubble close to a pressure release interface is adequately described by the summation of the four paths only. However, this formulation does not account for the multiple scattering effects caused by the proximity of the bubble to the interface. To do so, we utilize an image bubble and write the higher-order scattering terms in a form derived from the multiple scattering series (MSS). In an earlier publication,[13] we provide a detailed description of the MSS (see Sec. I of that reference) for a somewhat similar problem involving multiple scattering effects, which is valid for scattering from two arbitrary-sized bubbles with scattering functions, f_B, expressed in terms of spherical harmonics [see Eq. (12) in Ref. 13]. For two bubbles with identical f_B associated with monopole scattering such as given by Eq. (2), that expression reduces to a much simpler one, first obtained by Twersky.[14] For the problem at hand, we adapt Twersky's expression to describe multiple scattering between the bubble and its image, and define what we shall refer to as the multiple scattering correction (MSC), which is

$$\mathrm{MSC} = \frac{1}{1 - R\dfrac{e^{ik2d}}{2d} f_B}. \tag{7}$$

The product of Eqs. (6) and (7) gives the total scattered pressure at the receiver, p_T;

$$p_T = \frac{p_T^0}{1 - R\dfrac{e^{ik2d}}{2d} f_B}. \tag{8}$$

Equation (8) allows calculation of the scattering response of a single bubble close to a pressure release interface. With minor modification, it can be used with an f_B based on the exact solution developed by Anderson for scattering by a single bubble in the free field. The ability to incorporate the exact solution to the model becomes important for problems involving bubbles with radii comparable to the wavelength of the insonifying wave. Note that the same equation can be used to calculate the scattering response of a bubble close to a rigid smooth surface by setting $R=1$.

In this paper, Eq. (8), combined with Eq. (2), will be used to model the experiments, with numerical results reported in terms of bistatic target strength (TS_{bi}) for the bistatic cases and monostatic target strength (TS) for the monostatic cases, both with units of dB *re* 1 m^2. The scattered pressure is related to TS_{bi} using

$$\mathrm{TS}_{bi} = 10\log\left(\frac{|p_T|^2}{|p_{\mathrm{inc}}|^2}\cdot r_{BR}^2\right), \tag{9}$$

where $p_{\mathrm{inc}} = e^{ikr_{SB}}/r_{SB}$ and for the monostatic configuration $r_{SB} = r_{BR}$. Equivalently, the TS of a single bubble in the free field is given by

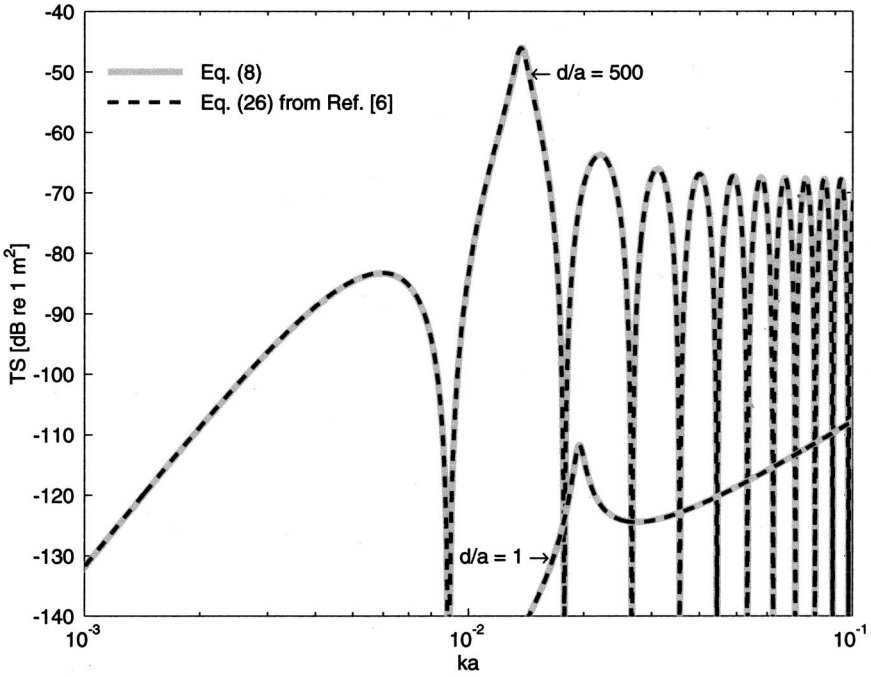


FIG. 3. Target strength for backscattering (i.e., monostatic configuration) from a bubble of radius 100 μm located close to a pressure release interface. Calculations performed using our Eq. (8) (solid gray line) are compared with those calculated using Eq. (26) from Ref. 6 (dashed black line). The slant range between the source/receiver and the bubble is 141 m, and the grazing angle is 45°.

$$\mathrm{TS}_{ff}=10\log\left(\frac{|p_{\mathrm{I}}|^2}{|p_{\mathrm{inc}}|^2}\cdot r_{BR}^2\right)=10\log|f_B|^2. \tag{10}$$

For a monostatic configuration and the aforementioned long-range limit, calculations performed using our Eq. (8) are in agreement with numerical results calculated using Eq. (26) from Ref. 6, including the increase of the resonance frequency of the bubble as it gets close to the surface. This is shown in Fig. 3, in which we plot the response of a single bubble at two different depths for variable frequency. The gray solid line is the frequency response (in terms of ka) calculated using Eq. (8), expressed in terms of TS using Eq. (9). The black dashed line (that completely overwrites the gray line) is 20 times the logarithm of the magnitude of Eq. (26) from Ref. 6. The calculations were performed for backscattering from a single bubble of radius $a=100\,\mu$m located at $[X_B,Z_B]=[100\,\mathrm{m},-d\,\mathrm{m}]$. The source/receiver was located at $[X_S,Z_S]=[X_R,Z_R]=[0\,\mathrm{m},-100\,\mathrm{m}]$. The slant range between the bubble and the source/receiver equals 141 m, while the grazing angle equals 45°.

Equation (8) provides a convenient way to study the effect of multiple scattering on the response of a bubble close to a flat, air–water interface. It readily shows that scattering is defined by the four paths and is affected by multiple scattering only for an appropriate combination of insonifying frequency and distance of the bubble from the interface. This is shown in Figs. 4(a) and (b). In Fig. 4(a) we plot the theoretical target strength calculated using Eqs. (8) and (9) for variable frequency and fixed bubble size (expressed in terms of ka on the y axis), and variable bubble depth d at each frequency [expressed in terms of $\log(kd)$ on the x axis]. In Fig. 4(b) we plot the decibel equivalent of MSC (i.e., $20\log|\mathrm{MSC}|$) for the same parameters as in Fig. 4(a). The numerical calculations were performed for backscattering from a bubble of radius $a=425\,\mu$m located at $[X_B,Z_B]=[100\,\mathrm{m},-d\,\mathrm{m}]$, with the source/receiver located at $[X_S,Z_S]=[X_R,Z_R]=[0\,\mathrm{m},-100\,\mathrm{m}]$.

In Fig. 4(a) an oscillation as a function of both ka and kd space is identified. In Sec. IV we show that this oscillation is due to the coherent addition of the four paths. In the same figure a multiple scattering effect is shown by the increase in resonance frequency for decreasing d. At $kd=32$ [or $\log(kd)=1.5$] the strongest scattering is at $ka\approx0.0136$ (equivalent to the resonance frequency of a bubble in the free field), while for decreasing kd the strongest scattering curves toward a higher ka. The effect of multiple scattering in the response of the bubble is more clearly demonstrated in Fig.

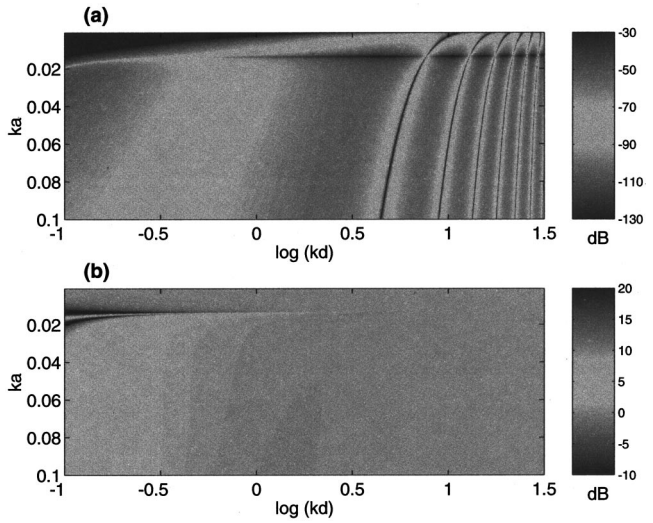


FIG. 4. (a) Target strength of a single bubble ($a=425\,\mu$m) close to a pressure release interface for variable frequency (ka) and depth (kd), where d is the depth below the air–water interface, and k is the acoustic wave number of the water medium. (b) Contribution of multiple scattering, calculated using the quantity $20\log|\mathrm{MSC}|$. Note the contribution is highest near resonance ($ka=0.0136$), and for $kd<1$.

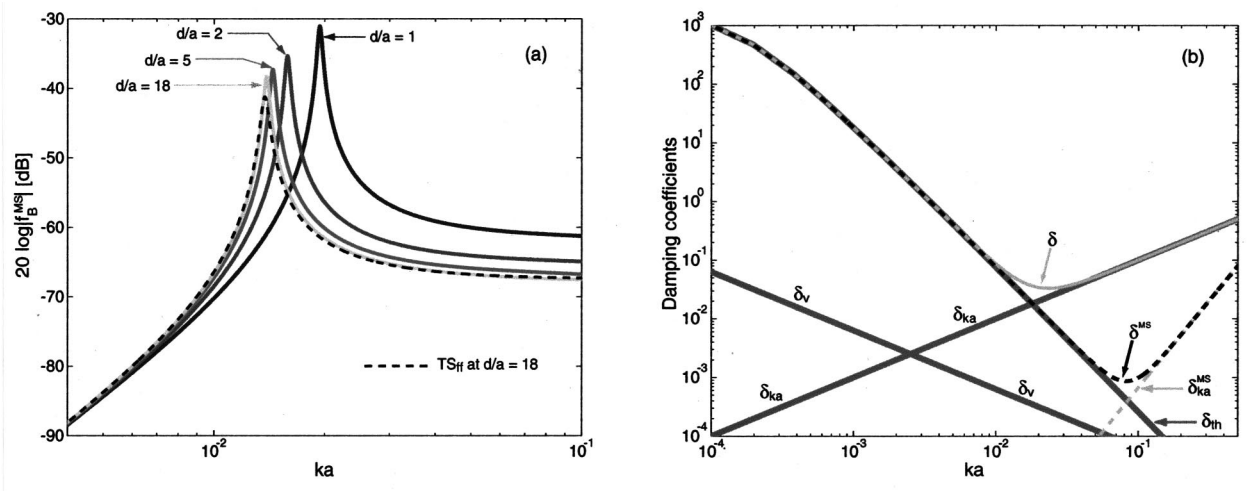


FIG. 5. The effect of multiple scattering between a single 425 μm bubble and a pressure release interface. (a) Plot of $20\log|f_B^{\mathrm{MS}}|$ for variable frequency at four different bubble depths. Note the change in the resonance frequency and scattering amplitude. (b) Plot of the damping coefficients of the bubble in free field (δ_{ka}, δ_{th}, and δ_v), along with their sum (δ) and the modified radiation and total damping coefficients ($\delta_{ka}^{\mathrm{MS}}$ and δ^{MS}, respectively) showing the contribution of multiple scattering to the damping of the bubble.

4(b), which shows that the higher-order scattering terms play a substantial role for $kd<1$ for most of the frequency range. There are two exceptions to this: (1) well below the bubble's free-field resonance frequency ($ka \ll 0.0136$), where multiple scattering has no effect, and (2) around the bubble's free-field resonance frequency, where multiple scattering affects the response beyond $kd=3$ [or $\log(kd)=0.48$]. Note that for $kd>3.5$ [or $\log(kd)>0.54$] there is no significant contribution from multiple scattering.

It is evident that as the bubble gets very close to the surface, multiple scattering will play a role. Equation (8) predicts an enhancement of scattering amplitude and a shift of the resonance frequency [features that stand out in Figs. 4(a) and (b)]. To investigate this further, we define the quantity $f_B^{\mathrm{MS}}=f_B\cdot\mathrm{MSC}$, which includes the influence of the multiple waves reflected from the air–water interface on the scattering function of the bubble and can be expressed as

$$f_B^{\mathrm{MS}}=\frac{a}{\left(\frac{f_{\mathrm{res}}}{f}\right)^2-1-i\delta-R\frac{a}{2d}e^{ik2d}}. \tag{11}$$

In Eq. (11), the real part of the term $R(a/2d)e^{ik2d}$ is the change in resonance frequency, while the imaginary part is the change in the bubble's damping. This was also recognized by Ye and Feuillade [see Eqs. (26)–(29) in Ref. 6] and is identical to the effect of multiple scattering from two monopole bubbles oscillating out of phase.[15] As an example, the modified coefficient of total damping, δ^{MS}, equals

$$\delta^{\mathrm{MS}}=\delta+R\frac{a}{2d}\sin(2kd). \tag{12}$$

For $R=-1$, Eq. (12) is identical to Eq. (29) from Ref. 6. Therefore, the net effect of the multiple interactions of the scattered waves with the free surface is the reduction of the radiation damping, δ_{ka}, which represents the dissipation of energy by the oscillating bubble in the form of reradiated spherical waves. An expansion (to third order) of the sine term results in an approximate expression of the modified radiation damping, $\delta_{ka}^{\mathrm{MS}}$, as $(2/3)(ka)(kd)^2$.

In Fig. 5(a) we plot $20\log|f_B^{\mathrm{MS}}|$ (in dB) for a 425 μm bubble and variable frequency, for four different bubble depths (as noted on the graph). For reference, the scattering response of a 425 μm bubble in the free field (i.e., $20\log|f_B|$) is also shown as a dashed line. In Fig. 5(b) we plot the three damping components of a 425 μm bubble in the free field (identified as δ_{ka}, δ_{th}, and δ_v), and their sum, δ, along with $\delta_{ka}^{\mathrm{MS}}$ and δ^{MS}, which is equal to the sum of $\delta_{ka}^{\mathrm{MS}}$, δ_{th}, and δ_v.

For frequencies well below the resonance frequency, scattering from the bubble is very weak, and thus, too, is interaction with the surface. Very little energy is dissipated through reradiation, and δ_{th} (or for bubbles of radius less than 10 μm the combination of δ_{th} and δ_v) dominates. For example, for a 425 μm bubble insonified by a 2790 Hz wave (i.e., $ka=0.005$), the free-field target strength equals -83.7 dB [see Fig. 5(a)], and δ_{th} is two orders of magnitude larger than δ_{ka} and δ_v [see Fig. 5(b)]. In such a case, the reduction of radiation damping has a negligible effect, and thus the response of a bubble close to the air–water interface can be modeled using the four paths only [i.e., the p_T^0 given by Eq. (6)] with little error. This is seen in Fig. 4(b) (in what appears as a uniformly colored strip equivalent to $20\log|\mathrm{MSC}|\approx 0$ dB for all kd and for $ka \ll 0.0136$) and in Fig. 6(a), which shows backscattering from a 425 μm bubble calculated using Eq. (6) (gray solid line) and Eq. (8) (black solid line) vs frequency for d equal to $1a$, $5a$, $18a$, and $100a$. The response of a single 425 μm bubble at depth $d=100a$ is also shown as a dashed black line. The vertical thin dashed line marks the resonance frequency, $f_{\mathrm{res}}=7663$ Hz ($ka=0.0137$), of a 425 μm bubble at $d=100a$. The vertical dashed–dotted line marks the insonifying frequency $f=44\,638$ Hz ($ka=0.08$). The geometry used to perform the calculations is exactly the same as that used to produce Fig. 4.

For a bubble of radius larger than 10 μm in the free field and excited in the vicinity of its resonance frequency, its response is affected mostly by δ_{th} and δ_{ka}. The effect of

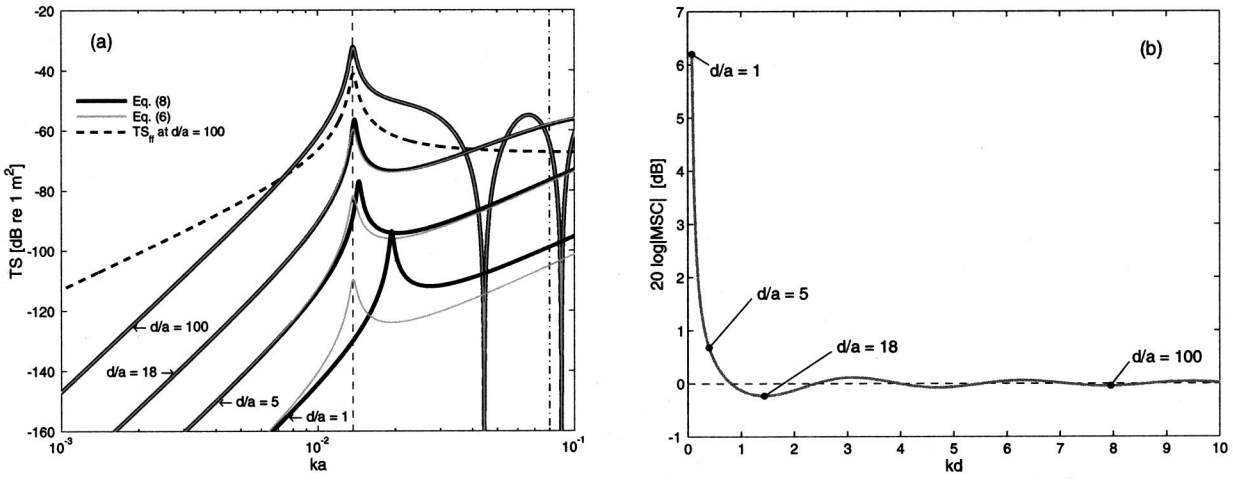


FIG. 6. Numerical results for backscattering from a 425 μm bubble close to a flat, pressure release surface. (a) Calculations performed using Eq. (8) (black solid line) and Eq. (6) (gray solid line) are plotted for four different d/a (as noted on the graph) and variable frequency. (b) The contribution of multiple scattering for a fixed frequency (44 638 Hz, or $ka=0.08$) vs d. The slant range between the source/receiver and the bubble is 141 m, and the grazing angle is 45°.

multiple scattering reduces δ_{ka} to $\delta_{ka}^{\mathrm{MS}}$, which allows thermal damping to play an increased role around the resonance [see Fig. 5(b)]. The decreased overall damping dissipates less energy, which manifests itself as an enhancement of the scattering amplitude [as seen in Figs. 5(a) and 6(a)]. For a comprehensive discussion on the effect of multiple scattering on the resonance frequency see Ref. 6.

Finally, for $ka \gg 0.0136$ the radiation damping dominates, and thus any change to it due to the effect of multiple scattering will have a direct result in the scattering amplitude. In Fig. 6(b) we plot $20\log|\mathrm{MSC}|$ vs d (expressed in terms of kd) for an insonifying frequency of 44 638 Hz ($ka=0.08$). The values for $d/a=1$ ($kd=0.22$), $d/a=5$ ($kd=0.4$), $d/a=18$ ($kd=1.44$), and $d/a=100$ ($kd=8$) are shown as dots. These show the exact difference between the response calculated using Eq. (8) and Eq. (6), shown in Fig. 6(a). The largest difference between the two solutions is at $d=a$ (a gain of 6.2 dB). This drops rapidly to a minimum (a loss of 0.23 dB) at $d=18a$; beyond this value the scattering amplitude exhibits a diminishing, small amplitude oscillation, going as $a\sin(2kd)/(2d)$, about the coherent sum of the four paths. Thus at this point, the four paths alone are sufficient to describe scattering by a bubble close to a flat, pressure release interface. Note that the change in resonance frequency and amplitude due to multiple scattering will be exactly the same for either a monostatic or a bistatic configuration.

III. EXPERIMENT

The experimental apparatus and procedure were presented in detail in an earlier publication (see Ref. 13); thus we provide only a brief description in this paper. The experiments were performed in a Plexiglas tank filled with fresh water, utilizing either two immersion transducers symmetrically arranged in bistatic configuration (shown in Fig. 7) or a single transducer in a monostatic configuration [by employing a transmit/receive (T/R) switch]. In both cases the transducers were rotated by an angle θ from the horizontal axis (see Fig. 8). The source was driven in pulse mode using a five-cycle tone burst with a center frequency of 120 kHz. The signal received was amplified (a gain of 40 dB), bandpass filtered (not shown in Fig. 8), and digitized. For each measurement 30 waveforms were averaged coherently, and the root-mean-squared voltage of the bubble signal was measured and squared. A free-field calibration of the system was performed using an 18.2 mm radius tungsten–carbide sphere as a reference target.

Bubbles were created using a method similar to the one described in Ref. 13, which produced bubbles of radius (425 ± 40) μm. This is our best estimate for the size of the single bubble measured with the calibrated acoustic system. The equivalent mean of a population of single bubbles independently measured using a microscope was 420 μm, with a standard deviation of 27 μm. For our experiments using single bubbles of radius 425 μm insonified by 120 kHz sound, ka equals 0.215, which exceeds the upper limit of the ka range for which Eq. (2) is typically used. (The difference between the TS calculated using the low ka approximation [i.e., Eq. (2)] and the one calculated using Anderson's general solution is small, about ≈ 0.25 dB.) In this paper however, we are motivated to use Eq. (2) because the monopole approximation allows for the derivation of very simple expressions (for both the monostatic and bistatic case) that embody all the physics of the problem of scattering from a bubble close to an air–water interface, including multiple scattering effects.

The bubbles were placed on a fine nylon thread that was attached to a thin wire frame. The thread was approximately 100 μm in diameter. When in bistatic mode the noise floor was equivalent to a scatterer with $\mathrm{TS}\approx -84$ dB, and the thread alone did not return a signal above that. The frame was fastened to a three-axis translation stage, allowing accurate control of the vertical movement of the frame. The primary variable was the depth of the bubble, d, and data were acquired for $d=1a$ to $d=100a$ in 1 mm increments.

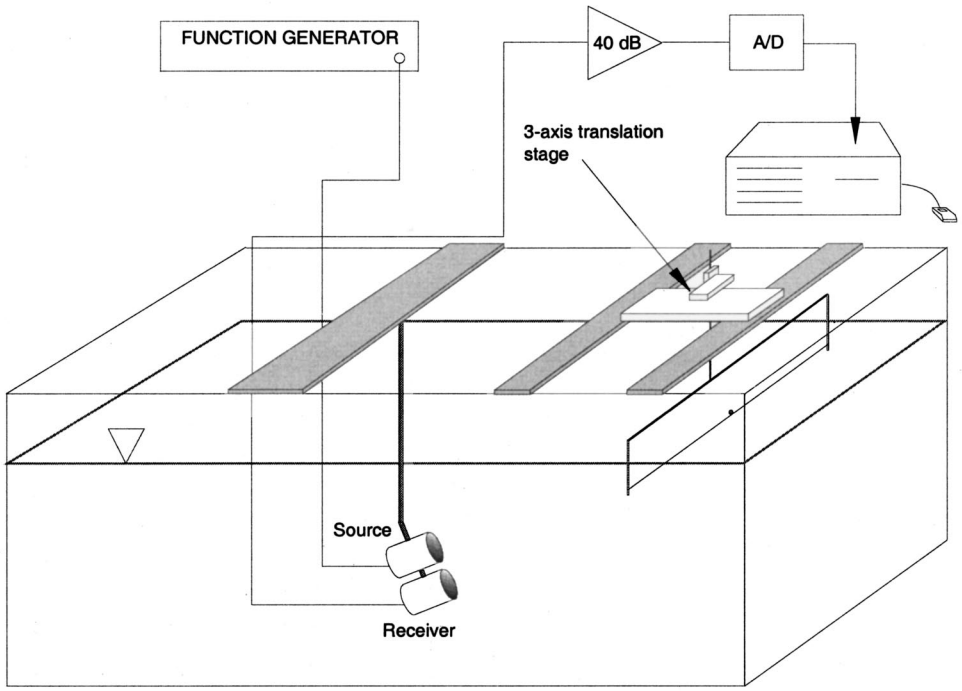


FIG. 7. Experimental apparatus used for the bistatic configuration. For the monostatic configuration one of the two transducers was used in conjunction with a T/R switch. The drawing is not to scale.

For both transducer arrangements, the two-way directivity pattern was obtained from measurements of our standard calibration sphere (away from the boundary) and fitting the results to a Gaussian beam profile. The modeled beam was used to correct the data that were off the axis of the beam. (The two-way -3 dB beamwidth for an individual transducer was 9.9°.)

For each of the two transducer configurations (i.e., bistatic and monostatic), two angles were investigated experimentally: $\theta \approx 18°$ and $\theta \approx 28°$. Note that in the monostatic cases the angle θ becomes the grazing angle. Four experiments were performed for each of the four combinations of transducer configuration and angle θ. The mean of these measurements will be compared with the model results in the next section. Each experiment lasted approximately 15 minutes, and we did not observe significant reduction of TS over this period due to bubble dissolution.

IV. COMPARISON OF THE MODEL WITH THE DATA AND DISCUSSION

Comparison of the model with the data (Fig. 9) verifies the validity of the model. In Fig. 9 the theoretical curves (solid lines) and experimental data (dots) are plotted for a fixed frequency of 120 kHz and variable d (recall that $d = |Z_B|$). The variable d is expressed in terms of d/a, which varies from 1 to 100, and kd, which varies from 0.22 to 21.5. Modeling of the experiments was achieved using Eq. (8) to calculate the total pressure at the receiver. Equation (9) was then used to express the calculated pressure in terms of target strength. The variables used to simulate each of the experiments are shown in Table I. Figures 9(a) and (c) compare theory and data for the bistatic configuration, while Figs. 9(b) and (d) compare theory and data for the monostatic configuration. The black line represents the response of a single bubble of radius 425 μm insonified by continuous waves (CW). The gray line represents its response to a five-cycle tone burst; this line also accounts for the noise floor, since the best estimate of the background noise (shown as the vertical dashed–dotted line) has been added to the response of the bubble. For reference, the target strength of a single 425 μm bubble in the free field (vertical dotted line) is also plotted on each of the four graphs. The dots are the ensemble linear intensity average of the four experiments at

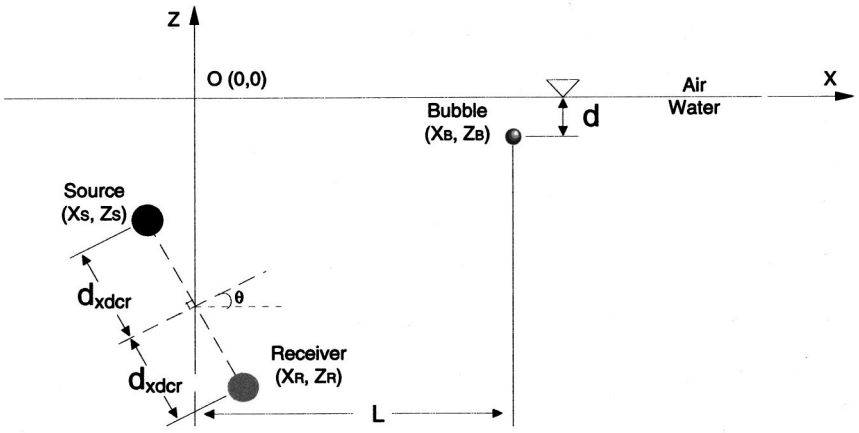


FIG. 8. Geometry describing the laboratory experiment (not to scale).

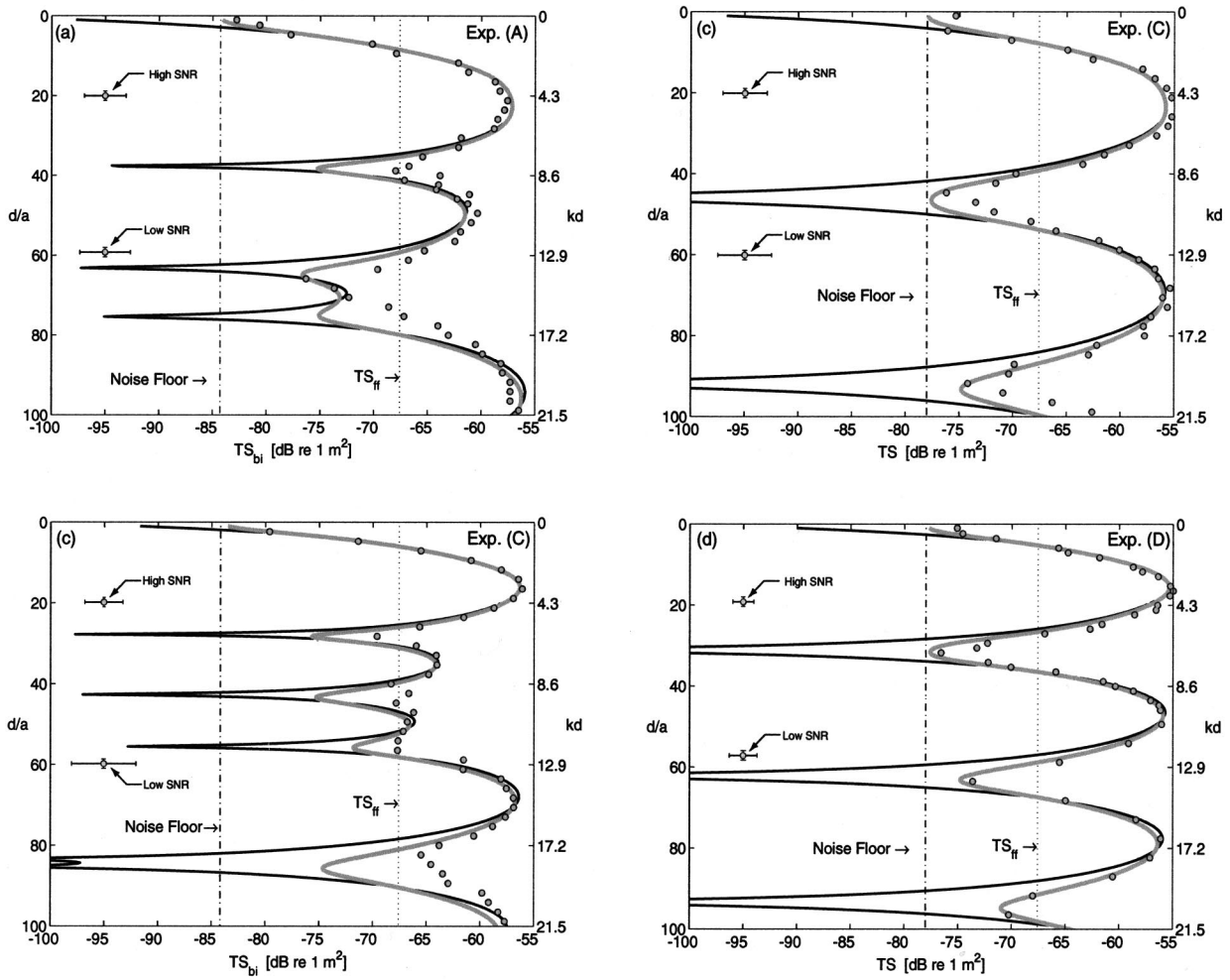


FIG. 9. Theoretical curves for scattering from a single bubble of radius 425 μm close to a flat, pressure release interface at 120 kHz compared with experimental data (dots). The black solid line is the response of the bubble due to CW excitation, and the gray solid line represents its response to a five-cycle tone burst. Note that the gray line also accounts for the presence of the noise floor (the best estimate of which is shown as the vertical dashed–dotted line). The configuration of the system for each of the experiments is detailed in Table I.

each d. The error bars shown on each graph are the uncertainties in the measurements; the uncertainty in d, which was estimated as 0.5 mm, is represented by the vertical error bars, and the uncertainties in the acoustic measurements at d are represented by the horizontal error bars. Depending on the signal-to-noise ratio (SNR), defined by the difference between the apparent bubble TS and the nominal background TS, we establish two uncertainties for the acoustic measurements; one for the high SNR, say at least 20 dB above the noise floor, and one for the low SNR [as shown in Figs. 9(a)–(d)]. For a comprehensive error analysis see Appendix D in Ref. 16. The experimental data are in good agreement with the simulations. For the bistatic experiments (A) and (C) data deviate from theory for $d/a>80$, due to a systematic error caused by the narrow two-way beam of the system, while for the monostatic experiments (B) and (D) the agreement between theory and data is good throughout the d/a range.

TABLE I. Pertinent variables used to simulate each of the four experiments.

Experiment	a (μm)	X_S (m)	Z_S (m)	X_R (m)	Z_R (m)	X_B (m)	$\approx\theta$
(A)	425	−0.016	−0.149	+0.016	−0.251	0.61	18°
(B)	425	0.0	−0.205	0.0	−0.205	0.61	18°
(C)	425	−0.035	−0.246	+0.035	−0.374	0.64	28°
(D)	425	0.0	−0.330	0.0	−0.330	0.63	28°

For bistatic scattering, experiments (A) and (C), the phase difference between the four paths depends on a combination of grazing angle, transducer separation, range from the bubble, and insonifying frequency, which result in the patterns shown in Figs. 9(a) and (c). A simpler expression for the bistatic case can be obtained by writing Eq. (6) as

$$p_T^0 = f_B \tilde{A}_{\text{inc}} \tilde{A}_{\text{scat}}, \tag{13}$$

where

$$\tilde{A}_{\text{inc}} = \frac{e^{ikr_{SB}}}{r_{SB}} - \frac{e^{ik(r_{SI}+r_{IB})}}{r_{SI}+r_{IB}} \tag{14}$$

and

$$\tilde{A}_{\text{scat}} = \frac{e^{ikr_{BR}}}{r_{BR}} - \frac{e^{ik(r_{BI}+r_{IR})}}{r_{BI}+r_{IR}}. \tag{15}$$

Equation (14) combines the propagation terms of the two waves that constitute the incident field, while Eq. (15) does

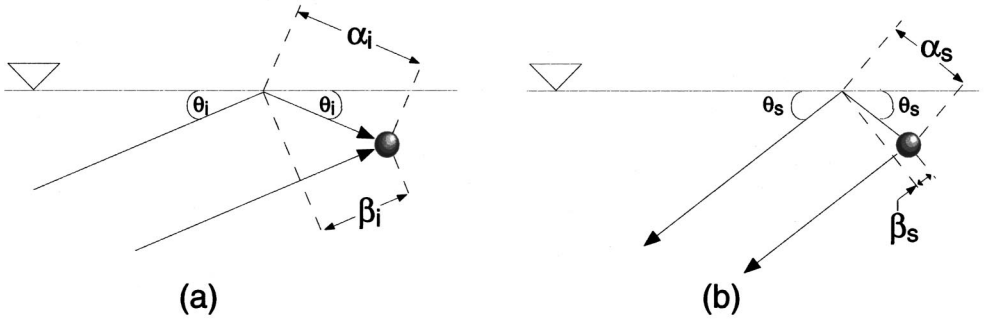


FIG. 10. Geometry of the (a) incident and (b) scattered paths for the long-range limit and bistatic configuration.

this for the scattered field. In the long-range limit, Eq. (13) can be written as

$$p_T^0|_{d/r_i\ll 1,d/r_s\ll 1}\approx f_B\left[\left(\frac{e^{ikr_i}}{r_i}\right)\left(\frac{e^{ik\beta_i}}{1+\frac{\beta_i}{r_i}}-\frac{e^{ik\alpha_i}}{1+\frac{\alpha_i}{r_i}}\right)\left(\frac{e^{ikr_s}}{r_s}\right)\times\left(\frac{e^{ik\beta_S}}{1+\frac{\beta_s}{r_s}}-\frac{e^{ik\alpha_s}}{1+\frac{\alpha_s}{r_s}}\right)\right], \tag{16}$$

where $r_i=r_{SI}$ and $r_s=r_{IR}$ are shown in Fig. 2, and the variables α_i, β_i, α_s, and β_s are shown in Fig. 10. An asymptotic expansion (to first order in α_i/r_i, β_i/r_i, α_s/r_s, and β_s/r_s) of Eq. (16) gives

$$p_T^0|_{d/r_i\ll 1,d/r_s\ll 1}\approx 4\frac{e^{ikr_i}}{r_i}f_B\frac{e^{ikr_s}}{r_s}\sin\left(kd\frac{d_{src}}{r_i}\right)\sin\left(kd\frac{d_{rcv}}{r_s}\right), \tag{17}$$

where $d_{src}=|Z_S|$ and $d_{rcv}=|Z_R|$ [see Fig. 2(I)]. Therefore, the response of the bubble in the bistatic case is governed by two sine functions with the same amplitude and a different period. The superposition of the two sinusoids results in the seemingly irregular pattern seen in Figs. 9(a) and (c).

If we furthermore assume that $r_i\approx r_s$, and set them both equal to their average [that is, define $r=(r_i+r_s)/2$], Eq. (17) reduces to

$$p_T^0|_{d/r\ll 1}\approx 2\frac{e^{ikr}}{r}f_B\frac{e^{ikr}}{r}\times\left[\cos\left(kD\frac{d}{r}\cos\theta\right)-\cos(2kd\sin\theta)\right], \tag{18}$$

where $D=2d_{xdcr}$ and θ is the angle of rotation of the transducers (see Fig. 8). Note that Eq. (7) can be incorporated into both Eqs. (17) and (18) to account for the effects due to multiple scattering between the bubble and the interface. Written in terms of bistatic TS [via Eq. (9)], for example, the substitution of Eqs. (18) into Eq. (8) yields

$$\mathrm{TS}_{bi}|_{d/r\ll 1}\approx \mathrm{TS}_{ff}+6+20\log\left|\cos\left(kD\frac{d}{r}\cos\theta\right)-\cos(2kd\sin\theta)\right|-20\log\left|1+\frac{e^{ik2d}}{2d}f_B\right|, \tag{19}$$

where TS_{ff} is given by Eq. (10).

In Fig. 11 we show the scattering response of a 425 μm bubble at a distance d from a smooth air–water interface using Eq. (6) (solid dark-gray line), Eq. (17) (solid light-gray line), and Eq. (18) (black dashed line). All three equations were expressed in terms of bistatic TS using Eq. (9). Figure 11(a) shows the results for the laboratory coordinates of experiment (A) (that is, $[X_S,Z_S]=[-0.016\,\mathrm{m},-0.149\,\mathrm{m}]$, $[X_R,Z_R]=[0.016\,\mathrm{m},-0.251\,\mathrm{m}]$, and $[X_B,Z_B]=[0.61\,\mathrm{m},$

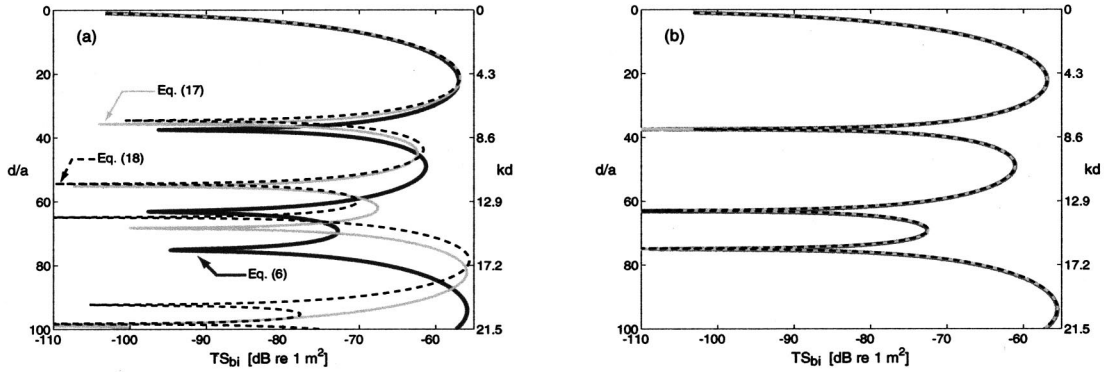


FIG. 11. Comparison of the exact solution, provided by Eq. (6) (solid dark-gray line), and the approximate solutions given by Eq. (17) (solid light-gray line) and Eq. (18) (dashed black line) for scattering from a 425 μm bubble close to a pressure release interface. (a) Comparison of the three solutions for the actual geometry of experiment (A) (see Table I). (b) Comparison of the three solutions for $d/r\ll 1$ and $r_i\approx r_s$ (see text for details). In (b) the three curves are indistinguishable.

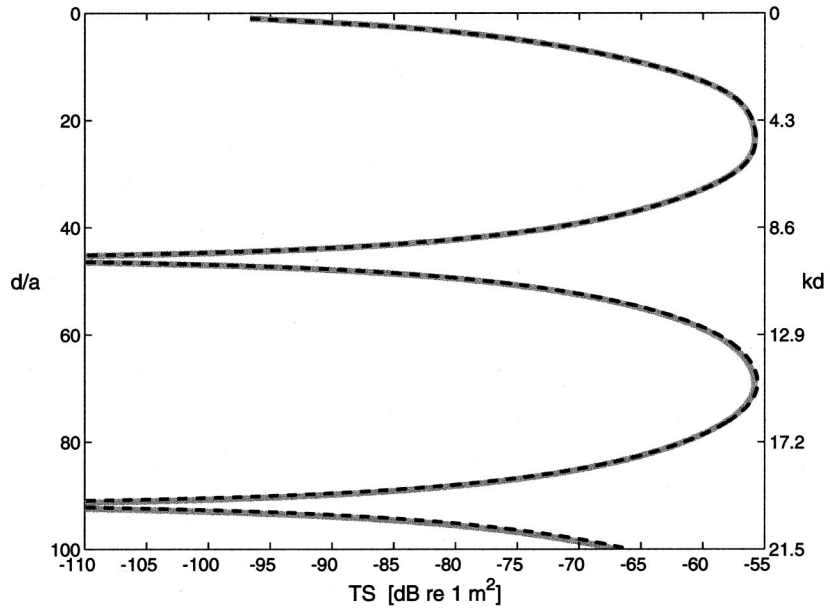


FIG. 12. Comparison of the exact solution (solid gray line) provided by Eqs. (8) and (9) for the monostatic case and the $d/r \ll 1$ approximation (black dashed line) given by Eq. (21). The two curves simulate scattering for the actual geometry of experiment (B) (see Table I).

$-d$ m]), and Fig. 11(b) shows the results for $[X_S, Z_S] = [-1.6\,\text{m}, -14.9\,\text{m}]$, $[X_R, Z_R] = [1.6\,\text{m}, -25.1\,\text{m}]$, and $[X_B, Z_B] = [61\,\text{m}, -d\,\text{m}]$. It is evident that when both the source and the receiver are much further from the bubble than d, the three solutions are in very good agreement. In Fig. 11(a) the conditions $d/r_i \ll 1$ and $d/r_s \ll 1$ are not satisfied, and thus there is a significant disagreement between the curve calculated using Eq. (17) and the one calculated using the exact solution. Furthermore, the condition $r_i \approx r_s$ is valid only for the low d/a range that was used for the calculation. This becomes evident from the increasing deviation of the two approximate curves for increasing d. Therefore, it was necessary to model experiments (A) and (C) using the exact solution.

For experiments (B) and (D) (i.e., monostatic scattering) a simplified expression can be obtained by setting $D=0$ in Eq. (18), which leads to

$$p_T^0|_{d/r \ll 1, D=0} \approx 4 \frac{e^{ikr}}{r} f_B \frac{e^{ikr}}{r} [\sin^2(kd \sin\theta)], \tag{20}$$

where now θ is the grazing angle. As in the bistatic case, the MSC can be readily incorporated in Eq. (20) to account for multiple scattering effects. In terms of target strength, for example, backscattering that includes MSC is

$$\text{TS}|_{d/r \ll 1} \approx \text{TS}_{ff} + 12 + 40 \log|\sin(kd \sin\theta)| - 20 \log\left|1 + \frac{e^{ik2d}}{2d} f_B\right|. \tag{21}$$

Equation (21) indicates that the target strength of a single bubble near a pressure release interface exhibits oscillatory behavior with nulls every $kd \sin\theta = n\pi$ (where $n = 0,1,2,3,...$), and has a maximum that is 12 dB above the target strength of a single bubble in the free field.

The good agreement between the exact solution given by Eqs. (8) and (9) and the approximation for monostatic configuration given by Eq. (21) can be seen in Fig. 12, in which we plot backscattering from a 425 μm bubble for the geometry of experiment (B) (see Table I). The gray solid line is the exact solution, and the dashed black line is the approximation for $d/r \ll 1$. Figure 12 clearly shows that the two solutions are in very good agreement. This implies that experiments (B) and (D) could have been modeled with minimal computational effort by using Eq. (21).

It is of interest to revisit the idea presented in Sec. II of the effect of multiple scattering on the scattering response of a bubble close to an air–water interface and for $ka \gg 0.0136$ (see last paragraph of Sec. II and Fig. 6). Figure 13 shows the contribution of multiple scattering to the coherent addition of the four paths for experiment (B), a case for which $ka \gg 0.0136$. The gray solid line was calculated using Eq. (21), while the black solid line was calculated utilizing Eqs. (20) and (9). It is evident that the gain due to multiple scattering is maximum (+5.5 dB) at $d/a = 1$ ($kd = 0.215$), which is marked with a dotted line. As the depth of the bubble increases the effect of multiple scattering decreases

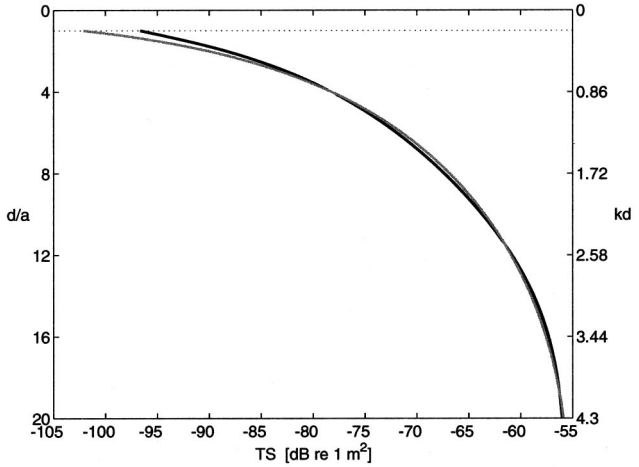


FIG. 13. Comparison of the coherent addition of the four paths (solid black line) and the prediction of Eq. (21) (solid gray line), which includes the contribution of multiple scattering. The two curves simulate scattering for the actual geometry of experiment (B) (see Table I).

and it reaches a minimum (a loss of -0.56 dB) at $d/a=7$ ($kd=1.5$). Beyond that minimum the scattering amplitude exhibits a diminishing, small amplitude oscillation about the coherent addition of the four paths.

V. SUMMARY

The problem of monostatic and bistatic scattering of sound from a single bubble of radius a that is located at a distance d from a flat, air–water interface was studied theoretically and experimentally. A model was developed utilizing a ray-synthesis approach to describe the four scattering paths, from source to bubble to receiver, that determine the response of the bubble. Multiple scattering effects were accounted for using an expression derived from the multiple scattering series. Simulations showed that in the long-range limit and for a monostatic configuration (i.e., a source and receiver that are co-located), calculations performed using our model are in agreement with those computed using an alternative model developed by Ye and Feuillade.[6] It was also shown that, for most of the ka range, multiple scattering plays a significant role only for $kd<1$, where k is the wave number of the water medium. However, for $ka\ll 0.0136$ multiple scattering has practically no effect. Around the bubble's free-field resonance frequency ($ka\approx 0.0136$) multiple scattering effects persist beyond $kd=3$.

Well-controlled experiments were performed by placing a single bubble of radius $a\approx 425\,\mu$m on a fine thread at a distance, d, from the air–water interface. The bubble was excited by five-cycle tone bursts with a center frequency of 120 kHz. The experiments utilized both a monostatic and a bistatic system. Theory and experiment were in good agreement, thus validating the model and verifying the effect of the four paths in the response of the bubble.

Simplified expressions were derived for scattering from a near-surface bubble for both monostatic and bistatic configurations, and in the long range limit. These expressions also included the effects due to multiple scattering between the bubble and the interface.

ACKNOWLEDGMENTS

This work was funded by the Office of Naval Research, Code 321 OA.

[1] P. H. Dahl, W. J. Plant, B. Nützel, A. Schmidt, H. Herwig, and E. A. Terray, ''Simultaneous acoustic and microwave backscattering from the sea surface,'' J. Acoust. Soc. Am. **101**, 2583–2595 (1997).
[2] P. A. Crowther, ''Acoustical scattering from near-surface bubble layers,'' in *Cavitation and Inhomogeneities in Underwater Acoustics*, edited by W. Lauterborn (Springer-Verlag, 1980), pp. 194–204.
[3] K. Sarkar and A. Prosperetti, ''Coherent and incoherent scattering by oceanic bubbles,'' J. Acoust. Soc. Am. **96**, 332–341 (1994).
[4] L. L. Foldy, ''The multiple scattering of waves. I. General theory of isotropic scattering by randomly distributed scatterers,'' Phys. Rev. **67**, 107–119 (1945).
[5] S. T. McDaniel, ''Sea surface reverberation: A review,'' J. Acoust. Soc. Am. **94**, 1905–1922 (1993).
[6] Z. Ye and C. Feuillade, ''Sound scattering by an air bubble near a plane sea surface,'' J. Acoust. Soc. Am. **102**, 798–805 (1997).
[7] M. Strasberg, ''The pulsation frequency of nonspherical gas bubbles in liquids,'' J. Acoust. Soc. Am. **25**, 536–537 (1953).
[8] E. Meyer and E. Skudrzyk, ''Sound absorption by gas bubbles,'' in *Sound Absorption and Sound Absorbers in Water*, Vol. 1, Chap. II, Department of the Navy, 1950.
[9] G. Gaunaurd and H. Huang, ''Acoustic scattering by an air-bubble near the sea surface,'' IEEE J. Ocean Eng. **20**, 285–292 (1995).
[10] M. Strasberg, ''Comments on 'Acoustic scattering by an air bubble near the sea surface,''' IEEE J. Ocean Eng. **21**, 233 (1996).
[11] C. Devin, Jr., ''Survey of thermal, radiation, and viscous damping of pulsating air bubbles in water,'' J. Acoust. Soc. Am. **31**, 1654–1667 (1959).
[12] V. C. Anderson, ''Sound scattering from a fluid sphere,'' J. Acoust. Soc. Am. **22**, 426–431 (1950).
[13] G. Kapodistrias and P. H. Dahl, ''Effects of interaction between two bubble scatterers,'' J. Acoust. Soc. Am. **107**, 3006–3017 (2000).
[14] V. Twersky, ''Multiple scattering of waves and optical phenomena,'' J. Opt. Soc. Am. **52**, 145–171 (1962).
[15] A. Sangani, ''A pairwise interaction theory for determining the linear acoustic properties of the dilute bubble liquids,'' J. Fluid Mech. **232**, 221–284 (1991).
[16] G. Kapodistrias, Ph.D. thesis, University of Washington, 2001.

Crack characterization using guided circumferential waves

Christine Valle
Department of Mechanical Engineering, University of Maine, Orono, Maine 04469-5711

Marc Niethammer
School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0355

Jianmin Qu
G. W. W. School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0405

Laurence J. Jacobs
School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0355

(Received 1 August 2000; revised 15 March 2001; accepted 14 May 2001)

This paper examines the propagation of guided circumferential waves in a hollow isotropic cylinder that contains a crack, with the goal of using these guided waves to both locate and size the crack. The crack is sized using a modified Auld's formula, which relates the crack's length to a reflected energy coefficient. The crack is then located by operating on the backscattered signal with a time-frequency digital signal processing (DSP) technique, and then comparing these results to those obtained if the cylinder is perfect. The guided circumferential waves are generated with a commercial finite element method (FEM) code. One objective of this work is to demonstrate the effectiveness of using sophisticated DSP techniques to describe the effect of scattering on dispersive waves, showing it is possible to characterize cracks systematically and accurately by quantifying this scattering effect. The results show that the need for high frequency signals to detect small cracks is significantly decreased by using these techniques. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1385899]

PACS numbers: 43.20.Fn, 43.20.Mv [DEC]

I. INTRODUCTION

It is well known that fatigue cracks can initiate and grow in the radial direction of an annular structure that is subjected to a large number of fatigue cycles. These annular structures are used extensively in a variety of industrial applications, such as the aerospace, oil and nuclear industries. A quantitative and systematic inspection methodology is needed to detect and characterize these cracks, before a catastrophic structural failure occurs.

Ultrasonic testing is a candidate nondestructive evaluation (NDE) methodology for this application. Unfortunately, traditional ultrasonic techniques (such as pulse-echo) are ineffective in this application because of problems associated with curvature and accessibility difficulties. Guided ultrasonic waves, i.e., waves that propagate in the direction of the layer while behaving as standing waves through the thickness of the layer, are a potential alternative. Nagy *et al.*[1] recently proposed using guided ultrasonic waves that propagate in the circumferential direction to detect radial cracks in weep holes of airframes.

The main advantage in using guided ultrasonic waves is that they can interrogate the entire specimen, including inaccessible portions. Unfortunately, the detected ultrasonic signals are very complicated, causing difficulties in signal interpretation. Previous researchers have studied guided waves in plates[2] and cylinders.[3] For example, Alleyne and Cawley[3] examine cylindrical waves that propagate down the axis of the cylinder *but remain standing in its circumferential direction* to detect cracks in long thin pipes. [That is, the expressions for the field quantities contain terms such as $\cos(n\theta)$ or $\sin(n\theta)$, where θ is the usual polar angle (see Fig. 1) and n is an integer. Thus quantities such as velocities or stresses are modulo $[2\pi]$ periodic.] However, these longitudinal waves are not well suited for determining the *radial* location of a crack in a short annular structure. This is especially true for a cylindrical component whose diameter is on the same order of magnitude as its length. Circumferential waves, that is guided waves *that propagate in the circumferential direction of the cylinder*, i.e., in the θ direction (see Fig. 1), are proposed for these applications. Such waves, contrary to the previous case, contain terms such as $e^{ik\theta}$, where θ has the same meaning as previously and k is a **real** number (and not an integer). Thus quantities such as velocities and stresses are not periodic modulo $[2\pi]$. These circumferential waves do not propagate in the z direction (in the axis of the cylinder or along its length). Guided circumferential waves have been studied in Refs. 4 and 5 and were first introduced by Viktorov[6,7] as the natural extension of Lamb waves from a flat plate to a curved plate. Liu and Qu[8] consider the time harmonic analysis of a (perfect) hollow cylinder, while Valle *et al.*[4] examine the time harmonic behavior of a double-layered, hollow cylinder. None of these studies consider transient circumferential waves propagating in a cracked cylindrical component.

This paper examines the propagation of (transient) guided circumferential waves in a metallic (steel) hollow

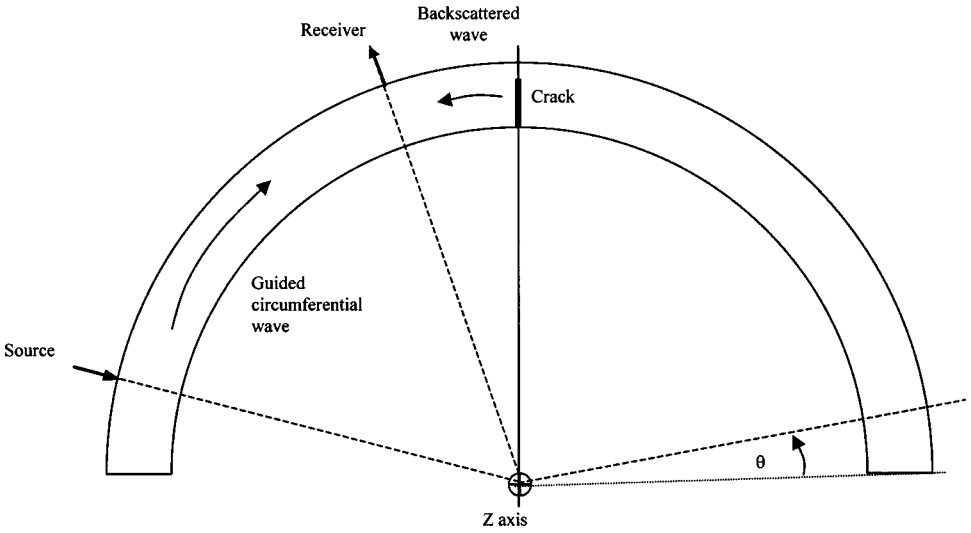


FIG. 1. Specimen geometry, taking advantage of symmetry of hollow cylinder, and simulating laser interrogation with a normal point load/point receiver system for backscattered energy detection.

cylinder that contains a crack, using these guided waves to both locate and size the crack. The crack is sized using a scattering formula first developed by Auld,[9] which is modified in the current work to analyze transient (non-time-harmonic) signals. The crack is then located by operating on the backscattered signal with a time-frequency digital signal processing (DSP) technique, and then comparing these results to those obtained from a perfect cylinder—one without a crack. All the time signals (guided circumferential waves) presented in this work are generated using a commercial FEM code (ABAQUS/Explicit)—except for accuracy checks done by comparing experimental and numerical results. While much fundamental work has been done in the past to develop numerically efficient methods to calculate the transient response of waveguides (both perfect and flawed),[10,11] the FEM was chosen here for its robustness, accuracy and convenience.

It is important to note that no effort is made in this work to selectively restrict the number of guided modes propagating in the cylinder.

One objective of this work is to demonstrate the effectiveness of using sophisticated DSP techniques to describe *guided, broadband, and multi-mode* circumferential waves in imperfect cylinders. Specifically, these DSP techniques are used to examine the effect of scattering on dispersive waves, showing it is possible to characterize cracks systematically and accurately by quantifying this scattering effect.

In this text, the term *wave* designates any ultrasonic signal that propagates into the annular structure under investigation. The term *mode* designates an ultrasonic signal fully represented in a time harmonic fashion by a single curve on the dispersion curves of the structure where the mode propagates (such as, the familiar Lamb modes a_0 and s_0). Therefore, a mode is always a wave, but a wave can consist of multiple modes (since, as an example, the a_0 and s_0 modes can co-exist for a given frequency range).

II. FEM MODEL OF THE SCATTERING OF GUIDED CIRCUMFERENTIAL WAVES CAUSED BY A CRACK

In the past, the computational cost of modeling high-frequency wave propagation problems was prohibitively high. It has been shown recently that the FEM is capable of modeling such problems accurately and efficiently. This computer power increase makes the FEM an excellent method for solving wave propagation problems that are intractable by analytical methods.[12,13]

This numerical study only models a portion (half) of the cylinder, because the signal processing techniques only make use of the backscattered data and thus the whole circumference need not be modeled. The backscattered data will reach the receiver before the transmitted data, and therefore the backscattered data can be windowed out. Also, FE simulations show that most of the signal attenuates by the time it has traveled 360°. Therefore, there is no need to model the complete cylinder. As such this geometry may be thought of as an infinitely long curved plate. ABAQUS/Explicit is used for all numerical calculations. Convergence is reached when the solution obtained is mesh-invariant (the solution remains the same when the mesh size is decreased). As a further check, solutions for the perfect cylinder are compared to the normal modal expansion (NME) calculations of Liu and Qu[14] (see Fig. 2), and to experimentally generated signals (using laser ultrasonics),[5] shown in Fig. 3. The excellent agreement between these solutions confirms the accuracy of these FEM models. A typical half cylinder mesh (the dimensions of the cylinder used in this research are shown in Table I) has 4000 elements (for loads of 0.5 MHz center frequency). This number of elements is obtained by successive refinement of the mesh until mesh invariance is obtained. The element size is 15 times smaller than the wavelength of the (nondispersive) shear wave at the center point of the bandwidth (0.5 MHz) and nearly 7 times smaller at the upper end of the bandwidth (1.5 MHz). While these ratios are low at the higher frequencies, mesh invariance and the numerical validation with Fortran show that they are sufficient for ABAQUS/Explicit.

Each element is a 4-node plain strain continuum element (CPE4R). Such an element provides a second-order interpolation, with reduced integration and hourglass control (hourglassing is a numerical phenomena by which a zero-energy mode propagates through and spoils the solution—see the ABAQUS Theory manual,[15] Sec. 3.1.1, for more details).

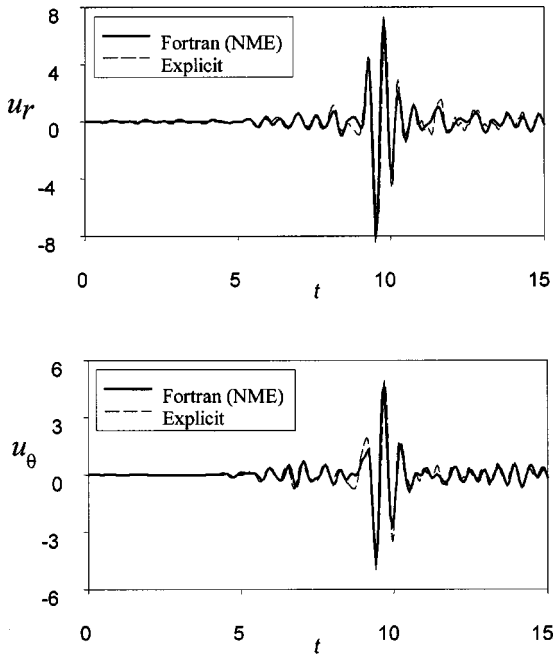


FIG. 2. Comparison between wave forms from Abaqus/Explicit and FORTRAN [normalized units (Ref. 12)].

Each node has 2-degrees of freedom (plane strain assumption).

Cracks in the cylinder are modeled by releasing nodes along the cylinder's wall and are assumed to have perfectly smooth faces, subjected to stress free boundary conditions. Such mesh discontinuities are infinitely thin, are a good representation of a fatigue crack (as opposed to notches), and allow for a variable crack length. In the far field (which is the domain of interest here). For near field calculations, singular elements must be used to accurately capture the singularity at the crack tip. It is important to note that the inverse problem solved in this study does operate under the assumption that the shape of the crack is known *à priori*. In addition, while

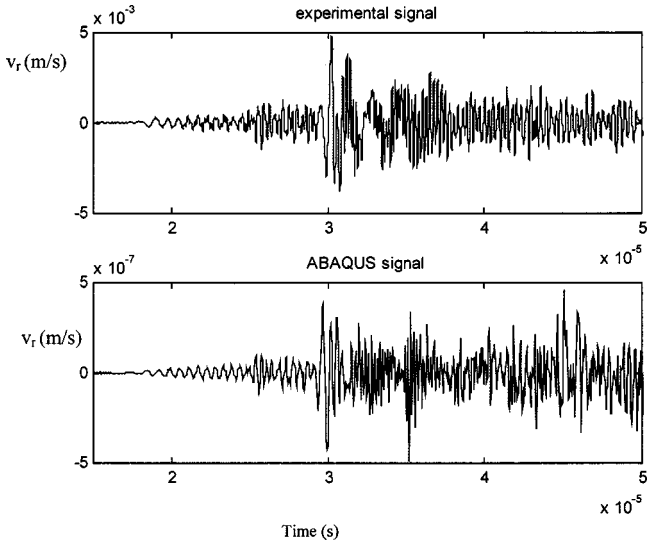


FIG. 3. Comparison between a laser-based experimental signal and the equivalent ABAQUS/Explicit signal.

TABLE I. Material and geometry data for the thin steel annulus.

v	c_T	c_L	a	b	η
0.2817	3120 m/s	5660 m/s	5.08 cm	5.38 cm	0.944 24

crack closure is of clear concern in practical cases, it will not be addressed in this work.

A finer mesh is needed to accurately model cylinders with smaller cracks (i.e., of length smaller than 20% of the wall of the waveguide). Convergence is usually obtained after increasing the number of elements in both the thickness and the waveguide axis directions. Therefore, the computational models presented here can detect cracks of length no smaller than 10% of the wall, that is, no smaller than 0.3 mm for a load with 0.5 MHz center frequency. More powerful computers can capture scattering due to even smaller cracks because they can accurately resolve the bigger meshes needed for loads of higher center frequencies.

The load in all cases is a point load, applied normally to the cylinder's surface. This load has a frequency range of 0.2 to 1.5 MHz, with most of the energy centered near 0.5 MHz. The solution of such a FEM model takes a few minutes on a Pentium III with a 433 MHz clock and 254 MB RAM.

III. AULD's FORMULA

Auld's formula[9,16] is developed using a two transducer, through transmission system applied to the structure to be interrogated. Transducer 1 produces an incident field of power P, and transducer 2 is the receiver [see Figs. 4(a) and (b)]. The ratio of received electrical signal strength, E_{II}, to incident electrical signal strength, E_I, is denoted by Γ. The change in the ratio, $\delta\Gamma$, due to a single scatterer (a crack in this case), is proportional to the reflection coefficient R and is given by Auld's formula:

$$\delta\Gamma = ((E_{II})_{\text{flaw}} - (E_{II})_{\text{noflaw}})/(E_I)_{\text{flaw}}. \tag{1}$$

This formula may be simplified for the case of backscattering, and with the hypothesis that the signals are time harmonic:

$$\delta\Gamma = -\frac{i\omega}{4P}\int_S (\sigma_{kj}^{(2)} u_k^{(1)} - \sigma_{kj}^{(1)} u_k^{(2)}) n_j \, dS, \tag{2}$$

where S is an arbitrary surface, which surrounds the scatterer (the crack in this case), j and k are dummy indices (the summation convention is used here) and n_j is the unit outward normal of S. In Eq. (2) the terms with superscript (1) relate to the fields in the absence of the scatterer, while terms with the superscript (2) relate to the fields in the presence of the scatterer. For the case where the scatterer is a traction free crack, Eq. (2) is simplified to

$$\delta\Gamma = -\frac{i\omega}{4P}\int_A (\sigma_{kj}^{(1)} \Delta u_k^{(2)}) n_l \, dS, \tag{3a}$$

where A is the crack area and

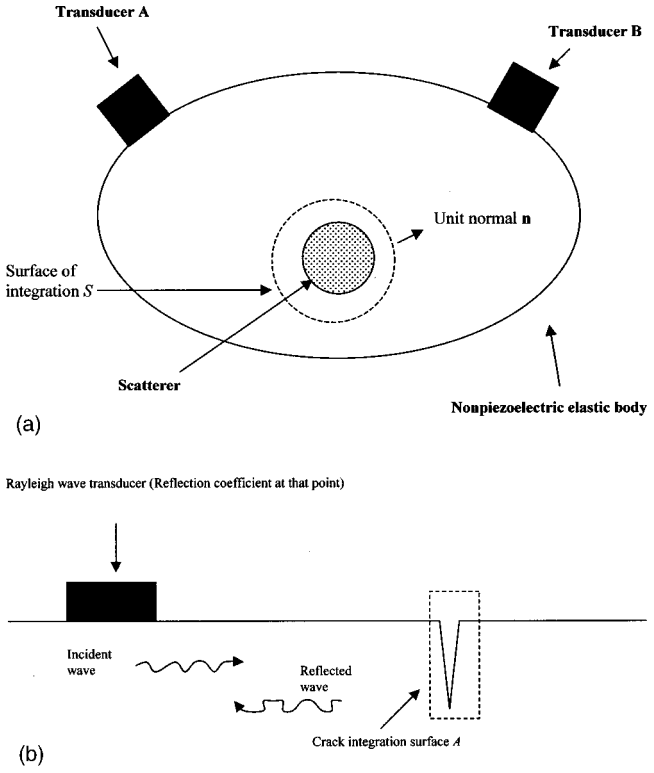


FIG. 4. (a) Auld's formula (general scattering geometry). (b) Auld's formula (backscattering case).

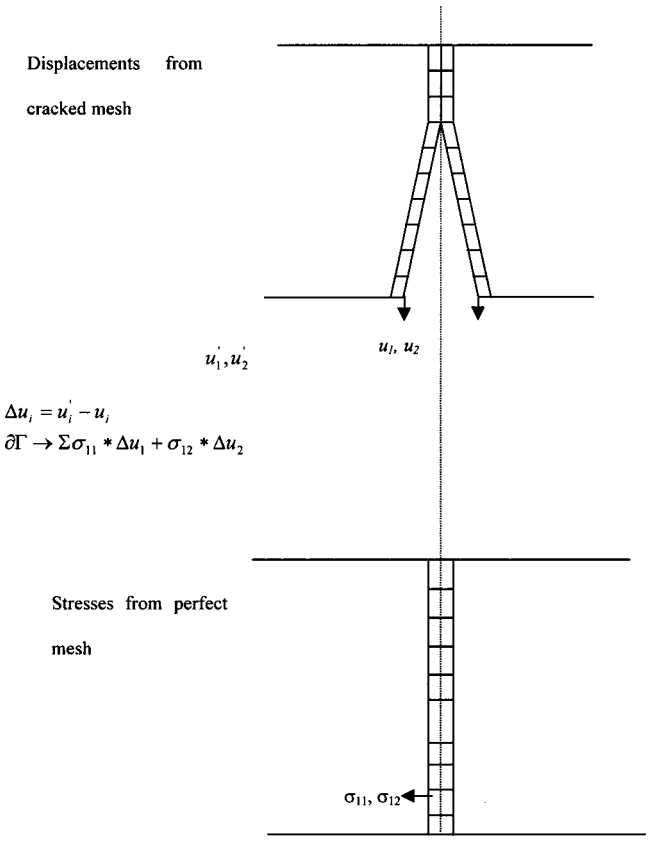


FIG. 5. Auld's formula—scheme of FEM computations.

$$\Delta u_k^{(2)} = u_k(x_2^-) - u_k(x_2^+). \tag{3b}$$

The two previous equations are used to calculate the backscattering from a surface crack at the inner radius of the cylinder.

Since the stress and displacement signals obtained from ABAQUS are transient and not time harmonic, they cannot be multiplied directly into the integrand of Eq. (3). The typical procedure[17,18] involves applying the fast Fourier transform (FFT) to the transient displacement and stress signals (the signals calculated with ABAQUS) first, then multiplying them (element by element, array multiplication and *not* vector multiplication), repeating the steps for each new value, and finally adding everything. Once the sum is obtained, it has to be multiplied by $i\omega$ (element by element again) and the resulting output is Auld's formula in the frequency domain. In addition, the inverse FFT has to be applied if a time domain result is desired. This procedure is lengthy and requires careful attention to the proper application of the FFT. In particular, zero-padding the time-domain data is absolutely necessary. Otherwise, the final signal violates causality.

An interesting alternative to the frequency-domain formulation is to adapt Eq. (3) so that it can be applied directly to time-domain signals. It is well known that the convolution in the time domain is equivalent to the multiplication in the frequency domain. In fact, many convolution programs use that property directly in their codes: they first zero-pad the data, apply the FFT, array multiply, and apply the inverse FFT to get the final outcome.[19] Equivalently, MATLAB uses a filter-based approach whereby the structures of the data are cross multiplied:

$$y(n) = f(n) * g(n) = \sum_{m=-\infty}^{\infty} f(n-m).g(m). \tag{4}$$

The advantage of using the cross multiplication, as done in MATLAB, is that much fewer operations are required (than in the alternative FFT-based formulation) and that the output is directly in the time domain. The only additional step that is required for agreement with Eq. (3) is a time differentiation, since multiplying by $i\omega$ in the frequency domain is equivalent to differentiating (with respect to time) in the time domain.

The time-domain reflection coefficient, R, is therefore calculated by using the following formula (up to a multiplicative constant):

$$R \propto \frac{d}{dt}\left(\sum_n u_i^n * \sigma_{i1}^n \right), \tag{5}$$

where n is the number of the element in consideration (for the stress tensor) and of that element's bottom node on the crack surface (for the displacement field), $*$ denotes convolution, and $i = 1,2$.

All the cracks considered in this study are radial (e.g., vertical along the 90° angle, or the y axis). Therefore, the stress tensor will always have a normal in the 0° angle direction, or the x axis (see Fig. 5). Therefore, the j index is always equal to 1.

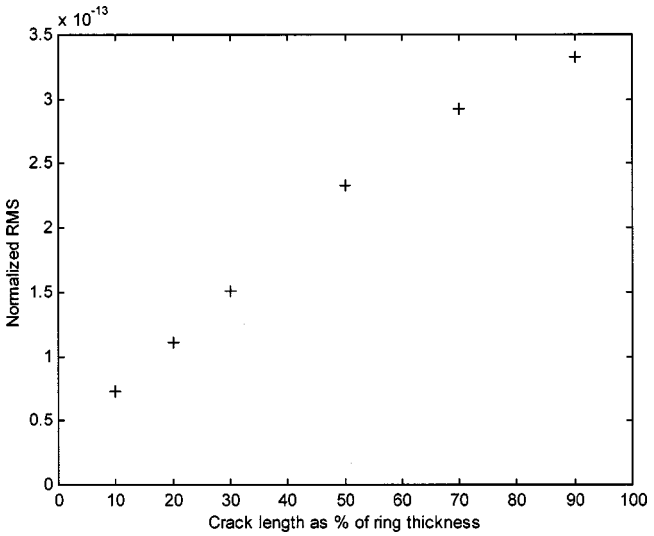


FIG. 6. Normalized RMS of FFT of Auld's reflection coefficient R versus crack depth.

IV. REFLECTION COEFFICIENT VERSUS CRACK LENGTH

Clearly, the longer the crack, the higher the overall level of the resulting reflection coefficient. One way to obtain a quantitative measure of this relationship is to plot the RMS (root-mean-square) value of the frequency spectrum of the reflection coefficient obtained via Auld's formula, for various crack lengths. That is, once the FE model is used to calculate the field around the crack, Auld's formula is then used to predict the length of the crack with this FE data. The procedure involves a series of steps.

(1) Construct a FEM model of the cylinder with a prescribed crack length.
(2) Solve for the pertinent displacement and stress fields around the crack (for the load case with a center frequency of 0.5 MHz).
(3) Obtain the time-domain (transient) reflection coefficient, R, from the procedure described in the preceding section and by applying Eq. (5), up to an arbitrary multiplicative constant.
(4) Apply the FFT to R to get its frequency spectrum.
(5) Get the RMS value of R's frequency spectrum, and normalize it with respect to the RMS value of the incident signal.
(6) Plot the normalized RMS value with respect to crack length, and repeat the entire procedure for other crack lengths.

The resulting plot is shown in Fig. 6 for six different crack lengths (10%, 20%, 30%, 50%, 70%, and 90% of cylinder wall). Figure 6 demonstrates that there is a quasilinear correlation between the crack length and the overall value of the reflection coefficient. For a perfect cylinder (crack length of 0%) there should be no backscattered energy and thus the normalized RMS should be 0. If the cylinder is completely cut (crack length of 100%) the backscattered energy should be equal to the incident energy. Figure 6 shows that the slope is steeper for crack depths smaller than 10% of the cylinder wall. The slope also tends to flatten slightly as the crack depth increases beyond 70%. Please note that mesh size (the number of elements needed to ensure convergence) of the ABAQUS model increases as the crack length decreases.

The quasilinear characteristic of the RMS plot shows that Auld's method, combined with the FEM model, gives a quantitative estimate of the crack length in a cracked steel cylinder using guided circumferential waves. Moreover, this method allows the sizing of cracks up to 300 μm even though the signal frequency is centered at 0.5 MHz (and never goes beyond 1.5 MHz). Note that a Rayleigh wave with a wavelength of 300 μm has a frequency of 9.7 MHz in steel.

V. TIME-FREQUENCY REPRESENTATIONS (TFRs) OF GUIDED WAVES

Now that a crack's length can be quantitatively measured, consider a procedure to determine the crack's location. Once again the main challenge is the dispersive and multimode nature of guided waves. For example, consider a time-domain scheme that uses arrival times to determine the unknown location of a crack on a cylinder's inner surface. Since this scheme must be based on the arrival times of a particular wave form feature, the dispersive nature of guided waves presents a large source of error. The reason for this (potential) breakdown is that the different frequencies of a dispersive wave travel with different group velocities, changing the wave's shape as it propagates. As a result, it is difficult (if not impossible) to track and identify the exact arrival time of the same feature of a propagating guided wave.

In order to alleviate this problem, changes in the frequency content of the signal need to be tracked as a function of time so that dispersion can be taken into account.

The time-frequency representation (TFR) of a signal is a quantitative measure of how a signal's frequency changes with respect to time. A TFR is obtained by dividing a time-domain signal into a series of small pieces in time; each of these pieces is windowed and then individually transformed into the frequency domain. Popular transforms include wavelets, the short time Fourier transform (STFT), and the Wigner–Ville distribution (WV). Please recall from the preceding discussion that the identification of an individual mode's arrival time is very difficult (if not impossible) from either the time-domain signal or its frequency spectrum, because the contributions from each mode in a multimode signal are not separable. However, the TFR enables the separation of the contribution of each mode, as a function of time *and* frequency simultaneously.

One problem inherent to a TFR is the time-frequency resolution limitation, that is, the impossibility to simultaneously have perfect resolution in both time and frequency. One way to increase time-frequency resolution, is through the reassignment (or reallocation) method.[20] Reassignment is not another TFR, but a way to reduce the spread of a TFR by concentrating its energy to its (the energy's) center of gravity. The reassignment method is not restricted to a specific TFR, but can be applied to any time-frequency shift invariant distribution of Cohen's class.[21] The energy density spectrum of the reassigned STFT, called the reassigned spectrogram, is

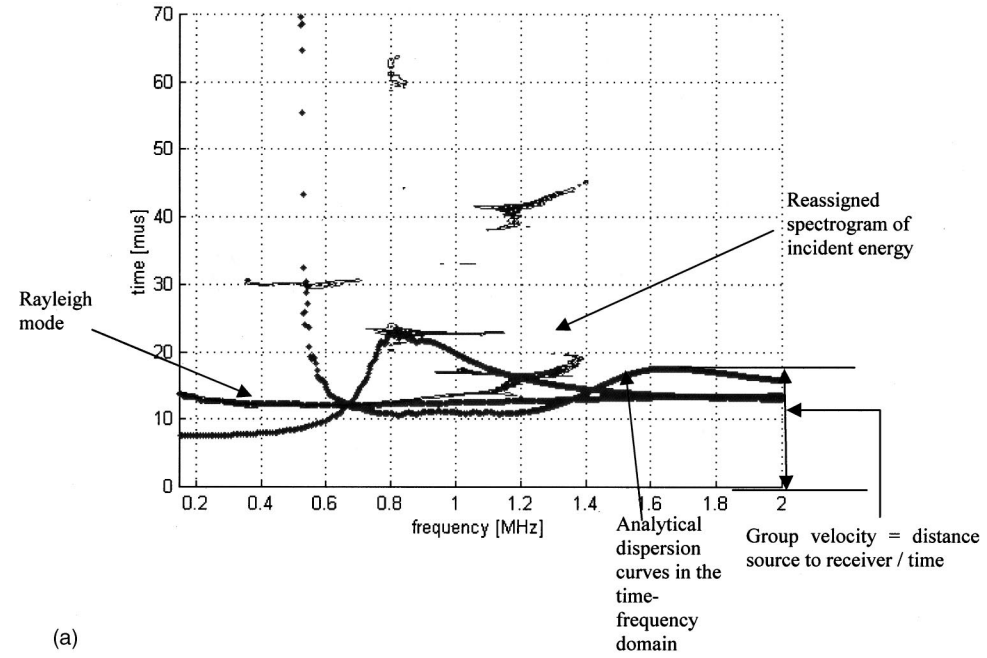


FIG. 7. (a) Comparison between the reassigned spectrogram of a FE-obtained guided circumferential wave and the analytical dispersion curves, in the time-frequency domain, of the same cylinder (both perfect). (b) Methodology to locate the crack: comparison between the reassigned spectrogram for the backscattered energy and the analytical dispersion curves in the time-frequency domain.

selected for this study, since previous work[21,22] shows that it is extremely effective in capturing the dispersive nature of guided, multimode ultrasonic signals in thin metallic plates.

Therefore, as was done in earlier studies for plates,[22] it is possible to compare analytical dispersion curves for the cylinder (calculated in the time-frequency domain) to a TFR calculated from either experimentally or numerically (FEM) generated guided circumferential wave signals. [See Fig. 7(a) for a comparison between the reassigned spectrogram of a synthetic, i.e., numerical, signal from a perfect cylinder and the analytical dispersion curves, calculated in the time-frequency domain, for that same perfect cylinder.] The ''ladder'' effect described earlier in plates[22] is clearly still present for guided circumferential waves, and although the Rayleigh mode is well represented by the TFR, the procedure breaks down beyond 0.8 MHz because the FE model's mesh is optimized for frequencies around 0.5 MHz.

One of the greatest advantages of the proposed technique is that both numerical and experimental TFRs are not limited to perfect geometries, while analytical dispersion curves (in general) have this limitation. It is thus possible to compare *in the time-frequency domain* analytically obtained dispersion curves (for a *perfect* cylinder) to a reassigned spectrogram obtained from a signal numerically calculated (or experimentally measured) in a *cracked* cylinder, using this comparison to quantify the crack's effect on these dispersion curves. *This comparison offers a systematic procedure to locate a crack that does not require the selective generation or detection of a particular mode.* Applying the reassigned spectrogram to *backscattered* ultrasonic energy and comparing it to analytical dispersion curves for the cylinder (calculated in the time-frequency domain) provides an excellent way to calculate the arrival times of specific modes (see Fig. 1 for the positioning of the source, the receiver and

the crack). A typical plot that results from the proposed procedure is given in Fig. 7(b).

From Fig. 7(b) it is possible to determine:

- which modes are affected by the crack,
- for what times and what frequencies,
- and even more importantly, by how much time is the arrival of each mode changed by the mode's interaction with the crack, for any given frequency.

This information is then used to determine the crack's location. The next section will illustrate how this information is extracted from the signals.

VI. CRACK LOCATION USING THE REASSIGNED SPECTROGRAM

This portion of the study develops FE models for two cases of the same steel cylinder, one with a crack length of 10% of the cylinder's wall (on the inner surface) and a load of center frequency of 0.5 MHz, and the second is the associated ''perfect'' cylinder (no crack). The time-domain signal backscattered from this crack is determined by subtracting the signal predicted with the cracked cylinder model from the signal of the perfect cylinder (i.e., the model without a crack), noting that each signal has the *same source and receiver locations*. The reassigned spectrogram is then applied to this backscattered signal, and this TFR is compared to the (perfect) cylinder's analytical dispersion curves (calculated, in the time versus frequency domain, for the *same source and receiver locations*). This comparison illustrates which modes are affected by the crack, as well as enabling the calculation of the time delay between the backscattered and the corresponding analytical dispersion curves [see Fig. 7(b)]. The reason for this time delay is that the backscattered signal only starts after the incident signal has propagated from the source to the receiver (as in the perfect cylinder case) and *then* to the crack *and* then back to the receiver. Therefore, the backscattered signal is zero for all times prior to the time that it takes to return to the receiver.

This time delay is given by the ratio between the group velocity of each specific mode, for a specific frequency, and twice the distance between the receiver and the crack. Note that there is a distortion in the backscattered reassigned spectrogram, when compared to the analytical dispersion curves, because the group velocity of most modes changes with frequency—this causes a change in the time delay, with frequency. As a result, the shape of the backscattered spectrogram looks slightly different from the shape of the corresponding analytical dispersion curves for the perfect cylinder [see Fig. 7(a)].

This distortion makes it difficult to identify exactly which feature of a given backscattered mode, for a specific frequency, corresponds to what feature of the analytical dispersion curve for that mode and for what frequency. For example, in Fig. 7(a), the highest peak in the second mode occurs at 0.82 MHz in the analytical dispersion curve, but finding this exact same feature in the backscattered spectrogram for the second mode is impossible since the peak is now spread over a small frequency range (1 to 1.2 MHz). Therefore, since the group velocity depends on frequency,

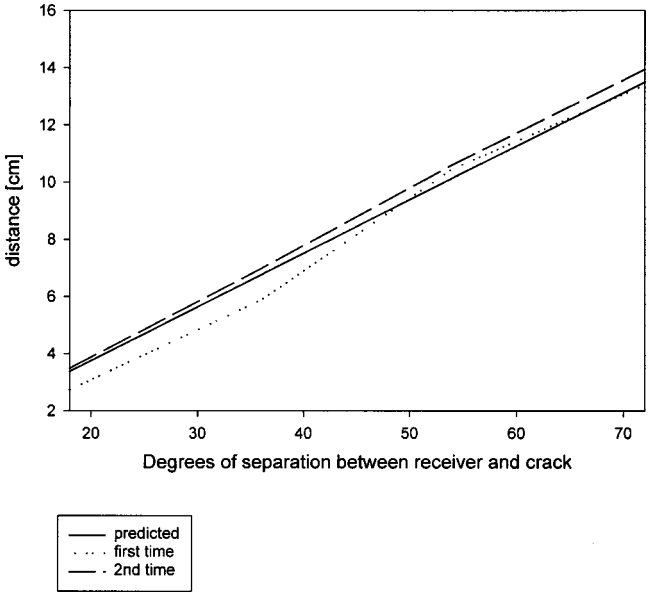


FIG. 8. Comparison between predicted distances receiver to crack (lower and upper bound) and the true distance, for the ring cracked at 10%.

the time delay between the reassigned spectrogram of a backscattered mode and its dispersion curve will also be hard to determine accurately. The only mode for which the preceding statement is not true is the Rayleigh mode, because its group velocity is constant for almost all frequencies, so the frequency distortion between the Rayleigh mode's analytical dispersion curve and its reassigned spectrogram is nearly nonexistent [beyond 0.4 MHz, shown in Fig. 7(a)]. *The time delay between the backscattered Rayleigh mode from the reassigned spectrogram and its predicted analytical arrival, times the group velocity of the Rayleigh mode, gives twice the distance between the receiver and the crack.*

Since the scattered Rayleigh mode, as calculated with the reassigned spectrogram, is not an infinitely thin line [as is clear from Figs. 7(a) and (b)], two distances are always calculated, one corresponding to the lowest bound in time (first time), and one corresponding to the upper bound in time (second time). At 0.5 MHz, the Rayleigh wave can be assumed to be almost nondispersive, thus if it reaches the receiver in a faster time, it has traveled a comparatively shorter distance. Hence the ''lower'' bound on the distance to, or location of, the flaw. Conversely the second and longer time corresponds to a longer distance traveled by the backscattered signal to the receiver. Hence the lower bound on the distance to the flaw. These distances can then be plotted against the true distance, i.e., the actual location of the crack. The comparison between those three distances is given in Fig. 8. There is very good agreement between them, even for this small (10% of wall thickness) crack.

The proposed methodology also works well in the nearfield of the crack (or source), although it is more difficult to accurately determine separation times between backscattered and analytical modes in this case; for example, the analytical modes will tend to be compressed all together toward the zero time axis if the receiver is close to the source. Since the method is based on a graphical calculation of time

delay, it is essential to have a definite separation between the analytical dispersion curves and their backscattered counterparts.

More importantly, the Rayleigh mode is not always the mode most sensitive to the crack. Indeed, in Fig. 7(b), one can see that for certain frequency ranges (for example, from 1 MHz on), the backscattered contribution from the Rayleigh mode simply vanishes. This means that for frequencies above about 1 MHz, the Rayleigh mode does not ''see'' the crack (this makes sense since the Rayleigh mode travels on the outside radius while the crack originates from the inside surface—therefore, as the frequency increases the Rayleigh mode is increasingly less likely to see the crack). Only the higher order modes (in this case the second mode) see the crack at that particular frequency (as was first shown in Valle *et al.*[4])—this is clearly visible in Fig. 7(b). Unfortunately, frequency distortion is a problem for these higher modes. This frequency distortion makes it difficult to use higher order modes to determine the location of the crack, unless one can select a frequency range with limited dispersion, so that the group velocity can be calculated accurately. Typically this frequency range exists, but it may be prohibitively high from a computational cost perspective. Please note that dispersion curves and FE signals tend to be costly for frequencies beyond 3 to 4 MHz.

This problem (the Rayleigh mode not being sensitive to cracks) only occurs for very small cracks. The reassigned spectrogram applied to time-domain signals created by a 0.5 MHz signal is clearly sufficient to locate this 300 μm crack (10% of the cylinder's thickness).

VII. SUMMARY AND CONCLUSIONS

This paper uses the FEM to model the propagation of guided circumferential waves in a cracked cylinder and presents sophisticated digital signal processing algorithms that characterize the waves' interaction with this crack. Specifically, this paper presents a technique to size the crack, and another to locate it.

For sizing the crack, a time domain analysis based on Auld's scattering formula for backscattered signals is used. By using the Auld's formula to calculate the RMS value of the frequency spectrum of the reflection coefficient for various crack lengths, it is shown that there is a quasilinear correlation between the crack depth and the overall RMS value of the reflection coefficient R. The RMS value of R increases with crack depth, and provides a quantitative means of characterizing cracks down to about 10% of the thickness of the cylinder. As a result, Auld's method combined with the FEM model gives a quantitative estimate of the crack depth from a measured signal.

To locate the crack, the reassigned spectrogram is presented as a digital signal processing algorithm that accurately captures a TFR of guided circumferential waves over a large frequency range. The reassignment method solves the time-frequency resolution problem inherent to using Fourier-based TFRs such as the spectrogram by redistributing the energy content of each mode to its center of gravity, with respect to time and frequency, and therefore clearly separates the contribution of each mode within the scattered signal. Applying the reassigned spectrogram to backscattered data, and comparing this to the analytical dispersion curves for the perfect cylinder provides an accurate measure of the time delay, per mode, between the incident signal of a perfect waveguide and its backscattered counterpart (if a crack is present). Through this time delay and the group velocity of the mode of interest at a specific frequency, the distance between the receiver and the crack can be calculated. This method can characterize cracks and notches with lengths as small as 10% of the waveguide wall. Computational power needs to be increased if even smaller sizes are of interest.

It is important to note that the results presented in this paper, both for crack sizing and crack locating, are dependent on the frequency of the input signal. Both methods (Auld's formula and the reassigned spectrogram) are applied to signals created with an input with a frequency range between 0.2 and 1.5 MHz (with most of the energy centered around 0.5 MHz) but they can detect cracks down to 300 μm—even though the wavelengths of these signals is much greater than 300 μm. Therefore, the need for high frequency signals to detect small cracks is significantly reduced by using these techniques.

Also, multiple cracks can be detected using those techniques. The reassigned spectrogram will locate them by showing a series of time delays. Auld's formula (as presented here with no restriction on the duration of the signals) can only show a combined total of damage accumulation—however, if it is used in combination with the reassigned spectrogram, the signals used in Auld's formula can be restricted to the pertinent time of flight and therefore can then characterize each crack singly, in the same fashion presented earlier in this paper.

ACKNOWLEDGMENTS

This work is supported by the Office of Naval Research M-URI Program ''Integrated Diagnostics'' (Contract No. N00014-95-1-0539). Two Amelia Earhart Fellowships from the Zonta Foundation, and support from the Deutscher Akademischer Austausch Dienst (DAAD) are also gratefully acknowledged by the first and second authors, respectively.

[1] P. B. Nagy, M. Blodgett, and M. Godis, ''Weep hole inspection by circumferential creeping waves,'' NDT & E Int. **27**, 131–142 (1994).
[2] D. E. Chimenti, ''Guided waves in plates and their use in material characterization,'' Appl. Mech. Rev. **50**, 247–284 (1997).
[3] D. N. Alleyne and P. Cawley, ''The long range detection of corrosion in pipes using lamb waves,'' Rev. Prog. Quant. Nondestr. Eval. **14**, 2073–2080 (1995).
[4] C. Valle, J. Qu, and L. J. Jacobs, ''Guided circumferential waves in layered cylinders,'' Int. J. Eng. Sci. **37**, 1369–1387 (1999).
[5] C. Valle, ''Guided circumferential waves in annular structures,'' Ph.D. thesis, Georgia Institute of Technology, Atlanta, Georgia, 1999.
[6] I. A. Viktorov, *Rayleigh and Lamb Waves* (Plenum, New York, 1967).
[7] I. A. Viktorov, ''Rayleigh-type waves on a cylindrical surface,'' J. Acoust. Soc. Am. **4**, 131–136 (1958).
[8] G. Liu and J. Qu, ''Guided circumferential waves in a circular annulus,'' J. Appl. Mech. **65**, 424–430 (1997).
[9] B. A. Auld, *Acoustic Fields and Waves in Solids*, 2nd ed. (R. E. Krieger, Melbourne, FL, 1990), Vols. 1 and 2.
[10] J. Zhu, A. H. Shah, and S. K. Datta, ''Transient response of a composite plate with delamination,'' J. Appl. Mech. **65**, 664–670 (1998).
[11] S. K. Datta, T. H. Ju, and A. H. Shah, ''Scattering of an impact wave by a crack in a composite plate,'' J. Appl. Mech. **59**, 596–603 (1992).

[12] F. Moser, L. J. Jacobs, and J. Qu, ''Modeling elastic wave propagation in wave guides with the finite element method,'' NDT & E Int. **32**, 225–234 (1999).
[13] M. J. S. Lowe, R. E. Challis, and C. W. Chan, ''The transmission of Lamb waves across adhesively bonded lap joints,'' J. Acoust. Soc. Am. **107**, 1333–1345 (2000).
[14] G. Liu and J. Qu, ''Transient wave propagation in a circular annulus subjected to transient excitation on its outer surface,'' J. Acoust. Soc. Am. **104**, 1210–1220 (1998).
[15] ABAQUS-Standard version 5.6 User's Manuals, Theory Manual, Example Problems Manual, and ABAQUS-Post Manual, 1996.
[16] B. A. Auld, ''General electromechanical reciprocity relations applied to the calculation of elastic wave scattering coefficients,'' Wave Motion **1**, 3–10 (1979).
[17] M. J. S. Lowe, P. Cawley, J-Y. Kao, and O. Diligent, ''Prediction and measurement of the reflection of the fundamental anti-symmetric Lamb wave from cracks and notches,'' Rev. Prog. Quant. Nondestructive Eval. **19**A, 193–200 (2000).
[18] M. Lowe, ''Characteristics of the reflection of lamb waves from defects in plates and pipes,'' Rev. Prog. Quant. Nondestr. Eval. **17**, 113–120 (1997).
[19] W. H. Press, B. P. Flannery, S. A. Tenkolsky, and W. T. Vetterling, *Numerical Recipes* (Cambridge University Press, Cambridge, 1988).
[20] F. Auger and P. Flandrin, ''Improving the readability of time-frequency and timescale representations by the reassignment method,'' IEEE Trans. Signal Process. **43**, 1068–1098 (1995).
[21] M. Niethammer, ''Application of time frequency representation to characterize ultrasonic signals,'' M.S. thesis, Georgia Institute of Technology, Atlanta, Georgia, 1999.
[22] M. Niethammer, L. J. Jacobs, J. Qu, and J. Jarzynski, ''Time-frequency representation of lamb waves using the reassigned spectrogram,'' J. Acoust. Soc. Am. **107**, L19–L24 (2000).

Coherent backscattering of elastic waves: Specific role of source, polarization, and near field

B. A. van Tiggelen[a)]
CNRS/Laboratoire de Physique et Modélisation des Milieux Condensés, Université Joseph Fourier, Maison des Magistères, B.P. 166, 38042 Grenoble Cedex 09, France

L. Margerin and M. Campillo
Laboratoire de Géophysique Interne et Tectonophysique, Observatoire de Grenoble, Université Joseph Fourier, B.P. 53, 38041 Grenoble Cedex, France

(Received 20 July 2000; revised 10 March 2001; accepted 2 May 2001)

Calculation of coherent backscattering of elastic waves in an infinite isotropic random medium is presented. Despite the simplicity of this geometry, this calculation highlights several specific aspects for seismic detection: near field detection, polarization, and the symmetry of the source. Line profiles and enhancement factors are seen to be time independent and are calculated for kinetic, shear, and compressional energy. © *2001 Acoustical Society of America.*
[DOI: 10.1121/1.1388017]

PACS numbers: 43.20.Gp, 43.20.Bi, 91.30.Dk [DEC]

I. INTRODUCTION

Coherent or enhanced backscattering finds its origin in interference effects in multiple scattering of waves. Basic is the reciprocity principle which states that partial waves propagating in opposite directions interfere constructively at the source from which they originate, no matter how complex the wave path.[1] These ideas were originally developed in condensed matter,[2] where such constructive interferences give rise to the so-called ''weak localization'' corrections to the electrical conductance. In the beginning of the 1980s, coherent backscattering of light was observed independently by different groups.[3,4] Since then, a lot of theoretical and experimental efforts have been undertaken to understand this novel effect in detail, such as the role of an external magnetic field, which breaks time-reversal symmetry,[5,6] chirality,[6] line shape,[7] enhancement factor,[8] stimulated emission,[9] and Raman scattering.[10] More recently, time-dependent enhanced backscattering was reported for acoustic waves.[11]

Wave scattering is also of interest in seismic applications. Since the pioneering work of Aki[12] and Aki and Chouet[13] three decades ago, seismic coda is interpreted as elastic waves scattered from randomly distributed inhomogeneities in the Earth's crust. The seismic coda refers to the exponential time tail observed in the seismograms of local earthquakes in the frequency band 1–10 Hz. Whereas early work[12,14,15] tried to model the coda as singly scattered waves in a uniform space, recent numerical studies suggested the importance of multiple scattering.[16–19] In the above-mentioned frequency band the mean free path of the elastic waves is estimated to be $l \approx 20-70$ km.[20] Given a typical velocity of 3.5 km/s the mean free time between two scattering events would be 20 s maximum. This time scale supports the view that seismic coda, which often lasts for more than 200 s, is a genuine multiple scattering phenomenon. A detailed theoretical study of enhanced backscattering of elastic waves is necessary to facilitate its observation in seismic data.

The presence of three polarizations and two different speeds makes any multiple scattering problem in elasticity considerably more complicated than in electromagnetism. The technical aspects have been studied in great detail by Weaver,[21] Turner and Weaver,[22] and Papanicolaou *et al.*[23] Seismic observations are characterized by a number of additional specific aspects, and the present work aims to address some of them.

First, seismic measurements are always carried out in the near field, i.e., near or in the scattering medium, in contrast to previous far-field scattering studies with light and acoustic waves. Our first study,[24] carried out for acoustic waves, has revealed that observation of enhanced backscattering of seismic waves requires detection within approximately one elastic wavelength (typically 100 m to 1 km) from the seismic source, making it a genuine near field effect. This characteristic distance is quite different from the one in the far field, where it equals the much longer mean free path.[8,11] The line shape of enhanced backscattering can be defined as the ratio of enhancement near the source, normalized to the energy measured far way. Studies with light[25] and acoustic waves[11] in the far field showed this ratio to be a function of time. In the near field, a *stable* line shape was predicted at times longer than the mean free time, with a maximal enhancement factor equal to 2 at the source and a typical distance of one wavelength for the enhancement to vanish. The present work investigates the precise line profile for enhanced backscattering in an infinite medium, using the full theory of elasticity. The complex role of the nearby free surface, which induces mode conversions, is left to future work.

The second specific aspect is that seismic sources are known to be highly polarized. The role of polarization has been shown to be crucial in optical studies, since the reciprocity principle applies only when detector and source have the same polarization.[1,7] The explosions have a diagonal seis-

a)Electronic mail: tiggelen@belledonne.polycnrs-gre.fr

mic moment tensor, releasing compressional energy only. Being relatively under control, they seem good candidates for enhanced backscattering studies. Unfortunately, because of the low shear velocity, multiple scattering rapidly converts most elastic energy into shear excitations,[26] and an observation of enhanced backscattering with explosions seems *a priori* hopeless. On the other hand, earthquakes are often modeled by a dislocation, with a traceless, symmetric seismic moment tensor, characteristic of a source involving a double couple.[27] The diagonal elements of the seismic moment tensor of an earthquake, corresponding to volume changes, are estimated to be small, typically of the order of 2%. This particular symmetry of the source has never been discussed in the context of multiple scattering and enhanced backscattering.

Our study is closely related to a recent study by De Rosny, Tourin, and Fink[28] and Weaver and Lobkis,[29] which involve the enhanced backscattering effect of elastic waves in a reverberant cavity. The main difference is that in reverberant cavities the mixing of the waves is caused by the chaotic nature of their boundaries, and not by elastic scatterers inside. Second, in a finite cavity, an additional characteristic time exists, the Heisenberg time, beyond which the enhanced backscattering factor changes from 2 to 3. This Heisenberg time is infinite for an infinite random medium. In this paper we show that the enhancement factor for coherent backscattering in an infinite elastic medium is necessarily less than or equal to 2. Sometimes the enhancement can be so small that observation is very unlikely, since the effect is easily overwhelmed by measurement errors. The study by De Rosny *et al.*[28] of two-dimensional enhanced backscattering of elastic waves in a chaotic silicon wafer has revealed a very unconventional line shape, vanishing exactly at the source and with two maxima at roughly one wavelength from the source. This line profile was explained by the *dipolar* nature of the source.[30]

II. SOURCE: EXPLOSIONS AND DISLOCATIONS

In general, seismic sources can be described by a force field $f_i(\mathbf{r},t)$ localized in space and time. This force field contains a wide range of frequencies ω and we shall represent it by its Fourier transform $f_i(\mathbf{r},\omega)$. For sources small compared to the elastic wavelength, the following multipole expansion applies:

$$f_i(\mathbf{r},\omega)=F_i(\omega)\,\delta(\mathbf{r})-M_{in}(\omega)\,\partial_n\delta(\mathbf{r})+\cdots. \tag{1}$$

The net force F_i and the seismic moment tensor M_{in} are defined by

$$F_i(\omega)\equiv\int d\mathbf{r}\, f_i(\mathbf{r},\omega),\quad M_{in}(\omega)\equiv\int d\mathbf{r}\, r_n f_i(\mathbf{r},\omega), \tag{2}$$

where the integral is carried out over the whole source. The process initiating the source has no external perturbations, so that the net force F_i and the net couple $M_{ij}-M_{ji}$ $(i\neq j)$ must vanish. Thus, the seismic moment tensor must be *symmetric*. The diagonal elements M_{ii} induce no couple but create compressional deformations, typical of explosions. An ideal, isotropic explosion can be characterized by the seismic moment tensor $M_{ij}=M\delta_{ij}$. Earthquakes are believed to be dislocations with very small volume changes. This reduces the model to one with two compensating dipolar couples, hence the name ''double couple.'' Only vulcano eruptions are examples of seismic sources for which a net force $F_i(\omega)$ is believed to exist.

The force field $f_j(\mathbf{r},\omega)$ acts as the source term in the elastic wave equation for the displacement vector $u_i(\mathbf{r})$, whose Green's tensor is denoted by G_{ij}.[27] From Eq. (1), it follows immediately that the elastic wave released at the source is

$$u_i(\omega,\mathbf{r})=M_{nj}(\omega)\,\partial_n G_{ij}(\omega,\mathbf{r}). \tag{3}$$

This source field will be used as a starting point for coherent backscattering calculations. In an isotropic homogeneous elastic medium the elastic Green's tensor is[27]

$$\begin{aligned}G_{ij}(\omega,\mathbf{r})=&-\frac{3\hat{r}_i\hat{r}_j-\delta_{ij}}{4\pi r^3}\int_{r/\alpha}^{r/\beta}dy\, y\exp(i\omega y)\\&-\frac{\hat{r}_i\hat{r}_j}{4\pi\alpha^2 r}\exp(i\omega r/\alpha)\\&+\frac{\hat{r}_i\hat{r}_j-\delta_{ij}}{4\pi\beta^2 r}\exp(i\omega r/\beta).\end{aligned} \tag{4}$$

The inequality $\alpha\neq\beta$ of the two elastic velocities gives rise to the first near-field term.

III. COHERENT BACKSCATTERING IN THE NEAR FIELD

In the following we assume that multiple scattering does not affect the frequency ω, and often drop the frequency label. The basic observable used in radiative transfer of elastic waves is the correlation function of two elastic displacement vectors $\langle u_i(\mathbf{r}-\frac{1}{2}\mathbf{x},t)u_j^*(\mathbf{r}+\frac{1}{2}\mathbf{x},t')\rangle$ at two different positions and for two different times. One time label can be transformed away in favor of a central carrier frequency ω. The time label that remains describes the time evolution of the ''slowly varying'' envelope of the pulse, which is much slower than the harmonic cycle.

The kernel $L_{ij\to kl}(\mathbf{r}_1,\mathbf{r}_2\to\mathbf{r}_3,\mathbf{r}_4,t)$ describes the transfer of the displacement correlation function from source (displacement indices i, j and positions $\mathbf{r}_1$, $\mathbf{r}_2$) to detector (displacement indices k, l and positions $\mathbf{r}_3$, $\mathbf{r}_4$). Standard radiative transfer theory neglects all interference effects in the vertex L and assumes that the field $\mathbf{u}$ and its complex conjugate $\mathbf{u}^*$ travel along exactly the same paths, so that their phase difference cancels. This leads to the diagrammatic ''ladder'' representation of L shown in Fig. 1(a). The solid lines denote a Green's tensor that quantifies the radiation from one scatterer to the next, displayed as crosses. We emphasize that the diagram denotes the propagation of the displacement Stokes tensor $\langle u_j u_j^*\rangle$. Both the release of energy by the source and the measurement carried out at the detector can be described by some linear operation of this Stokes tensor which will be specified in the following.

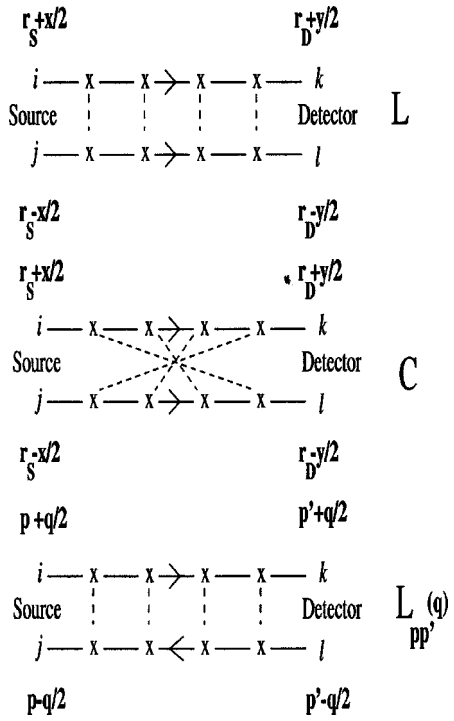


FIG. 1. Diagrammatic representations of incoherent propagation of elastic displacements (top), and interference between time-reversed wave (middle). Crosses denote scatterers, a solid line denotes the Green's function of the elastic wave equation. Dashed lines connect identical scatterers. The bottom diagram denotes the incoherent propagation in Fourier space.

The basic physics of the coherent backscattering is the interference between time reversed paths. If the sequence of scattering events, which starts at the source and ends at the detector is time reversed, the diagram 1(b) is obtained, representing the vertex C. The reciprocity principle gives a very simple relation between both vertices,[1]

$$C_{ijkl}(t,\mathbf{r}_1,\mathbf{r}_2\rightarrow\mathbf{r}_3,\mathbf{r}_4)=L_{ilkj}(t,\mathbf{r}_1,\mathbf{r}_4\rightarrow\mathbf{r}_3,\mathbf{r}_2), \tag{5}$$

i.e., polarization labels and positions of bottom line simply interchange.

The interference contribution C is not contained in classical radiative transfer, but will significantly modify the energy coming back to the source. This can best be seen for the simplified case of a measurement of $\langle|u_k(\mathbf{r}_d)|^2\rangle$, which is proportional to the average kinetic elastic energy at the detector, and a genuine monopole source $f_i(\omega)\delta(\mathbf{r}-\mathbf{r}_s)$. The last property means that the source emits the wave field $\mathbf{G}(\omega,\mathbf{r}-\mathbf{r}_s)\cdot\mathbf{f}(\omega)$ and not the one described by Eq. (3). In that case, L_{iikk} and C_{iikk} correspond *exactly* to the observation. Equation (5) implies

$$C_{iikk}(t,\mathbf{r}_s,\mathbf{r}_s\rightarrow\mathbf{r}_d,\mathbf{r}_d)=L_{ikki}(t,\mathbf{r}_s,\mathbf{r}_d\rightarrow\mathbf{r}_d,\mathbf{r}_s). \tag{6}$$

This relation means that $C=L$ at all times for detection at the source ($\mathbf{r}_s=\mathbf{r}_d$) provided that the displacement component is measured along the polarization of the source ($i=k$). Since C may be expected to vanish away from the source, we would find a local enhancement of kinetic energy of *exactly* a factor of 2. In addition, at very long times we may expect incoherent radiation to be completely depolarized, $L_{ikki}(t)\sim\delta_{ik}$. This means that the interference C vanishes for any component orthogonal to the polarization of the source. This would be very useful to determining the polarization of a seismic source with waves in the regime of seismic coda, where multiple scattering is believed to occur.

A. Coherent backscattering near a seismic source

The above-mentioned arguments have been shown to apply to experiments carried out in the far field for both light[3] and acoustic waves,[11] where the factor of 2 was confirmed. For light also the absence of intensity enhancement in the orthogonal polarization channel has been established. The seismic situation is more complicated in view of the aspects of the source and the near field detection.

We restrict our analysis to elastic wave diffusion in an *infinite* medium, whose translational symmetry facilitates a study of the vertex L in the Fourier space. Evidently, the study with plane waves is a natural choice in optics.[1,6] The convention is shown in Fig. 1(c), introducing the four momenta $\mathbf{p}\pm\frac{1}{2}\mathbf{q}$ and $\mathbf{p}'\pm\frac{1}{2}\mathbf{q}$. Note the bottom line that travels in the opposite direction since it represents the complex conjugate of the displacement vector, so that the law of momentum conservation is obeyed, as required by translational symmetry. At times longer than the mean free time, the four-rank tensor L is dominated by its second-rank eigenfunction S_{ij} that corresponds to the eigenvalue closest to zero. This eigenvalue is given by $-Dq^2+1/\tau_a$ with D the diffusion constant and τ_a allowing for some small dissipation. (See Refs. 1 and 31 for more technical detail.) We get for the vertex L in the momentum representation,

$$L_{ijkl,\mathbf{pp}'}(t,\mathbf{q})\rightarrow e^{-Dq^2t-t/\tau_a}S_{ij}(\mathbf{p},\mathbf{q})S^*_{lk}(\mathbf{p}',\mathbf{q}). \tag{7}$$

The symmetric form of the vertex, which is invariant with respect to interchanging the source and the detector, is a manifestation of the reciprocity principle. The tensor $S_{ij}(\mathbf{p},\mathbf{q})$ determines the polarization of the diffuse field propagating in the direction $\mathbf{p}$, and fixes the elastic Stokes parameters completely.[22,23] It is given by[31]

$$S_{ij}(\mathbf{p},\mathbf{q})=G_{ij}(\mathbf{p}+\tfrac{1}{2}\mathbf{q})-G^*_{ji}(\mathbf{p}-\tfrac{1}{2}\mathbf{q}) \tag{8}$$

in terms of the elastic Green's tensor defined in Eq. (4), the asterisk denoting the complex conjugate.[32]

For our purposes we need the real space formulation, which can be obtained by the Fourier transformation,

$$\begin{aligned}
&L_{ijkl}(t,\mathbf{r}_1,\mathbf{r}_2\rightarrow\mathbf{r}_3,\mathbf{r}_4)\\
&\quad=\int\frac{d\mathbf{p}}{(2\pi)^3}\int\frac{d\mathbf{p}'}{(2\pi)^3}\int\frac{d\mathbf{q}}{(2\pi)^3}\\
&\qquad\times e^{i(\mathbf{p}+\mathbf{q}/2)\cdot\mathbf{r}_1}e^{-i(\mathbf{p}-\mathbf{q}/2)\cdot\mathbf{r}_2}e^{-i(\mathbf{p}'+\mathbf{q}/2)\cdot\mathbf{r}_3}\\
&\qquad\times e^{i(\mathbf{p}'-\mathbf{q}/2)\cdot\mathbf{r}_4}e^{-Dq^2t-t/\tau_a}S_{ij}(\mathbf{p},\mathbf{q})S^*_{lk}(\mathbf{p}',\mathbf{q})\\
&\quad\sim\frac{e^{-t/\tau_a}}{(Dt)^{3/2}}\int\frac{d\mathbf{p}}{(2\pi)^3}\int\frac{d\mathbf{p}'}{(2\pi)^3}e^{i\mathbf{p}\cdot\mathbf{r}_1}\\
&\qquad\times e^{-i\mathbf{p}\cdot\mathbf{r}_2}e^{-i\mathbf{p}'\cdot\mathbf{r}_3}e^{i\mathbf{p}'\cdot\mathbf{r}_4}\,\mathrm{Im}\,G_{ij}(\mathbf{p})\,\mathrm{Im}\,G_{lk}(\mathbf{p}')\\
&\quad=\frac{e^{-t/\tau_a}}{(Dt)^{3/2}}\,\mathrm{Im}\,G_{ij}(\mathbf{r}_1-\mathbf{r}_2)\,\mathrm{Im}\,G_{lk}(\mathbf{r}_4-\mathbf{r}_3).
\end{aligned} \tag{9}$$

The approximation carried out in the second line holds for large lapse times, where typically $q \approx 0$. The factor $e^{-t/\eta_a}/(Dt)^{3/2}$ is reminiscent of a diffusion process in three dimensions. In the rest of this paper we neglect absorption. The vertex C can be easily constructed using the reciprocity relation (5),

$$C_{ijkl}(t,\mathbf{r}_1,\mathbf{r}_2 \rightarrow \mathbf{r}_3,\mathbf{r}_4) = \frac{1}{(Dt)^{3/2}} \operatorname{Im} G_{il}(\mathbf{r}_1-\mathbf{r}_4) \operatorname{Im} G_{jk}(\mathbf{r}_2-\mathbf{r}_3). \tag{10}$$

This tensor describes the displacement correlation function $\langle u_k(\mathbf{r}_3) u_l^*(\mathbf{r}_4)\rangle$ which is due to constructive interference, for an arbitrary source. It is nonzero only near the source ($\mathbf{r}_1 \approx \mathbf{r}_4$ and $\mathbf{r}_3 \approx \mathbf{r}_2$). This analysis confirms the conclusion, which was reached previously for acoustic waves, that the line profile C/L around the source is *independent* of time.[24] The characteristic length in the line profile $\operatorname{Im} G_{il}(\mathbf{x})$ is typically the wavelength of compressional or shear waves, i.e., in the near field the phenomenon of coherent backscattering is diffraction limited. For a given distance x from the source, the relative enhancement factor C/L is highly dependent on frequency, typically proportional to $\sin^2(\omega x/\alpha)/x^2\omega^2$. This specific near-field feature may facilitate its observation.

In the following sections L_{kl} and C_{kl} are, respectively, the incoherent and coherent contribution to the displacement correlation function $\langle u_k u_l^*\rangle$ as measured at or near the source.

B. Enhancement of kinetic energy

Seismic studies address the three components of the displacement field $\mathbf{u}(t)$. Let us assume that the detection process produces the correlation functions $\langle u_k(\mathbf{r},t) u_l^*(\mathbf{r},t)\rangle$ between two displacement vectors at the same point. For $k=l$ this is the kinetic elastic energy density, proportional to $\omega^2\mathbf{u}^2$. For the source described in Eq. (3), incoherent and coherent contributions to $\langle u_k u_l^*\rangle$ at the detector may be obtained from Eqs. (9) and (10), respectively, by carrying out the spatial differentiation,

$$\begin{aligned} L_{kl}(\omega,\mathbf{r}_s,\mathbf{r}_d) &\sim \frac{-\omega^2}{(Dt)^{3/2}} \operatorname{Im} G_{kl}(\omega,\mathbf{0}) \\ &\quad \times M_{ni}(\omega) M_{mj}(\omega) \partial_m \partial_n \\ &\quad \times \operatorname{Im} G_{ij}(\omega,\mathbf{x}=0), \\ C_{kl}(\omega,\mathbf{r}_s,\mathbf{r}_d) &\sim \frac{\omega^2}{(Dt)^{3/2}} M_{mi}(\omega) M_{nj}(\omega) \\ &\quad \times \partial_m \operatorname{Im} G_{il}(\omega,\mathbf{x}) \partial_n \operatorname{Im} G_{jk}(\omega,\mathbf{x}). \end{aligned} \tag{11}$$

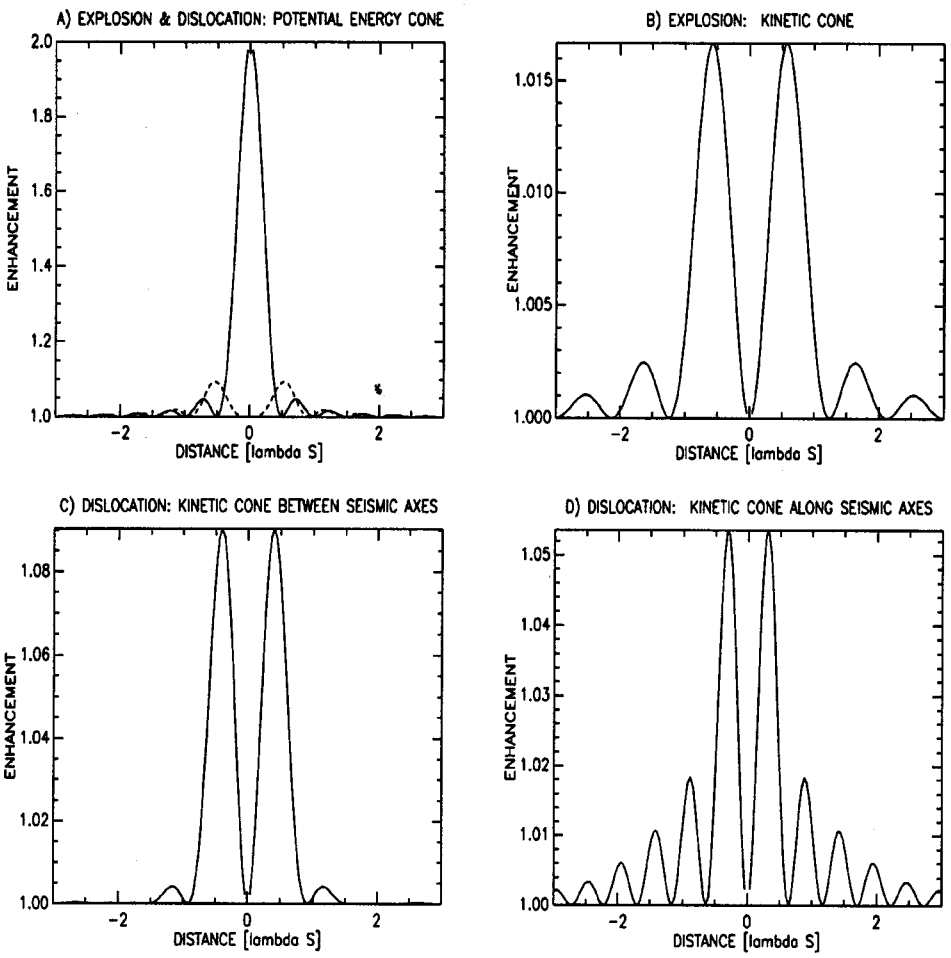


FIG. 2. Enhancement profiles as a function of the distance from the source, in units of the shear wave length, for an infinite Poissonian medium. (a) The enhancement of the potential energy for an explosion (solid) and a dislocation (dashed). (b) Enhancement profile for the kinetic energy near an explosion. (c) and (d) Enhancement profiles for the kinetic energy between and along the seismic axes of a dislocation source.

As usual, the implicit summation over the repeating indices is assumed, and $\mathbf{x}=\mathbf{r}_d-\mathbf{r}_s$ denotes this distance. The incoherent background L_{kl} is independent of the distance of the source. Explicitly, we have

$$L_{kl}(\omega,t)=\delta_{kl}\frac{\omega^6}{48\pi^2(Dt)^{3/2}}\left(\frac{1}{\alpha^3}+\frac{2}{\beta^3}\right)\left[\left(\frac{1}{\alpha^5}-\frac{1}{\beta^5}\right)\times\frac{2\,\mathrm{Tr}\,\mathbf{M}^2+(\mathrm{Tr}\,\mathbf{M})^2}{15}+\frac{1}{3\beta^5}\mathrm{Tr}\,\mathbf{M}^2\right]. \tag{12}$$

The factor δ_{kl} implies that the orthogonal displacements are decorrelated by high orders in multiple scattering. For $\alpha>\beta$, true for a Poissonian solid, the incoherent background is dominated by shear waves, that is the β terms in Eq. (12).

The coherent background C_{kl} decays as an oscillating power law away from the source. Somewhat surprisingly, we find that $C_{k=l}(\mathbf{x}=0)=0$, that is at the source we have a *destructive* interference. This is not a violation of the reciprocity principle, since the different operations performed by the source [this operation is specified in Eq. (3) and involves a spatial derivative] and the detector (the detector is supposed to measure the displacement itself, and thus no spatial derivative is involved) break their mutual symmetry.

We proceed by considering the coherent enhancement of kinetic energy cone in two special cases.

1. Explosion

An isotropic explosion is characterized by a seismic moment tensor $M_{ij}=M(\omega)\,\delta_{ij}$. Using Eq. (11) we find

$$C_{kl}(\omega,\mathbf{x},t)=\hat{r}_k\hat{r}_l\frac{M(\omega)^2\omega^6}{16\pi^2(Dt)^{3/2}\alpha^8}j_1^2\left(\frac{\omega x}{\alpha}\right),$$
$$L_{kl}(\omega,\mathbf{x},t)=\delta_{kl}\frac{M(\omega)^2\omega^6}{48\pi^2(Dt)^{3/2}\alpha^5}\left(\frac{1}{\alpha^3}+\frac{2}{\beta^3}\right). \tag{13}$$

The coherent profile vanishes at $x=0$, i.e., at the source, and has its first maxima on a shell at $x\approx\frac{1}{3}\lambda_P$ away from the source. Only radial, longitudinal P waves contribute to the coherent background whereas the incoherent background is dominated by shear waves. This lowers the maximal enhancement factor of the total kinetic energy to only $0.1(\beta/\alpha)^3\approx 0.017$. Figure 2(b) shows the line profile for the kinetic energy around an explosion. Therefore, an observation of enhanced backscattering of kinetic energy, with an explosion source, is rather unlikely; these measurements will be dominated by statistical fluctuations.

2. Dislocation

A dislocation source may be defined by a slip vector along the x axis and a plane normal to the z axis.[27] It offers a generally accepted model for an earthquake. Its seismic moment tensor is $M_{ij}=M(\omega)(\hat{x}_i\hat{y}_j+\hat{y}_i\hat{x}_j)$. The xyz frame defines the azimuthal angle ϕ in the nodal plane and the poloidal angle θ in the usual way. Starting with Eq. (11), a long but straightforward calculation gives C_{kk} and L_{kk} (summed over all components k) near the source, as

$$C_{kk}(\omega,\mathbf{x},t)=\frac{M(\omega)^2\omega^4\sin^2\theta}{16\pi^2(Dt)^{3/2}\beta^8}[S_1(\omega,x)^2\sin^2\theta\sin^2 2\phi+S_2(\omega,x)^2(1-\sin^2\theta\sin^2 2\phi)],$$
$$L_{kk}(\omega,\mathbf{x},t)=\frac{M(\omega)^2\omega^4}{16\pi^2(Dt)^{3/2}}\left(\frac{1}{\alpha^3}+\frac{2}{\beta^3}\right)\left(\frac{2}{15\alpha^5}+\frac{8}{15\beta^5}\right) \tag{14}$$

with the two line-shape functions,

$$S_1(\omega,x)=-\frac{9}{y_s^4}[y_s\cos y_s-\sin y_s-y_p\cos y_p+\sin y_p]+\frac{4}{y_s^2}\left(\frac{\beta}{\alpha}\right)^2\sin y_p-\frac{3}{y_s^2}\sin y_s-\frac{1}{y_s}\left(\frac{\beta}{\alpha}\right)^3\cos y_p,$$
$$S_2(\omega,x)=+\frac{6}{y_s^4}[y_s\cos y_s-\sin y_s-y_p\cos y_p+\sin y_p]-\frac{2}{y_s^2}\left(\frac{\beta}{\alpha}\right)^2\sin y_p+\frac{3}{y_s^2}\sin y_s-\frac{1}{y_s}\cos y_s, \tag{15}$$

given in terms of the dimensionless distances $y_s\equiv\omega x/\beta$ and $y_p\equiv\omega x/\alpha$.

Part of the double-couple structure of the source may be recognized from the angular dependence of the line profiles, which are different along and between the two seismic axes of the dislocation (i.e., the two axes in its nodal plane with respect to which the seismic moment tensor is diagonal), and vanish perpendicular ($\theta=0$) to the nodal plane. The two top figures in Fig. 3 show two different views of the enhancement of kinetic energy in the seismic plane ($\theta=\pi/2$). Four major maxima may be recognized between the seismic axes. Despite the fact that both coherent and incoherent radiation are dominated by shear waves, the enhancement factor still does not exceed 10%, and vanishes at the source [see Fig. 2(d)].

C. Enhancement of potential energy

The potential energy density of an elastic plane wave $\mathbf{u}$ in an infinite medium is proportional to $\alpha^2(\mathrm{div}\,\mathbf{u})^2+\beta^2(\mathrm{rot}\,\mathbf{u})^2$. The first term represents the energy of the longitudinal P waves and the second, of the transverse S waves. In this section we will calculate the coherent enhancement profiles of both terms.

1. Enhancement of compressional energy

The compressional energy is proportional to $\alpha^2(\mathrm{div}\,\mathbf{u})^2$. Its coherent enhancement near a source with the seismic moment tensor $M_{ij}(\omega)$ may be calculated from the vertex $C_{ijkl}(t,\mathbf{r}_1,\mathbf{r}_2\rightarrow\mathbf{r}_3,\mathbf{r}_4)$ given in Eq. (10), as follows:

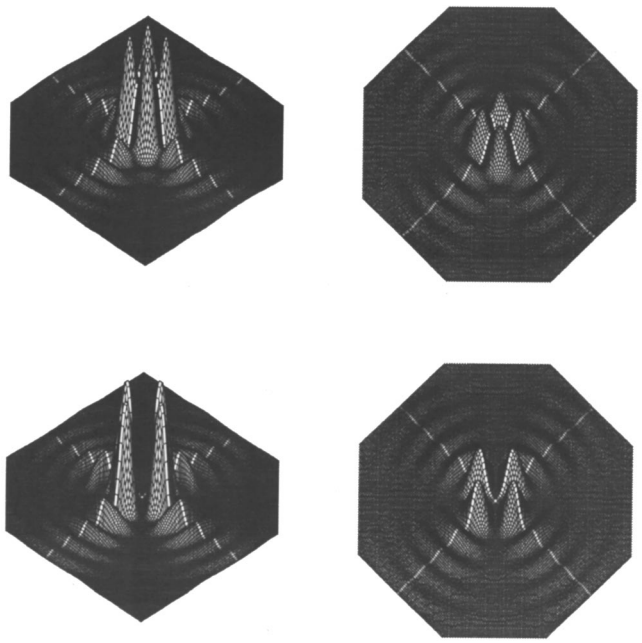


FIG. 3. Enhancement profiles of kinetic energy (top) and potential energy (bottom) in the seismic plane of a dislocation, from two different views. Distance is in units of the shear wavelength. The white axes denote the two seismic axes.

$$
\begin{aligned}
C_p &= \alpha^2 M_{ni}(\omega) M_{mj}(\omega) \partial_n^{(1)} \partial_m^{(2)} \partial_k^{(3)} \partial_l^{(4)} C_{ijkl}(t, \mathbf{r}_1, \mathbf{r}_2 \rightarrow \mathbf{r}_3, \mathbf{r}_4) \\
&= \frac{1}{(Dt)^{3/2}} M_{ni}(\omega) M_{mj}(\omega) \partial_n \partial_l \operatorname{Im} G_{il}(\omega, \mathbf{x}) \partial_m \partial_k \\
&\quad \times \operatorname{Im} G_{jk}(\omega, \mathbf{x}) \\
&= \frac{\omega^6}{16\pi^2 (Dt)^{3/2} \alpha^8} \left[j_2(y_p) \hat{\mathbf{x}} \cdot \mathbf{M}(\omega) \cdot \hat{\mathbf{x}} - \frac{j_1(y_p)}{y_p} \operatorname{Tr} \mathbf{M}(\omega) \right]^2 .
\end{aligned}
\tag{16}
$$

We recall that $y_p = \omega x / \alpha$. Similarly, the incoherent background L becomes

$$
\begin{aligned}
L_p(\omega, \mathbf{x}, t) = \frac{\omega^6}{16\pi^2 (Dt)^{3/2} \alpha^3} &\left[\left(\frac{1}{\alpha^5} - \frac{1}{\beta^5} \right) \right. \\
&\left. \times \frac{2 \operatorname{Tr} \mathbf{M}^2 + (\operatorname{Tr} \mathbf{M})^2}{15} + \frac{1}{3\beta^5} \operatorname{Tr} \mathbf{M}^2 \right].
\end{aligned}
\tag{17}
$$

For an isotropic explosion $M_{ij} = M(\omega) \delta_{ij}$ the enhancement ratio simplifies to

$$
\frac{C_p}{L_P} = j_0(y_p)^2 . \tag{18}
$$

This result is exactly the same as for an acoustic medium and a monopole source,[24] including the factor of 2 enhancement over the background. When measuring the compressional energy, the symmetry between the explosion source and detector is restored, so that the reciprocity principle applies again [see Fig. 2(a)].

For a dislocation source $\operatorname{Tr} \mathbf{M} = 0$ the enhancement of compressional energy becomes

$$
\frac{C_P}{L_P} \approx \frac{15}{16} \left(\frac{\beta}{\alpha} \right)^5 \sin^4 \theta \sin^2 2\phi j_2(y_p)^2 , \tag{19}
$$

where the azimuthal angle ϕ is defined in the seismic plane. The operation of measuring the compressional energy (i.e., $(\operatorname{div} \mathbf{u})^2$) yields a very small value when applied to the double-couple source. This is due to the fact that the dislocation sources release most of their energy as shear waves. Hence, the source and detector are highly "orthogonal" (in a sense that will be made explicit in Sec. IV) so that the enhancement vanishes rapidly (as x^4) near the source. The small factor $(\beta/\alpha)^5$ in Eq. (19) expresses the efficient $P-S$ mode conversion, which seriously hampers the enhancement of compressional energy near a dislocation source.

2. Enhancement of shear energy

The operation of measuring the shear energy is $\beta^2 \epsilon_{ijk} \epsilon_{inl} \partial_j u_k \partial_n u_l^*$, with ϵ_{ijk} the antisymmetric Levi-Cevitta tensor. Together with the operation carried out by the source, coherent and incoherent contributions to the shear energy at the detector become, respectively,

$$
\begin{aligned}
C_S &= \beta^2 M_{ni}(\omega) M_{mj}(\omega) \epsilon_{pqk} \epsilon_{prl} \partial_q^{(3)} \partial_r^{(4)} \partial_n^{(1)} \partial_m^{(2)} \\
&\quad \times C_{ijkl}(t, \mathbf{r}_1, \mathbf{r}_2 \rightarrow \mathbf{r}_3, \mathbf{r}_4) \\
&= \frac{\omega^6}{16\pi^2 (Dt)^{3/2} \beta^8} j_2(y_s)^2 |\hat{\mathbf{x}} \times [\mathbf{M}(\omega) \cdot \hat{\mathbf{x}}]|^2 , \\
L_S &= \frac{\omega^6}{8\pi^2 (Dt)^{3/2} \beta^3} \left[\left(\frac{1}{\alpha^5} - \frac{1}{\beta^5} \right) \frac{2 \operatorname{Tr} \mathbf{M}^2 + (\operatorname{Tr} \mathbf{M})^2}{15} \right. \\
&\quad \left. + \frac{1}{3\beta^5} \operatorname{Tr} \mathbf{M}^2 \right].
\end{aligned}
\tag{20}
$$

Not unexpectedly for an explosion the coherent enhancement vanishes entirely. For a dislocation source it vanishes only near the source (as x^4). The enhancement of shear energy does not suffer from mode conversions but is only 0.09 at the first maximum located at $1/2\lambda_s$ on the seismic axis [see Fig. 2(a)]. In Fig. 3 (bottom) we display the enhancement factor of the total potential energy $S+P$ in the seismic plane which, in a Poissonian elastic medium, is largely dominated by the shear energy.

IV. RECOVERING THE FACTOR OF 2

In most cases that we have studied, the enhancement factor vanishes exactly at the source. Our explanation is that the operation carried out at the detector has a symmetry which is different to the source.

To clarify this point, we investigate if there exists a measurement of the type $N_{nj} \partial_n u_j$ that restores the factor of 2 familiar from the optical and acoustic studies. In view of Eq. (3), the choice of $\mathbf{N} = \mathbf{M}$, i.e., an operation equal to the seismic moment tensor of the source, seems to be a good candidate, since it restores the symmetry between source and detector. The second point that we have to establish is uniqueness of this choice.

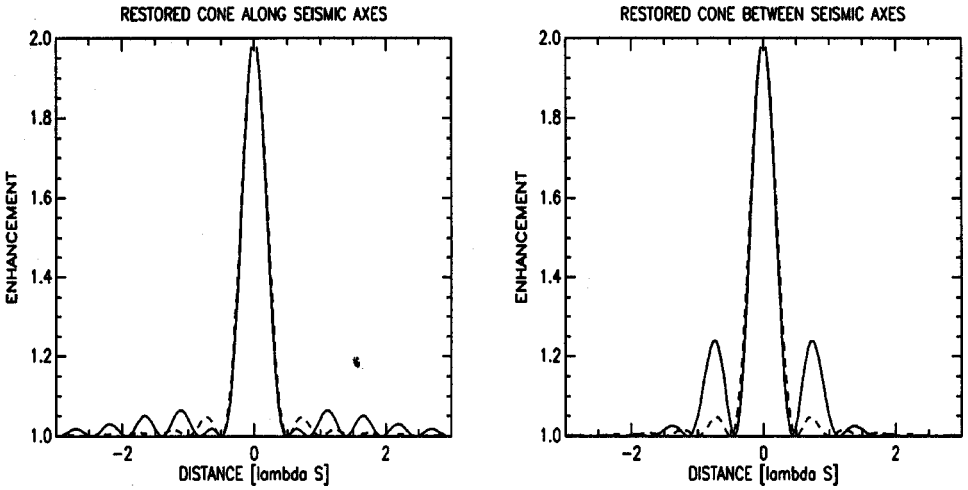


FIG. 4. Restoration of enhanced backscattering by carrying out exactly the same operation at the detector as at the dislocation source. Dashed line denotes the line profile for an acoustic source in an infinite medium. Distance is in units of the shear wavelength.

Upon measuring $N_{nj}\partial_n u_j$, the coherent enhancement becomes

$$C = M_{ni}M_{mj}N_{pk}N_{ql}\partial_n^{(1)}\partial_m^{(2)}\partial_p^{(3)}\partial_q^{(4)}C_{ijkl}(t,\mathbf{r}_1,\mathbf{r}_2\rightarrow\mathbf{r}_3,\mathbf{r}_4)$$

$$= \frac{1}{(Dt)^{3/2}}\times(M_{ni}N_{ql}\partial_n\partial_q \operatorname{Im} G_{il}(\mathbf{x}))^2. \tag{21a}$$

For the incoherent background the same operation yields

$$L = \frac{1}{(Dt)^{3/2}} M_{ni}M_{mj}\partial_n\partial_m \operatorname{Im} G_{ij}(\mathbf{x}=0)$$

$$\times N_{pk}N_{ql}\partial_p\partial_q \operatorname{Im} G_{kl}(\mathbf{x}=0). \tag{21b}$$

From Eq. (21) we conclude that if $\mathbf{M}=\mathbf{N}$ then $C(\mathbf{x}=0)=L$. This confirms our conjecture that the enhancement factor is fully restored if the detector carries out exactly the same operation as the source. The uniqueness of this choice for a dislocation source among all possible choices of $\mathbf{N}$ with a vanishing trace may be readily demonstrated. Using the Green's tensor (4) for an homogeneous isotropic elastic medium and Eq. (21), the enhancement factor becomes

$$\frac{C(\mathbf{x}=0)}{L} = \frac{(\operatorname{Tr}\mathbf{M}\cdot\mathbf{N})^2}{\operatorname{Tr}\mathbf{M}^2\operatorname{Tr}\mathbf{N}^2}. \tag{22}$$

The Cauchy inequality implies that $C(\mathbf{0})/L \leqslant 1$, with the equality holding *if and only if* $\mathbf{M}=\mathbf{N}$. This conclusion allows us to recover the seismic moment tensor by optimizing the enhancement near the source. Equation (22) also shows explicitly that the orthogonality of operations at source and detector leads to vanishing enhancement at the source.

Our final task is to calculate the line profile corresponding to the ideal choice. The scalar function $M_{ni}M_{mj}\partial_n\partial_m G_{ij}(\mathbf{x})$ for a general seismic moment tensor is presented in the Appendix. In Fig. 4 we show the restored line profile for a dislocation source, and compare it to the enhancement factor (18) which has been calculated for an explosion. Near the source they look very similar, but the "restored" profile is still anisotropic, with four significant secondary maxima at $x=0.7\lambda_s$ in between the seismic axes of the nodal plane.

V. CONCLUSION

We have calculated the spatial line profiles of the coherent enhancement factors of various elastic energies near the source. We have emphasized the seismic aspects of this elastic problem, in particular the specificity of the near-field detection. The first consequence of the near-field detection is the typical decay distance of the enhancement, which equals the elastic wavelength(s) and not the mean free path, as for the far-field detection. Second, the near field has a huge impact on the precise line profiles of the different measurable elastic energies. The operation performed at the detector (measuring kinetic energy, potential energy, etc.) as well as the precise symmetry of the source al play a crucial role. We have shown that *one unique* operation carried out at the detector shows the maximal possible enhancement of two. The enhancement of kinetic and potential energy vanishes at the source and does not exceed 10%. It is also hampered by mode conversions in multiple scattering. Only for an explosion source and a measurement of compressional energy is an enhancement of two is expected.

ACKNOWLEDGMENTS

We thank Julien de Rosny, Nicolas Trégourès, Mathias Fink, and Renaud Hennino for helpful discussions.

APPENDIX: LINE PROFILE OF RESTORED ENHANCEMENT

The restored line profile is given by Eq. (21a), in which figures the scalar function $M_{ni}M_{nj}\partial_n\partial_m \operatorname{Im} G_{ij}(\mathbf{x})$, which we state here without derivation for the case of a pure dislocation that obeys $\operatorname{Tr}\mathbf{M}=0$. Using the notation $y_s=\omega x/\beta$, $y_p=\omega x/\alpha$, and $R=\alpha/\beta$,

$$
\begin{aligned}
&\frac{1}{\omega^3\alpha^5} M_{ni} M_{nj} \partial_n \partial_m \operatorname{Im} G_{ij}(\mathbf{x}) \\
&\quad = 3(10\hat{\mathbf{x}}\cdot\mathbf{M}^2\cdot\hat{\mathbf{x}} - 10(\hat{\mathbf{x}}\cdot\mathbf{M}\cdot\hat{\mathbf{x}})^2 - 2)\left(\frac{\cos y_p}{y_p^4} - R\frac{\cos y_s}{y_p^4} + \frac{\sin y_s - \sin y_p}{y_s^3}\right) \\
&\qquad - 3(2\hat{\mathbf{x}}\cdot\mathbf{M}^2\cdot\hat{\mathbf{x}} - 5(\hat{\mathbf{x}}\cdot\mathbf{M}\cdot\hat{\mathbf{x}})^2)\left[\sin y_p\left(\frac{5}{y_p^5} - \frac{1}{y_p^3}\right) - \sin y_s\left(\frac{5}{y_p^5} - \frac{R^2}{y_p^3}\right) - \frac{5}{y_p^4}\cos y_p + \frac{5R}{y_p^4}\cos y_s\right] \\
&\qquad + 2(6\hat{\mathbf{x}}\cdot\mathbf{M}^2\cdot\hat{\mathbf{x}} - 6(\hat{\mathbf{x}}\cdot\mathbf{M}\cdot\hat{\mathbf{x}})^2 - 1)\frac{\sin y_p}{y_p^3} - 2(\hat{\mathbf{x}}\cdot\mathbf{M}^2\cdot\hat{\mathbf{x}} - 3(\hat{\mathbf{x}}\cdot\mathbf{M}\cdot\hat{\mathbf{x}})^2)\left(\frac{\cos y_p}{y_p^2} - \frac{3\sin y_p}{y_p^3}\right) \\
&\qquad - 3(4\hat{\mathbf{x}}\cdot\mathbf{M}^2\cdot\hat{\mathbf{x}} - 4(\hat{\mathbf{x}}\cdot\mathbf{M}\cdot\hat{\mathbf{x}})^2 - 1)R^2\frac{\sin y_s}{y_p^3} + 3(\hat{\mathbf{x}}\cdot\mathbf{M}^2\cdot\hat{\mathbf{x}} - 2(\hat{\mathbf{x}}\cdot\mathbf{M}\cdot\hat{\mathbf{x}})^2)\left(R^3\frac{\cos y_s}{y_p^2} - 3R^2\frac{\sin y_s}{y_p^3}\right) \\
&\qquad + (2\hat{\mathbf{x}}\mathbf{M}^2\cdot\hat{\mathbf{x}} - 2(\hat{\mathbf{x}}\cdot\mathbf{M}\cdot\hat{\mathbf{x}})^2 - 1)R^3\frac{\cos y_s}{y_p^2}(\hat{\mathbf{x}}\cdot\mathbf{M}^2\cdot\hat{\mathbf{x}} - (\hat{\mathbf{x}}\cdot\mathbf{M}\cdot\hat{\mathbf{x}})^2)\left(R^4\frac{\sin y_s}{y_p} + 2R^3\frac{\cos y_s}{y_p^2} - 2\frac{\cos y_p}{y_p^2}\right) \\
&\qquad \times(\hat{\mathbf{x}}\cdot\mathbf{M}\cdot\hat{\mathbf{x}})^2\left(\frac{\sin y_p}{y_p} + 2\frac{\cos y_p}{y_p^2}\right). \qquad \text{(A1)}
\end{aligned}
$$

[1] B. A. van Tiggelen and R. Maynard, ''Reciprocity and Coherent Backscattering of Light,'' in *Wave Propagation in Complex Media*, edited by G. Papanicolaou (Springer, New York, 1998).

[2] B. L. Altschuler, A. G. Aronov, and B. Z. Spivak, ''The Aharonov–Bohm effect in disordered conductors,'' JETP Lett. **33**, 94–97 (1981); D. Yu. Sharvin, and Yu. V. Sharvin, ''Magnetic flux quantization in a cylindrical film,'' *ibid.* **34**, 272–275 (1981).

[3] Y. Kuga and A. Ishimaru, ''Retroreflection from a dense distribution of spherical particles,'' J. Opt. Soc. Am. A **1**, 831–835 (1984); M. P. van Albada and A. Lagendijk, ''Observation of weak localization of light in a random medium,'' Phys. Rev. Lett. **55**, 2692–2695 (1985); P. E. Wolf and G. Maret, ''Weak localization and coherent backscattering of photons in disordered media,'' *ibid.* **55**, 2696–2699 (1985); M. Kaveh, M. Rosenbluh, I. Edrei, and I. Freund, ''Weak localization and light scattering from disordered solids,'' *ibid.* **57**, 2049–2052 (1986).

[4] *New Aspects of Eletromagnetic and Acoustic Wave Diffusion*, edited by POAN Research Group [Springer-Tracts Mod. Phys. **144**, 17–21 (1998)].

[5] F. A. Erbacher, R. Lenke, and G. Maret, ''Multiple light scattering in magneto-optically active media,'' Europhys. Lett. **21**, 551–556 (1993); D. Lacoste and B. A. van Tiggelen, ''Coherent backscattering in a magnetic field,'' Phys. Rev. E **61**, 4556–4565 (2000).

[6] F. C. MacKintosh and S. John, ''Coherent backscattering of light in the presence of time-reversal noninvariant and parity nonconserving media,'' Phys. Rev. B **37**, 1884–1897 (1988).

[7] E. Akkermans, P. E. Wolf, and R. Maynard, ''Coherent backscattering of light by disordered media: Analysis of peak-line shape,'' Phys. Rev. Lett. **56**, 1471–1473 (1986).

[8] D. S. Wiersma, M. P. van Albada, B. A. van Tiggelen, and A. Lagendijk, ''Experimental evidence for the occurrence of recurrent scattering of light in diffusion,'' Phys. Rev. Lett. **74**, 4193–4196 (1995).

[9] D. S. Wiersma, M. P. van Albada, and A. Lagendijk, ''Coherent backscattering of light from amplifying random media,'' Phys. Rev. Lett. **75**, 1739–1742 (1995).

[10] G. Labeyrie, F. de Tomasi, J.-C. Bernard, C. A. Müller, C. Miniatura, and R. Kaiser, ''Coherent backscattering of light by cold atoms,'' Phys. Rev. Lett. **83**, 5266–5269 (1999).

[11] G. Bayer and T. Niederdränk, ''Weak localization of acoustic waves in strongly scattering media,'' Phys. Rev. Lett. **70**, 3884–3887 (1993); A. Tourin, Ph. Roux, A. Derode, B. A. van Tiggelen, and M. Fink, ''Time-dependent coherent backscattering of acoustic waves,'' *ibid.* **79**, 3637–3639 (1997); K. Sakai, K. Yamamoto, and K. Takagi, ''Observation of acoustic coherent backscattering,'' Phys. Rev. B **56**, 10930–10933 (1997).

[12] K. Aki, ''Analysis of the seismic coda of local earthquakes as scattered waves,'' J. Geophys. Res. **74**, 615–631 (1969).

[13] K. Aki and B. Chouet, ''Origin of coda waves, source, attenuation and scattering effects,'' J. Geophys. Res. **80**, 3322–3342 (1975).

[14] H. Sato and M. C. Fehler, *Seismic Wave Propagation and Scattering in the Heterogeneous Earth* (Springer, Heidelberg, 1995).

[15] M. Herraiz and A. F. Spinoza, ''Coda waves: A review,'' Pure Appl. Geophys. **125**, 499–577 (1987).

[16] I. R. Abubakirov and A. A. Gusev, ''Estimation of scattering properties of the lithosphere of Kamchatka, based on Monte-Carlo simulations of record envelopes of a near earthquake,'' Phys. Earth Planet Inter. **64**, 52–67 (1990).

[17] M. Hoshiba, ''Simulation of multiple scattered coda wave excitations based on the Energy Conservation Law,'' Phys. Earth Planet. Inter. **67**, 126–136 (1991).

[18] H. Sato, ''Multiple isotropic scattering model including P–S conversions for the seismogram envelope formation,'' Geophys. J. Int. **117**, 487–494 (1994).

[19] R. S. Wu and K. Aki, ''Multiple scattering and energy transfer of seismic waves: Separation of scattering effects from intrinsic attenuation,'' Pure Appl. Geophys. **128**, 49–80 (1988).

[20] L. Margerin, M. Campillo, N. M. Shapiro, and B. A. van Tiggelen, ''Residence time of diffuse waves in the crust as a physical interpretation of coda Q,'' Geophys. J. Int. **138**, 343–352 (1999).

[21] R. L. Weaver, ''Diffusivity of ultrasound in polycrystals,'' J. Mech. Phys. Solids **38**, 55–86 (1990).

[22] J. A. Turner and R. Weaver, ''Radiative transfer and multiple scattering of diffuse ultrasound in polycrystalline media,'' J. Acoust. Soc. Am. **96**, 3675–3683 (1994).

[23] G. C. Papanicolaou, L. V. Ryzhik, and J. B. Keller, ''Stability of the P-to-S energy ratio in the diffuse regime,'' Bull. Seismol. Soc. Am. **86**, 1107–1115 (1996).

[24] L. Margerin, M. Campillo, and B. A. van Tiggelen ''Coherent backscattering of acoustic waves in the near field,'' Geoph. J. Int. (submitted).

[25] R. Vreeker, M. P. van Albada, R. Sprik, and A. Lagendijk, ''Femto-second time-resolved measurements of weak localization of light,'' Phys. Lett. A **132**, 51–55 (1988).

[26] R. L. Weaver, ''On diffuse waves in solid media,'' J. Acoust. Soc. Am. **71**, 1608–1609 (1982).

[27] T. Lay and T. C. Wallace, *Modern Global Seismology* (Academic, San Diego, 1995).

[28] J. de Rosny, A. Tourin, and M. Fink, ''Coherent backscattering of an elastic wave in a chaotic cavity,'' Phys. Rev. Lett. **84**, 1693–1695 (2000).

[29] R. L. Weaver and O. I. Lobkis, ''Enhanced backscattering and modal echo of reverberant elastic waves,'' Phys. Rev. Lett. **84**, 4942–4945 (2000).

[30] J. de Rosny, ''Milieux Réverbérants et Réversibilité,'' Ph.D. thesis, University of Paris, 2000.

[31] P. Sheng, *Introduction of Wave Scattering, Localization and Mesoscopic Phenomena* (Academic, San Diego, 1995).

[32] A detailed analysis (see, e.g., Ref. 31) shows that this Green's function should be the one of the effective medium, in which the elastic waves have a finite mean free path l due to scattering. This notion is irrelevant in the near field, since the wavelengths are much smaller than the mean free path.

On the character of acoustic waves at the interface between hard and soft solids and liquids

Christ Glorieux[a] and Kris Van de Rostyne
Laboratorium voor Akoestiek en Thermische Fysica, Departement Natuurkunde, Katholieke Universiteit Leuven, Celestijnenlaan 200D, B-3001 Leuven, Belgium

Keith Nelson
Department of Chemistry, Massachusetts Institute of Technology, Room 6-235, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139

Weimin Gao, Walter Lauriks, and Jan Thoen
Laboratorium voor Akoestiek en Thermische Fysica, Departement Natuurkunde, Katholieke Universiteit Leuven, Celestijnenlaan 200D, B-3001 Leuven, Belgium

(Received 5 May 2000; revised 26 October 2000; accepted 15 April 2001)

Laser ultrasonics is used to optically excite and detect acoustic waves at the interface between a liquid and a solid or coated solid. Several case studies show that this technique is feasible to investigate experimentally the theoretically predicted fundamental properties of different aspects of interface waves at liquid–solid interfaces and to characterize the elastic properties of soft solids. The theoretical prediction that the leaky Rayleigh (LR)-type root of the characteristic determinant becomes forbidden when the shear velocity of the solid lies below the bulk velocity of the liquid was experimentally confirmed. The depth profiling and nondestructive testing potential of Scholte waves was experimentally illustrated and explained by the properties of the wave displacement profile. [DOI: 10.1121/1.1396333]

PACS numbers: 43.20.Hq, 43.20.Ks, 81.70.Cv, 43.35.Bf, 43.35.Cg [DEC]

I. INTRODUCTION

Interface acoustic waves between a solid material and a liquid have been the subject of numerous studies, both theoretically[1–7] and experimentally.[8–11] Depending on the configuration used, two kinds of interface waves may propagate on a plane solid–liquid interface: the leaky Rayleigh wave and the Scholte wave (also called ''Stoneley wave'' or ''Scholte–Stoneley wave''). Both types of wave are of considerable interest in seismology, engineering, and nondestructive testing (NDT). Recently, it has been suggested[12] that both wave types always exist, for any combination of solid and liquid.

Two situations can be distinguished. A ''hard solid–fluid'' configuration refers to a fluid velocity v_{L2} smaller than the shear velocity v_{T1} of the solid material ($v_{L2}<v_{T1}<v_{L1}$, indices 1 for solid, 2 for liquid, L for longitudinal, and T for transverse). For this case, it is well established that the leaky Rayleigh wave propagates with a velocity slightly lower than v_T and is attenuated because of energy radiation into the liquid. The Scholte wave velocity is lower than the liquid velocity and, when neglecting the viscoelasticity of both liquid and solid, travels unattenuated.

For a ''soft solid–fluid'' interface, the solid shear velocity is smaller than the liquid velocity ($v_{T1}<v_{L2}<v_{L1}$), as is the case in configurations like Plexiglas–water and PVC (polyvinyl chloride)–water. We will further present an analysis of the Scholte determinant for a configuration Plexiglas/water, which reveals that the Rayleigh-type root is not a physical solution. This yields an exception to the statement in Ref. 12 that the leaky Rayleigh wave always exists.

In this paper we give an overview of theoretical aspects of leaky Rayleigh and Scholte waves at different solid–liquid interfaces and confirm the calculations with experimental evidence. Equations for the displacements caused by the interface waves in the solid and the liquid are deduced in Sec. II. In Secs. III and IV, the interface wave displacements are calculated for a configuration of glass–water and Plexiglas–water, respectively, and the theoretical results are compared with experimental signals.

The use of Scholte wave data for the characterization of material parameters is a well-established technique in seismology for the characterization of sediments.[13] In particular, the shear wave velocity of the sediment bottom can be accurately determined by applying inversion techniques to the Scholte determinant. This inversion is only possible if the Scholte wave velocity is strongly linked to the shear wave velocity,[11] which requires that the substrate is sufficiently ''soft'' compared to the liquid. This is discussed in Sec. V. The possibility of using Scholte wave dispersion for coating characterization is illustrated in Sec. VI.

II. THE SCHOLTE EQUATION

For the derivation of the Scholte determinant and the calculation of acoustic displacements and stresses in layers or half-spaces, it is usual to introduce potentials (see, e.g., Ref. 1). Alternatively, one can use the Laplace formalism.[2]

[a]On leave at Massachusetts Institute of Technology, Cambridge, MA 02139.

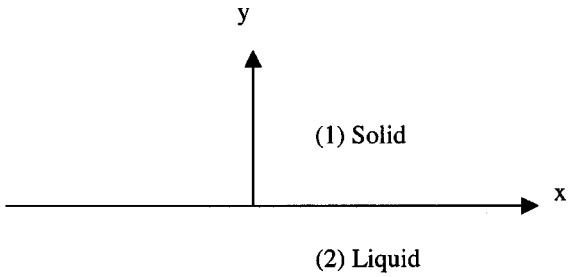


FIG. 1. Configuration used for the calculations. The waves are running in the positive x direction.

Here, we use an intermediate approach. The configuration under investigation is depicted in Fig. 1. Two elastic half-spaces $y>0$ (solid, index "1") and $y<0$ (liquid, index "2") are separated by a plane interface ($x-z$ plane). We assume no z dependence in any of the quantities and neglect viscosity. The Scholte wave at the boundary between a viscoelastic solid and an ideal fluid, modeling, e.g., a water–sediment interface has been studied extensively by Favretto-Anrès.[10,11,14] The most significant effect of viscoelasticity in the solid is the dispersion of the Scholte wave in the lower-frequency region (below 1 MHz for Plexiglas), which is of minor importance for our experiments (between 5 and 50 MHz). Under these assumptions, the displacement in the x- and y direction can be written in general as[1]

$$\begin{aligned} u_x&=(L_x\ \exp[\pm(-p_L y)]+T_x\ \exp[\pm(-p_T y)])\\ &\quad\times\exp(i\omega t-iqx),\\ u_y&=(L_y\ \exp[\pm(-p_L y)]+T_y\ \exp[\pm(-p_T y)])\\ &\quad\times\exp(i\omega t-iqx), \end{aligned} \tag{1}$$

with the upper sign for the solid and the lower sign for the liquid, T_x, T_y, L_x, L_y coefficients to be determined from the boundary conditions and $T_{x2}=T_{y2}=0$ in the liquid. Furthermore, it can be shown that (see, e.g., Ref. 2)

$$p_{L,T}=\sqrt{q^2-q_{L,T}^2},\quad \text{with}\ \ q_{L,T}=\frac{\omega}{v_{L,T}}. \tag{2}$$

The stresses σ are given by[2]

$$\begin{aligned} \frac{\sigma_{yx}(y)}{\rho_1 v_T^2}&=\left(\frac{\partial u_x}{\partial y}+\frac{\partial u_y}{\partial x}\right),\\ \frac{\sigma_{yy}(y)}{\rho_1 v_{L1}^2}&=\left(\frac{\partial u_y}{\partial y}+\left[\frac{v_{L1}^2-2v_{T1}^2}{v_{L1}^2}\right]\frac{\partial u_x}{\partial x}\right). \end{aligned} \tag{3}$$

This leads to

$$\begin{aligned} \frac{\sigma_{yx}(y)}{\rho_1 v_T^2}&=(\mp p_L L_x\ \exp(\pm(-p_L y))\\ &\quad\mp p_T T_x\ \exp(\pm(-p_T y))-iqL_y\ \exp(\pm(-p_L y))\\ &\quad-iqT_y\ \exp(\pm(-p_T y)))\exp(i\omega t-iqx),\\ \frac{\sigma_{yy}(y)}{\rho_1 v_{L1}^2}&=\Bigg(\mp p_L L_y\ \exp(\pm(-p_L y))\\ &\quad\mp p_T T_y\ \exp(\pm(-p_T y))-iq\left[\frac{v_{L1}^2-2v_{T1}^2}{v_{L1}^2}\right]L_x\\ &\quad\times\exp(\pm(-p_L y))-iq\left[\frac{v_{L1}^2-2v_{T1}^2}{v_{L1}^2}\right]T_x\\ &\quad\times\exp(\pm(-p_T y))\Bigg)\exp(i\omega t-iqx). \end{aligned} \tag{4}$$

The equations of motion are given by (propagating waves, no source)[2]

$$\begin{aligned} &\left(\frac{\partial^2}{\partial t^2}-v_L^2\frac{\partial^2}{\partial x^2}-v_T^2\frac{\partial^2}{\partial y^2}\right)u_x-(v_L^2-v_T^2)\frac{\partial^2}{\partial x\partial y}u_y=0,\\ &-(v_L^2-v_T^2)\frac{\partial^2}{\partial x\partial y}u_x+\left(\frac{\partial^2}{\partial t^2}-v_T^2\frac{\partial^2}{\partial x^2}-v_L^2\frac{\partial^2}{\partial y^2}\right)u_y=0. \end{aligned} \tag{5}$$

Inserting the x- and t-dependence for a harmonic interface wave yields

$$\begin{aligned} &\left(-\omega^2+v_L^2q^2-v_T^2\frac{\partial^2}{\partial y^2}\right)u_x+iq(v_L^2-v_T^2)\frac{\partial}{\partial y}u_y=0,\\ &iq(v_L^2-v_T^2)\frac{\partial}{\partial y}u_x+\left(-\omega^2+v_T^2q^2-v_L^2\frac{\partial^2}{\partial y^2}\right)u_y=0, \end{aligned} \tag{6}$$

which, applied to the solid at the interface, results in

$$\begin{aligned} &[(v_{L1}^2-v_{T1}^2)p_{L1}^2]L_{x1}(v_{L1}^2p_{L1}^2-v_{T1}^2p_{T1}^2)T_{x1}\\ &\quad-p_{L1}iq(v_{L1}^2-v_{T1}^2)L_{y1}-p_{T1}iq(v_{L1}^2-v_{T1}^2)T_{y1}=0,\\ &-p_{L1}iq(v_{L1}^2-v_{T1}^2)L_{x1}-p_{T1}iq(v_{L1}^2-v_{T1}^2)T_{x1}\\ &\quad+(v_{T1}^2p_{T1}^2-v_{L1}^2p_{L1}^2)L_{y1}+[(v_{T1}^2-v_{L1}^2)p_{T1}^2]T_{y1}=0. \end{aligned} \tag{7}$$

Two continuity conditions for stress, one continuity condition for the normal displacement at the liquid solid, and one kinematic condition (equation of motion) for the liquid render four extra equations in the six unknowns $L_{x,y,1,2}$ and $T_{x,y,1}$

$$\begin{aligned} &\sigma_{xy1}(y=0)=0,\\ &\sigma_{yy,1}(y=0)=\sigma_{yy,2}(y=0),\\ &u_{y,1}(y=0)=u_{y,2}(y=0),\\ &p_{L2}\sigma_{yy,2}=-\rho_2\omega^2u_{y,2}. \end{aligned} \tag{8}$$

After elimination of the liquid unknowns, a 4×4 set of equations in $L_{x,y,1}$ and $T_{x,y,1}$ is obtained

$$A\cdot\begin{pmatrix} L_{x1}\\ T_{x1}\\ L_{y1}\\ T_{y1}\end{pmatrix}=0 \tag{9}$$

with

$$A=\begin{bmatrix} p_{L1}^2 & q^2 & -p_{L1}iq & -p_{T1}iq \\ p_{L1}iq & p_{T1}iq & q^2 & p_{T1}^2 \\ p_{L1} & p_{T1} & iq & iq \\ iq(v_{L1}^2-2v_{T1}^2) & iq(v_{L1}^2-2v_{T1}^2) & v_{L1}^2 p_{L1}-\dfrac{\rho_2\omega^2}{\rho_1 p_{L2}} & v_{L1}^2 p_{T1}-\dfrac{\rho_2\omega^2}{\rho_1 p_{L2}} \end{bmatrix}. \tag{10}$$

From Eqs. (9) and (10), solutions for propagating waves for the Scholte and leaky Rayleigh mode can be explicitly calculated, but solutions exist only if the (Scholte) determinant is zero, resulting in a condition for the Scholte and leaky Rayleigh roots $q(\omega)$. The resulting dependence between the four equations allows elimination of one equation. To find propagating wave solutions, we can fix L_{x1} to a nonzero arbitrary value. We can obtain the solution for the three other constants T_{x1}, L_{y1}, and T_{y1} by solving

$$A_{\text{red}}\cdot\begin{pmatrix} T_{x1} \\ L_{y1} \\ T_{y1} \end{pmatrix}=-L_{x1}\cdot\begin{pmatrix} p_{L1}^2 \\ p_{L1}iq \\ p_{L1} \end{pmatrix}. \tag{11}$$

The liquid displacement constants can be found by

$$L_{y2}=(L_{y1}+T_{y1}), \tag{12}$$

$$L_{x2}=-\frac{iq}{p_{L2}}L_{y2}=-\frac{iq}{p_{L2}}(L_{y1}+T_{y1}). \tag{13}$$

Inserting these constants in Eqs. (1) and (4) gives the displacement and stress profiles in the solid and in the liquid. The phase velocities of these solutions are found by solving the characteristic Scholte determinant equation $\Delta_S=0$ using Eq. (10). Care should be taken in choosing the signs of the square roots of the wave numbers defined in Eq. (2). In principle, all roots are allowed, i.e., $\pm p_{L2}$, $\pm p_{L1}$, $\pm p_{T1}$, giving eight possible choices (or, in the formalism of Ref. 12, eight Riemann sheets). Symmetry of the determinant reduces it to four. When finding a root, attention should be paid to whether the root is physical, by checking whether or not the corresponding wave is divergent.[2]

III. THE "HARD SOLID–FLUID" INTERFACE

The equations of Sec. II allow the calculation of the dispersion curve and the displacements in the solid and the liquid. First we consider a "hard solid–fluid" configuration (i.e., velocities $v_{L2}<v_{T1}<v_{L1}$), e.g., glass–water. We have simulated the wave profiles for this case, with $\rho_2=1000\ \text{kg m}^{-3}$ and $v_{L2}=1500\ \text{m s}^{-1}$ for water, and $\rho_1=2500\ \text{kg m}^{-3}$ and $v_{L1}=5712\ \text{m s}^{-1}$ and $v_{T1}=3356$ for glass. The root of the Scholte determinant at $1496.1\ \text{m s}^{-1}$ corresponds to the Scholte wave, whose normal displacement profile $u_y(y)$ is shown in the solid and in the liquid in Fig. 2. The Scholte wave travels unattenuated along the x axis and has an exponential decay in the y direction, both in the solid and in the liquid. Comparing the decay in the solid and the liquid, it is clear that almost all Scholte wave energy is localized in the liquid and the Scholte wave nearly behaves as a plane bulk wave (weak localization, velocity only slightly lower than bulk velocity) traveling in the liquid along the interface. This situation is typical when the solid is much harder than the liquid (see, e.g., Ref. 9) and renders solid characterization difficult. This will be discussed in further detail in Sec. V.

The leaky Rayleigh wave was found at a velocity of $(3091+109i)\ \text{m s}^{-1}$ (with i the imaginary unit). The leakage of the wave energy in the liquid causes attenuation along the propagation direction (the x direction). In the solid (Fig.

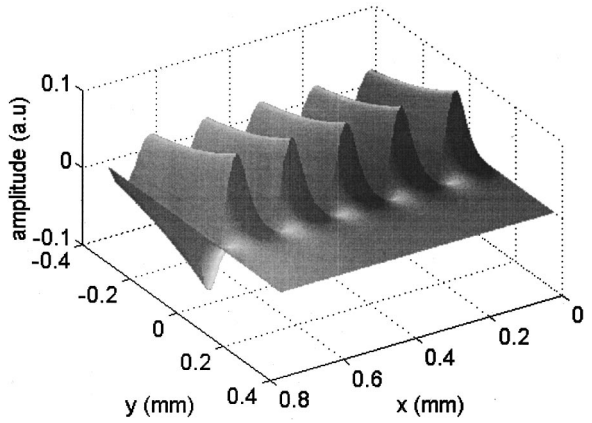


FIG. 2. Instantaneous distribution of the normal displacement component u_y for the Scholte wave (10 MHz) in the solid and in the liquid for a configuration of glass/water ($v=1496.1\ \text{m s}^{-1}$).

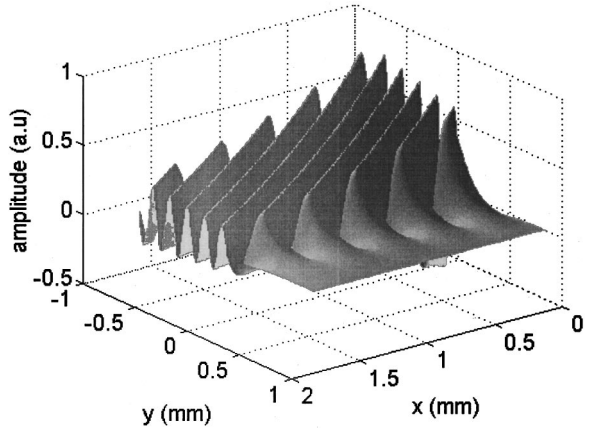


FIG. 3. Instantaneous distribution of the normal displacement component u_y for the leaky Rayleigh wave (10 MHz) in the solid and in the liquid for a configuration of glass/water [$v=(3091+109i)\ \text{m s}^{-1}$]. The Rayleigh wave is damped due to energy leakage into the liquid, under the Rayleigh angle (64° with respect to the interface).

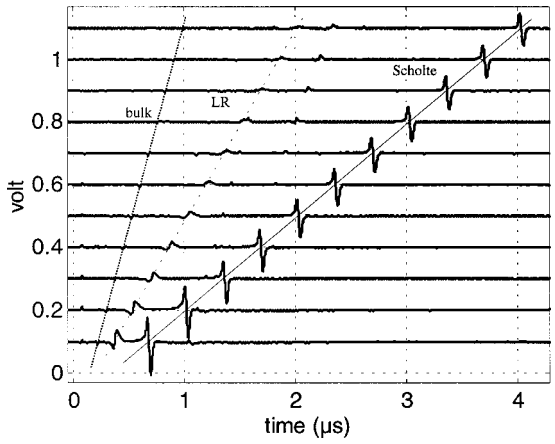


FIG. 4. Measured time traces of deflection signals detected at the interface between water and glass, for pump–probe distances varying between 1 and 6 mm (in steps of 0.5 mm; traces are shown with an increasing offset for increasing pump–probe distance). The full line denotes the time of arrival of the Scholte wave signal [(1488 ± 2) m s^{-1}], the dashed line refers to the leaky Rayleigh wave signal [(3091 ± 12) m s^{-1}], and the dotted line to the bulk wave in the glass. The deflection signals are bipolar, as they correspond to the tangential derivative of the (monopolar) normal displacement.

3, $y>0$) the decay along the y axis is clear. Figure 3 shows clearly that energy of the leaky Rayleigh wave is radiated under the form of a longitudinal wave in the liquid propagating under the Rayleigh angle direction (64° with respect to the interface). In the figure, there is an apparent divergence of the u_y amplitude along the y axis. However, along the propagation direction, i.e., under the Rayleigh angle, the amplitude stays perfectly constant.

Several methods exist to excite both leaky Rayleigh and Scholte waves. See, e.g., Ref. 11 and references therein. We have used laser-induced thermoelastic excitation. This method has shown to be very efficient in generating both leaky Rayleigh[15–17] and Scholte waves.[17,18] In our experiment, a pulsed (8-ns) Nd:YAG laser beam was focused onto a line (width approximately 30 μm) at the interface between the transparent solid and the liquid (which was colored by a colorant to absorb the laser energy). The absorption causes a strong thermal expansion, which in turn launches interface waves propagating along the interface. The displacements were detected optically, monitoring the deflection of a He–Ne probe beam. With this detection method, the spatial derivative du_y/dx of the normal displacement component was measured. The sensitivity of the instrument was about 110-V photodetector signal per radian. The experimental setup has been described in detail elsewhere (see, e.g., Refs. 17 and 18). The detected signals were broadband (approximately 50-MHz bandwidth) and the frequency was only limited by the focus of the pump beam. During the experiment, the probe beam position was fixed and the pump–probe distance was varied by moving the pump beam. From a spectral analysis of the interface waves for different pump–probe distances, the phase velocity of the signals could be calculated.

Experimentally observed Scholte and leaky Rayleigh waves for the configuration of glass/water are shown in Fig. 4, over a propagation distance of 5 mm. From the time signals, clearly the leaky Rayleigh wave arrival (dashed line) and Scholte wave arrival (full line) can be distinguished. Also, strong attenuation of the leaky Rayleigh wave is visible, in contrast with a weak attenuation for the Scholte wave. Theoretically, the Scholte wave is undamped (the wave number is real). However, the experimentally observed damping is due to a combination of nonlinearity (Mach number up to 0.01) and viscosity.[19] The bipolarity of the signals is typical for a deflection setup. In order to obtain the normal displacement u_y, the signals should be integrated. The velocities found from the zero-crossing point of the different time signals were (1488 ± 2) m s^{-1} for the Scholte wave signal (full line in Fig. 4) and $[(3091\pm12)+(110\pm30)i]$ m s^{-1} for the leaky Rayleigh wave signal (dashed line in Fig. 4). Together with our experimental value for the longitudinal velocity (5712 ± 40) m s^{-1} of the glass substrate (determined by a pulse-echo transducer measurement), by doing a zero search of the equation $\Delta_S(v_{T1})=0$, we could determine the transverse velocity $v_{T1}=(3356\pm15)$ m s^{-1} and the Poisson ratio of 0.23 ± 0.01 of the glass substrate. Fourier analysis of the time signals revealed no dispersion in either Scholte or leaky Rayleigh waves in the experimentally accessible frequency range, confirming theory. Using the experimentally determined values for the glass substrate, we calculated for the imaginary part of the leaky Rayleigh wave velocity 109 m s^{-1}, which corresponds well with the experimentally observed value of (110 ± 30) m s^{-1}.

IV. THE "SOFT SOLID–FLUID" CASE

For a "soft solid–fluid" configuration (i.e., $v_{T1}<v_{L2}<v_{T1}$), the picture changes. We have done calculations for a configuration of Plexiglas–water. The bulk parameters used for Plexiglas were 1190 kg m^{-3} for the density, 2692 m s^{-1} for the longitudinal velocity, and 1407 m s^{-1} for the transversal velocity. The calculation unambiguously predicts a Scholte wave with a velocity of 1067 m s^{-1}. The behavior of the Scholte wave is shown in Fig. 5. Compared to the Scholte wave in the hard solid–fluid configuration (Fig. 2), the Plexiglas–water Scholte wave is localized much closer to the interface. A second root in the Scholte determinant was found at 1381 m s^{-1}. However, the plot of the amplitude

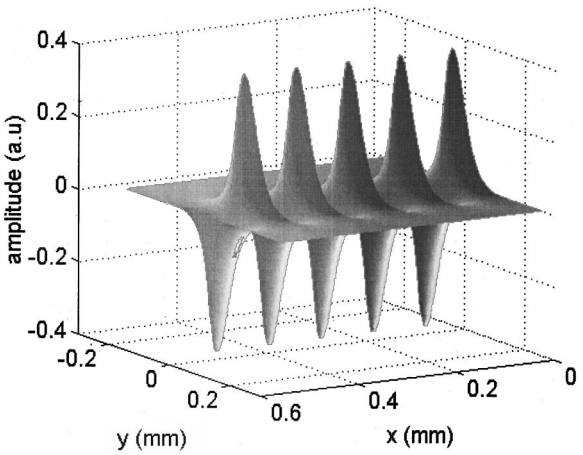


FIG. 5. Instantaneous distribution of the normal displacement component u_y for the Scholte wave (10 MHz) in the solid and in the liquid for a configuration of Plexiglas/water ($v=1067$ m s^{-1}).

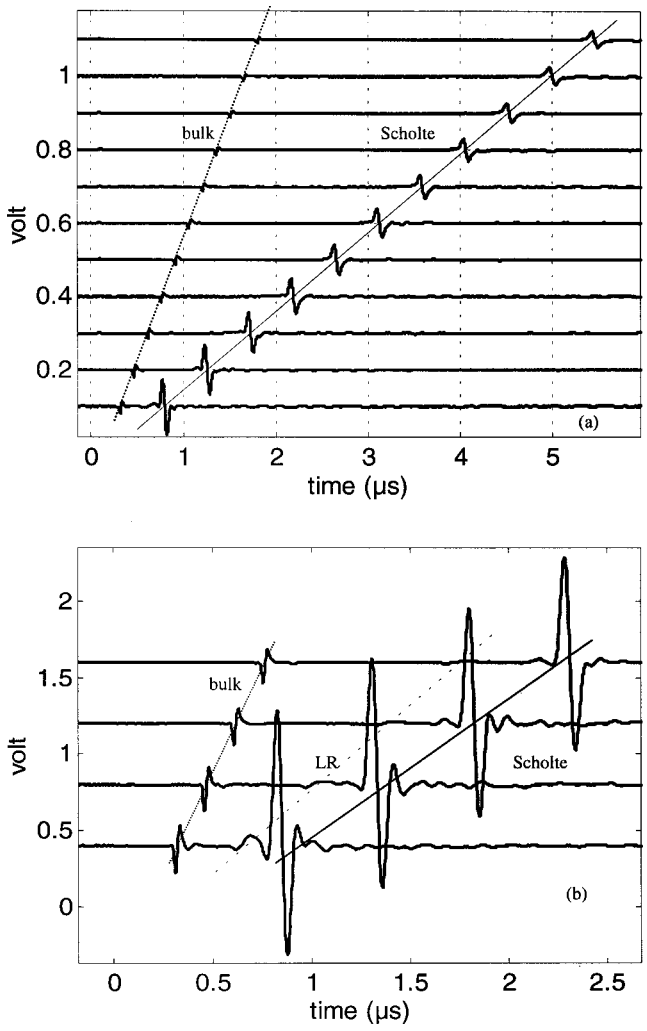


FIG. 6. (a) Experimental deflection signals detected at the interface between water and Plexiglas, for the pump–probe distance varying between 0.84 and 5.84 mm (in steps of 0.5 mm). The full line denotes the time of arrival of the Scholte wave signal [(1067 ± 2) m s^{-1}]. The dotted line corresponds to a bulk wave arrival. No wave with velocity around 1381 m s^{-1} is observed. (b) Measured time signals at the interface between Plexiglas and methanol, for a pump–probe distance varying between 0.90 and 2.40 mm (in steps of 0.5 mm). The dashed line denotes the time of arrival of the leaky Rayleigh wave signal [(1377 ± 45) m s^{-1}]. The dotted line corresponds to a bulk wave arrival and the full line with the arrival of the Scholte wave [(1011 ± 1) m s^{-1}].

dependence of the u_y component in the solid reveals a divergence. Since the root has no imaginary (leaking) part, this solution has to be rejected. This root has been incorrectly identified as the leaky Rayleigh wave in Ref. 12. Though its identification remains difficult, it is possible that the wave observed in Ref. 12 was a transverse lateral wave, which had been mode converted at the (solid–air)–(solid–liquid) transition from a Rayleigh wave incident at the liquid–air boundary in their configuration.

Figure 6(a) shows time traces of experimentally detected deflection signals in this soft solid–fluid configuration Plexiglas/water, for pump–probe distances varying over 5 mm. The Scholte wave arrival time is denoted by the full line and the Scholte wave velocity was found at (1067 ± 12) m s^{-1}. From this value for the Scholte wave velocity, from the longitudinal wave velocity of water (1500 m s^{-1}), and from the longitudinal wave velocity of Plexiglas (2692 ± 40) m s^{-1} (determined by a pulse-echo transducer measurement), we found a corresponding value of (1407 ± 10) m s^{-1} for the shear velocity of the Plexiglas substrate and a value of 0.31 ± 0.01 for the Poisson ratio. A slight decay can be observed in the Scholte time signals, which is mainly due to the viscoelasticity of the Plexiglas material and nonlinear and viscous effects, which were not included in the theoretical model of Sec. II. The fast wave arrival in Fig. 6(a) (and in Fig. 4) corresponds to a bulk wave in Plexiglas which is also excited by the pump beam, and which generates a signal due to stress-induced refractive index changes in the region where the bulk wave crosses the optical path of the probe beam. No wave with a phase velocity of 1381 m s^{-1} (as predicted in Ref. 12) is observed.

In order to investigate in more detail the forbidden leaky Rayleigh-type root, we have investigated the existence of waves versus the composition of water–methanol mixtures at the interface with Plexiglas. The forbidden character of the leaky Rayleigh root for Plexiglas–water is related to the acoustic velocity of water (1500 m s^{-1}) exceeding the transverse velocity of Plexiglas (1407 m s^{-1}). This is not the case for Plexiglas–pure methanol ($\rho_{L2}=790$ kg m^{-3} and $v_{L2}=1115$ m s^{-1}). Thus, with increasing methanol concentration there should be a transition between the forbidden ($v<v_{L2}$, towards pure water) and allowed ($v>v_{L2}$, towards pure methanol) regime. This is confirmed by the experiment. In Fig. 6(b), experimentally detected time signals are shown for Plexiglas loaded with pure methanol. A leaky Rayleigh wave can be observed with a propagation velocity (1377 ± 45) m s^{-1}, apart from the Scholte wave, which has a velocity of (1011 ± 1) m s^{-1}. Together with the longitudinal velocity of Plexiglas [$v_{L1}=(2692\pm40)$ m s^{-1}], and the methanol density and bulk velocity [$v_{L2}=(1115\pm5)$ m s^{-1}, determined by a pulse-echo transducer measurement], this corresponds to [by solving the equation $\Delta_S(v_{T1})=0$] a transverse Plexiglas velocity $v_{T1}=(1421\pm15)$ m s^{-1}) from the Scholte wave velocity and to $v_{T1}=(1352\pm50)$ m s^{-1} from the leaky Rayleigh wave propagation velocity. The respective Poisson ratios are 0.307 ± 0.004 and 0.331 ± 0.016. These results are consistent with the value obtained before from the Scholte wave velocity at the Plexiglas–water interface [$v_{T1}=(1407\pm10)$ m s^{-1}, Poisson ratio 0.31 ± 0.01].

When the concentration of water is increased, the acoustic velocity of the methanol/water gradually increases. Different mixtures were made and the acoustic bulk velocities were measured with a time-of-arrival experiment using a conventional ultrasonic transducer (see Fig. 7). For an extended set of molar fractions of the mixture (see inset of Fig. 7), the roots of the Scholte determinant were calculated. Experimentally determined velocities (by linear regression of the times of arrival of the zero-crossing point of the signals vs pump–probe distance) are given in full symbols (triangles and circles). While for methanol concentrations above 80% the leaky Rayleigh wave is observed, below 80% the observed wave has a velocity which corresponds with the longitudinal velocity of the liquid, rather than with the velocity corresponding to the forbidden root. Interestingly, the calcu-

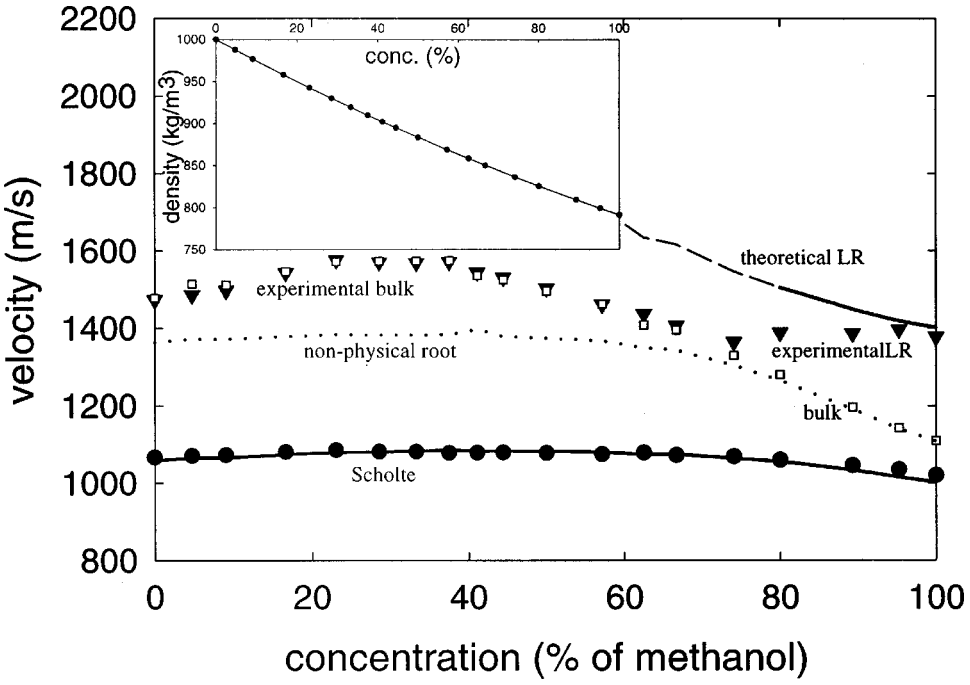


FIG. 7. Calculated (lines) and experimental (symbols) velocities as a function of methanol concentration in the water/methanol mixture at the interface with Plexiglas. Open squares denote the bulk velocities of the mixtures measured with a piezoelectric transducer. From these velocities, and from the densities shown in the inset, the theoretical Scholte and leaky Rayleigh velocities are calculated (full lines). Full circles denote the experimentally observed Scholte wave velocity and the full triangles denote the detected bulk (below 80% methanol concentration) or the leaky Rayleigh mixed with the bulk wave (about 80% methanol concentration). Below 80%, the acoustic velocity of the mixture exceeds the transverse velocity of Plexiglas and the leaky Rayleigh wave evolves to a high-velocity mode (dashed line). The dotted line corresponds to an unphysical root of the Scholte determinant, identified in Ref. 12 as the leaky Rayleigh root.

lations reveal that with decreasing methanol concentration, the leaky Rayleigh root evolves to a slightly damped high-velocity mode with a peculiar character. An illustration of the normal displacement component of this mode at a methanol concentration of 67% is given in Fig. 8. This mode is not observed experimentally below 80% methanol concentration. Above 80% the experimental velocities lie close, but systematically a little bit below the theoretically predicted ones. However, the small discrepancy could be due to experimental uncertainties caused by the overlap between the bulk wave and the leaky Rayleigh mode. Or, there could be an effective mixing of both kinds of waves. Figure 8 illustrates that the high-velocity leaky Rayleigh wave solution exhibits a very weak localization in the solid compared to ''ordinary'' leaky Rayleigh waves which may render this kind of wave difficult to excite with our localized laser excitation.

V. SCHOLTE WAVE CHARACTERIZATION OF SOLIDS

As the Scholte wave is localized at the solid–liquid interface, it potentially can yield information about the solid. By inversion of the Scholte determinant, the Scholte wave velocity is related to the shear velocity of the solid material.[11,13] Additionally, when introducing viscoelasticity in the solid, the Scholte decay in the *y* direction in the liquid can yield additional information about the attenuation coefficient in the solid. This was done in Ref. 13. Technologically, the interest in the Scholte wave for solid characterization arises from the intrinsic absence of attenuation of the Scholte wave, in contrast with the leaky Rayleigh wave.

It should be noted, however, that this inversion is only possible in cases where the substrate is relatively ''soft'' compared with the liquid, because only in that case is there a strong localization of the Scholte wave at the interface and relatively much energy is localized in the solid, as shown in the previous section (see Fig. 5). In the other case, when the

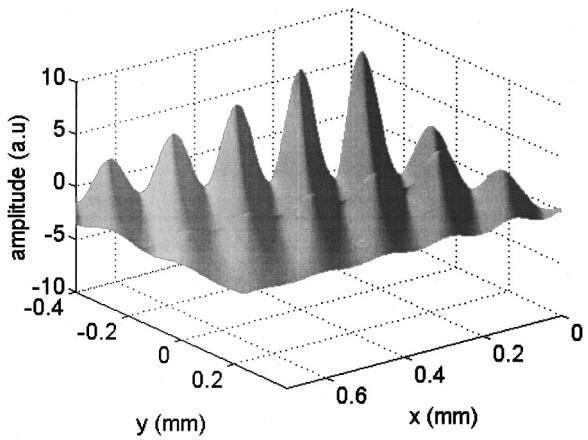


FIG. 8. Instantaneous distribution of the normal displacement component u_y for the interface wave (10 MHz) in the solid and in the liquid for a configuration of Plexiglas loaded with a water/methanol mixture (67% of methanol). The wave has a velocity of $(1616+97i)$ m s^{-1}.

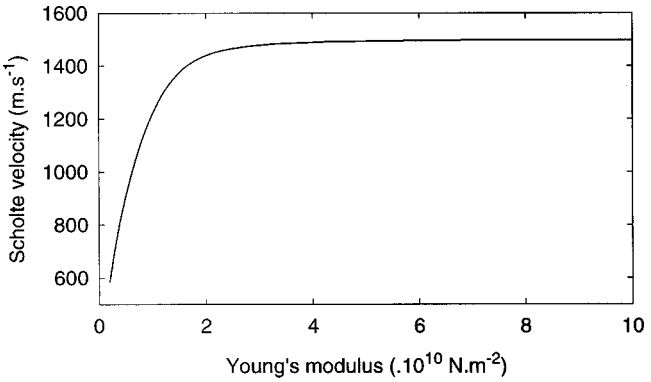


FIG. 9. Scholte phase velocity as a function of the substrate Young's modulus, with Poisson's ratio fixed at $\sigma=0.35$ and density at $\rho_1=1500$ kg m^{-3}. The liquid used is water with $v_{L2}=1500$ m/s and $\rho_2=1000$ kg/m^3.

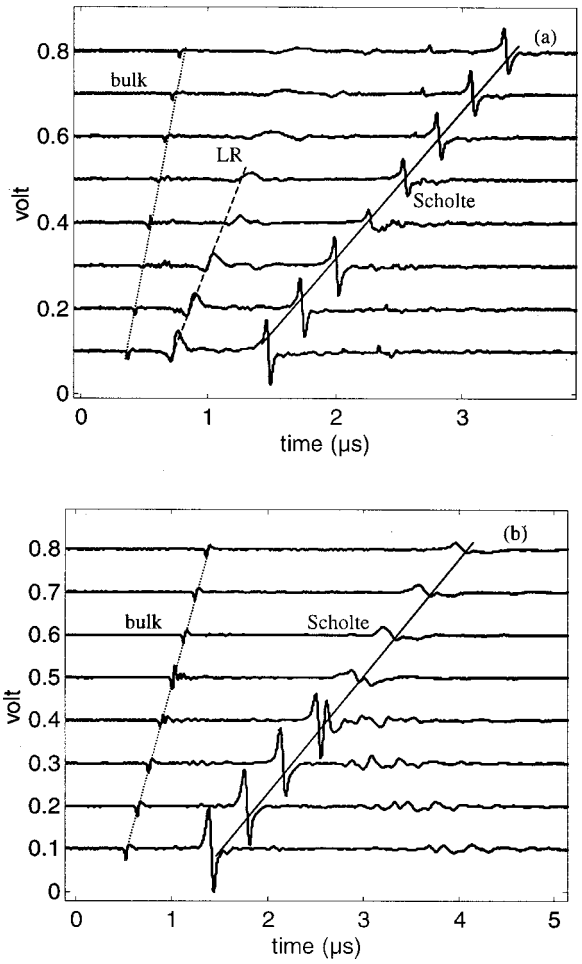


FIG. 10. (a) Time traces of deflection signals of waves propagating at the glass/water interface. Traces are shifted for convenience. From bottom to top, pump–probe distances are 2.2, 2.6, 3.0, 3.4, 3.8, 4.2, 4.6, and 5.0 mm. A scratch is located at pump–probe distance around 3.4 mm. The full line denotes the time of arrival of the Scholte wave signal, the dashed line of the leaky Rayleigh wave signal. The dotted line corresponds to a bulk wave arrival. (b) Waves propagating at the Plexiglas/water interface. Plots are shifted for convenience. From bottom to top: pump–probe distances of 1.6, 2.0, 2.4, 2.8, 3.2, 3.6, 4.0, and 4.4 mm. A scratch is located at pump–probe distance around 2.8 mm. The full line denotes the time of arrival of the Scholte wave signal. The dotted line corresponds to a bulk wave arrival.

solid is hard compared to the liquid (the hard solid–fluid configuration), the inversion is no longer possible because the Scholte wave contains almost no information about the solid (Fig. 2). This is illustrated in Fig. 9, where the Scholte wave velocity is plotted as a function of the substrate Young's modulus, with the Poison ratio of the polymer fixed to be 0.35 and the density to $\rho_1 = 1500$ kg m^{-1} (as a liquid, we took water: $\rho_2 = 1000$ kg m^{-3} and $v_{L2} = 1500$ m s^{-1}). As the solid becomes stiffer, the Scholte velocity saturates at the liquid bulk velocity.

In order to verify this experimentally, we designed the following experiment. A scratch of approximately 100-μm width and depth was made in a piece of glass and a piece of Plexiglas. The two solids were liquid loaded and surface waves were excited. Detected signals can be seen in Fig. 10(a) for the glass/water configuration and in Fig. 10(b) for the Plexiglas/water configuration. When increasing the pump–probe distance, a point is reached where the pump is passing the scratch. In the glass–water configuration we see the signal is distorted when exciting on the scratch. However, when further increasing the pump–probe distance, it is clear that, though having passed the scratch, the Scholte wave experiences almost no influence from the presence of the scratch. In the Plexiglas case, however, the Scholte wave is completely distorted after passing the scratch. From this, it can be concluded that Scholte waves can be used for nondestructive material characterization, though only for soft materials.

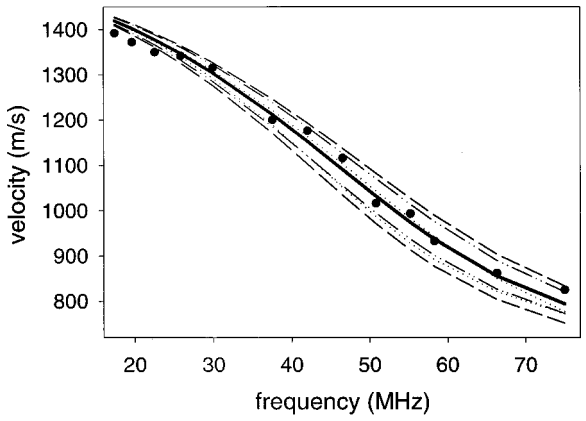


FIG. 11. Scholte velocity dispersion curve for a liquid (water) loaded soft (polycarbonate) coating on hard substrate (glass). The dotted curve is the experimental dispersion spectrum. Fitting yields: thickness $d = (3.6 \pm 0.2)$ μm, $v_{LC} = (1373 \pm 50)$ m s^{-1}, and $v_{TC} = (950 \pm 90)$ m s^{-1}. The full line denotes the best-fitting dispersion curve. The dashed lines illustrate the errors: v_{LC} is perturbed to 1323 m s^{-1} and 1423 m s^{-1} (long-dashed lines), v_{TC} is perturbed to 860 m s^{-1} and 1040 m s^{-1} (dotted lines), and d is perturbed to 3.4 μm and 3.8 μm (dashed-dotted lines).

VI. SCHOLTE DISPERSION IN THE LIQUID–COATED-SUBSTRATE CONFIGURATION

As is the case for Rayleigh waves, the penetration depth of Scholte waves is inversely proportional to the frequency, offering the potential for characterization of coatings and for depth profiling. This was illustrated experimentally by de Billy *et al.*[20] Moreover, the wide bandwidth of laser excitation should allow investigating relatively thin layers. Figure 11 shows the experimental and fitted frequency dependence of the Scholte wave velocity for a liquid/coated-substrate configuration, in the case of a 3.5-μm polycarbonate coating, which was softer than the glass substrate. The dispersion spectrum was determined by performing a heterodyne diffraction experiment[21] at a selected series of acoustic wavelengths. At high frequencies, the Scholte wave is confined to the soft coating and is slowed down. At low frequencies, the Scholte wave decay depth increases and, since the substrate is hard, the Scholte wave velocity lies very close to the velocity of the longitudinal wave in the liquid. In order to find the theoretical roots, the characteristic determinant had to be constructed. This can be done analogously to the construction of the Scholte determinant in Sec. II, but a 7×7 determinant is now obtained. The fit of $v(\omega)$ was performed by minimizing the quadratic difference between the experimental dispersion curve and the calculated one. During the fit, the film bulk properties (thickness d, bulk longitudinal and

transversal velocities, respectively, v_{LC} and v_{TC}) were assumed as varying fitting parameters, keeping constant the substrate (glass, $v_{L1}=5712$ m s^{-1} and $v_{T1}=3356$ m s^{-1} and $\rho_1=2500$ kg m^{-3}) and liquid (water) parameters and the density of the film ($\rho_C=1180$ kg m^{-3}). This three-parameter fit allowed simultaneous determination of the thickness $d=(3.6\pm0.2)$ μm, longitudinal [$v_{LC}=(1373\pm50)$ m s^{-1}], and transverse [$v_{TC}=(950\pm90)$ m s^{-1}] velocity of the polymer film. We determined the errors by analyzing the parameter dependence of the chi-squared function, illustrated in Fig. 11. Possible degeneracy between different fitting parameters was absent, illustrated by the different effect of perturbations on the respective parameters on the shape of the dispersion curve. Because of the intrinsic absence of attenuation in the Scholte wave, this method of coating characterization offers a valuable alternative to the characterization by the (quasi)-Rayleigh wave (as was done in Ref. 22) in cases where the liquid loading (and thus attenuation of the Rayleigh-type wave) is unavoidable (e.g., in seismic applications). To obtain good fit results, it is necessary to have sufficient dispersion in the frequency range measured, which can only be fulfilled if the coating material is soft compared to the liquid and the substrate material is hard, or vice versa.

VII. CONCLUSION

We have used the efficient excitation and sensitive detection characteristics of laser ultrasonics to investigate the character of acoustic waves traveling along the interface between a liquid and a solid or coated solid. It was illustrated that leaky Rayleigh and Scholte wave laser-ultrasonic spectroscopy is a suitable technique for the characterization of soft solids and coated solids down to the micron scale. Peculiar phenomena occurring when the liquid is supersonic with respect to the shear velocity of the solid have been investigated in detail theoretically and experimentally.

ACKNOWLEDGMENTS

This work was supported in part by U.S. National Science Foundation Grant No. DMR-9710140, and by Fonds voor Wetenschappelijk Onderzoek-Vlaanderen (FWO-V).

[1] I. A. Viktorov, *Rayleigh and Lamb Waves*, edited by L. Balamuth (Plenum, New York, 1967).
[2] V. Gusev, C. Desmet, W. Lauriks, C. Glorieux, and J. Thoen, J. Acoust. Soc. Am. **100**, 1514–1528 (1996).
[3] H. Überall, "Surface waves in acoustics," in *Physical Acoustics*, edited by W. P. Mason and R. D. Thurston (Academic, New York, 1973), Vol. 10, pp. 1–60.
[4] J. D. Achenbach, *Wave Propagation in Elastic Solids* (North-Holland, Amsterdam, 1973).
[5] L. M. Brekhoviskikh, *Waves in Layered Media* (Academic, New York, 1980).
[6] G. W. Farnell, "Properties of elastic surface waves," in *Physical Acoustics*, edited by W. P. Mason and R. D. Thurston (Academic, New York, 1970), Vol. 6, pp. 109–165.
[7] B. Poirée and F. Luppé, J. Acoust. **4**, 575–588 (1991).
[8] L. Adler, M. de Billy, and G. Quentin, J. Appl. Phys. **53**(12), 8756–8758 (1982).
[9] S. Nasr, J. Duclos, and M. Leduc, J. Acoust. Soc. Am. **87**, 507–512 (1990).
[10] N. Favretto-Anrès and J. Sessarego, Acustica **85**, 505–516 (1999).
[11] N. Favretto-Anrès and G. Rabau, J. Sound Vib. **203**(2), 193–208 (1997).
[12] F. Padilla, M. de Billy, and G. Quentin, J. Acoust. Soc. Am. **106**, 666–673 (1999).
[13] B. H. Ali and M. K. Broadhead, in *Full Field Inversion Methods in Ocean- and Seismo-Acoustics*, edited by O. Iachok, A. Caiti, P. Gerstoft, and H. Schmidt (Kluwer Academic, Dordrecht, 1995), pp. 371–376.
[14] N. Favretto-Anrès, Acustica **82**, 829–838 (1996).
[15] H. Coufal, R. Grygier, P. Hess, and A. Neubrand, J. Acoust. Soc. Am. **92**, 2980–2983 (1992).
[16] C. Desmet, V. Gusev, W. Lauriks, C. Glorieux, and J. Thoen, Opt. Lett. **22**, 69–71 (1997).
[17] K. Van de Rostyne, C. Glorieux, W. Gao, W. Lauriks, and J. Thoen, Acta Phys. Sin. **8**, 219–224 (1999).
[18] C. Desmet, V. Gusev, W. Lauriks, C. Glorieux, and J. Thoen, Appl. Phys. Lett. **68**(21), 2939–2941 (1996).
[19] C. Glorieux, K. Van de Rostyne, V. Gusev, W. Gao, W. Lauriks, and J. Thoen (unpublished).
[20] M. de Billy and G. Quentin, J. Acoust. Soc. Am. **88**, 535–541 (1990).
[21] A. A. Maznev, K. A. Nelson, and J. A. Rogers, Opt. Lett. **23**(16), 1319–1321 (1998).
[22] A. Neubrand and P. Hess, J. Appl. Phys. **71**(1), 227–238 (1992).

Determination of Lamb mode eigenvalues

V. Pagneux [a)]
Laboratoire d'Acoustique de l'Université du Maine, UMR CNRS 6613, Avenue Olivier Messiaen, 72085 Le Mans Cedex 9, France

A. Maurel
Laboratoire Ondes et Acoustique, UMR CNRS 7587, Ecole Supérieure de Physique et de Chimie Industrielles, 10 rue Vauquelin, 75005 Paris, France

(Received 22 November 2000; revised 10 June 2001; accepted 12 June 2001)

An original method is presented to determine the complex Lamb wave spectrum by using a numerical spectral method applied to the elasticity equations. This method presents the advantage to directly determine complex wave numbers for a given frequency via a classical matricial eigenvalue problem, and allows the wave numbers to be determined at relatively high frequencies (i.e., corresponding to many propagating modes). It does not need initial guess values for the wave numbers, contrary to the usual method of root finding of the Rayleigh–Lamb frequency equations (dispersion relation) in the complex plane. Results are presented and the method is discussed. [DOI: 10.1121/1.1391248]

PACS numbers: 43.20.Jr, 43.20.Mv, 43.40.Dx, 02.70.Hm [DEC]

I. INTRODUCTION

Lamb waves are involved in the nondestructive testing of plate structures because of their guided nature. The characterization of flaws through the scattering of elastic waves in such plates has received considerable attention in the past 20 years,[1] but Lamb wave propagation in an inhomogeneous medium has not been widely investigated.

A possible method for solving the problem of Lamb wave propagation in a medium with geometrical or material discontinuities is based on an eigenfunction expansion of the displacement and the stress, where the considered eigenfunctions are the Rayleigh–Lamb modes in an infinite plate. This expansion is combined with a mode-matching technique to treat the discontinuities. This method solves problems related to a semi-infinite plate[2,3] or to two dissimilar semi-infinite plates welded along their lateral boundaries.[4]

This method requires determining the complex wave number spectrum k_n associated with the eigenfunctions used in the expansion. In the context of Lamb waves, it is difficult to determine eigenfunctions and associated wave numbers, because of the spectrum complexity. It is well established that the Lamb wave spectrum in a free solid layer, composed of elastic material, consists of complex wave numbers k_n, real wave numbers corresponding to propagating Lamb waves, and complex wave numbers related to evanescent Lamb waves.[5–9] Usually, the dispersion relation $D(\omega,k)=0$ (the so-called Rayleigh–Lamb frequency equations) is numerically solved to determine the wave spectrum.

For propagating modes, the problem remains simple since the wave number is known to be real. On the other hand, when the mode-matching technique is used, evanescent modes have to be taken into account and a significant part of the entire complex spectrum has to be determined. The direct strategy of finding the roots of the dispersion relation $D(k,\omega)=0$, at a given frequency ω, is not well suited for systematic computation since the locations of the wave numbers are not known, *a priori*, in the complex plane. Consequently, initial guess values are not available for classical root finding routines. One technique to overcome this lack[4,2] consists in calculating the spectrum at zero frequency, where the dispersion relation is more tractable, and then to gradually increase the frequency to the desired value by using $dk=-(\partial_\omega D/\partial_k D)d\omega$ and/or $k(\omega)$ as an initial guess for $k(\omega+d\omega)$. Apart from the difficulties arising from the vanishing of the denominator, this ''step by step'' technique is time consuming because a series of spectra has to be computed for frequencies from zero to the desired frequency.

Here, we present an original method that directly projects the ordinary differential equation governing the Lamb modes on a spectral basis of orthogonal functions. Instead of solving the transcendental equation $D(k,\omega)=0$, one calculates the solutions of a classical eigenvalue problem in the form $(M-kI)\mathbf{X}=0$, where M is a matrix resulting from the projection of the differential equation. We obtain approximate eigenvalues that can be used as starting values for a more precise solution. This technique is classically used in the theory of fluid dynamics instability.[10,11] In this case, the coefficients of the differential equation governing the transverse modes are nonconstant; the method of finding the roots of the dispersion relation is then not natural since it is impossible to get an *analytical* dispersion relation.

The Lamb problem is posed in Sec. II. Then, spectral decomposition is performed (Sec. III), and a second-order polynomial system on k is derived (Sec. IV A). This system can be rewritten as a reduced eigenvalue problem for k_n^2 (Sec. IV B). Results are presented and discussed in Sec. V.

II. LAMB MODE PROBLEM

The Lamb mode problem (see Fig. 1) consists of searching for a solution of the elasticity equation in the waveguide defined by $-h\leqslant y\leqslant h$ with free boundaries, and for which

a)Electronic mail: vincent.pagneux@univ-lemans.fr

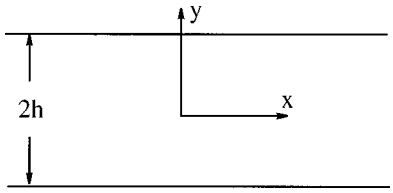


FIG. 1. Geometry of the Lamb wave problem.

displacements are in the (x, y) plane. The time dependence is $e^{-i\omega t}$ and will be omitted in the sequel. The equation of motion is

$$-\rho\omega^2\mathbf{w}=\mu\,\Delta\mathbf{w}+(\lambda+\mu)\boldsymbol{\nabla}(\mathrm{div}\,\mathbf{w}), \tag{2.1}$$

where ρ is the density, (λ, μ) are the Lamé's constants, and $\mathbf{w}=(\hat{u},\hat{v})$ is the vector of displacements, whose components are of the "modal" form:

$$\begin{pmatrix}\hat{u}(x,y)\\ \hat{v}(x,y)\end{pmatrix}=\begin{pmatrix}u(y)\\ v(y)\end{pmatrix}\exp(ikx). \tag{2.2}$$

The faces $y=\pm h$ are free of traction, corresponding to boundary conditions:

$$\begin{aligned}&\tau_{xy}(x,\pm h)=\mu(\partial_y\hat{u}+\partial_x\hat{v})=0,\\ &\tau_{yy}(x,\pm h)=\lambda\,\partial_x\hat{u}+(\lambda+2\mu)\partial_y\hat{v}=0.\end{aligned} \tag{2.3}$$

Defining $k_t=\sqrt{\rho/\mu}\omega$, $k_l=\sqrt{\rho/(\lambda+2\mu)}\omega$, and $\gamma=(\lambda+2\mu)/\mu$, the system (2.1) can be written for (u,v):

$$k^2u-ik\frac{\gamma-1}{\gamma}v'-\left(k_l^2u+\frac{u''}{\gamma}\right)=0, \tag{2.4a}$$

$$k^2v-ik(\gamma-1)u'-(k_t^2v+\gamma v'')=0, \tag{2.4b}$$

and the boundary conditions (2.3) become

$$u'(\pm h)=-ikv(\pm h) \tag{2.5a}$$

$$v'(\pm h)=-ik\frac{\gamma-2}{\gamma}u(\pm h), \tag{2.5b}$$

where the prime and double prime stand for d/dy and d^2/dy^2.

III. SPECTRAL DECOMPOSITION

The eigenproblem (2.4)–(2.5) is known to be separable into symmetric and antisymmetric solutions, where symmetric (*resp.*, antisymmetric) modes correspond to even (*resp.*, odd) u and odd (*resp.*, even) v.[5] In the following, superscripts s and a, respectively, refer to symmetric and antisymmetric modes. Basis functions ϕ_n for u^s, v^a, and ψ_n for v^s, u^a, with $n\geqslant 1$, are chosen such that

$$\begin{aligned}&\phi_n''+\alpha_n^2\phi_n=0,\quad \text{and}\quad \phi_n'(0)=\phi_n'(h)=0,\\ &\psi_n''+\beta_n^2\psi_n=0,\quad \text{and}\quad \psi_n(0)=\psi_n'(h)=0,\end{aligned} \tag{3.1}$$

that yields

$$\begin{aligned}&\phi_n=\sqrt{\frac{\epsilon_n}{h}}\cos(\alpha_n y),\quad \text{with}\quad \begin{cases}\epsilon_1=1,\ \epsilon_n=2,\ \text{for}\ n\geqslant 2,\\ \alpha_n=\dfrac{(n-1)\pi}{h},\end{cases}\\ &\psi_n=\sqrt{\frac{2}{h}}\sin(\beta_n y),\quad \text{with}\quad \beta_n=\frac{(n-1/2)\pi}{h};\end{aligned} \tag{3.2}$$

ϕ_n and ψ_n are such that $(\phi_n|\phi_m)=\delta_{nm}$ and $(\psi_n|\psi_m)=\delta_{nm}$, where the scalar product is defined by $(f|g)=\int_0^h f(y)g(y)dy$.

Functions ϕ_n (*resp.*, ψ_n) form a complete basis to describe any even (*resp.*, odd) function because they are eigenfunctions of a classical Sturm–Liouville problem.[12] Thus, symmetric and antisymmetric solutions can be decomposed in these bases as

$$\begin{aligned}&u^s(y)=\sum_{n\geqslant 1}U_n^s\phi_n(y),\quad \text{and}\quad \mathbf{U^s}=(U_n^s),\\ &v^s(y)=\sum_{n\geqslant 1}V_n^s\psi_n(y),\quad \text{and}\quad \mathbf{V^s}=(V_n^s),\\ &u^a(y)=\sum_{n\geqslant 1}U_n^a\psi_n(y),\quad \text{and}\quad \mathbf{U^a}=(U_n^a),\\ &v^a(y)=\sum_{n\geqslant 1}V_n^a\phi_n(y),\quad \text{and}\quad \mathbf{V^a}=(V_n^a).\end{aligned} \tag{3.3}$$

The next step is then to obtain the projection of Eqs. (2.4) on the basis functions. This is presented in the following paragraphs.

A. Symmetric modes

The scalar product of (2.4a) by ϕ_n and (2.4b) by ψ_n is performed. Then, the projection of the derivatives are (using the same procedure as in Ref. 13):

$$\begin{aligned}((v^s)'|\phi_n)&=[v^s\phi_n]_0^h-(v^s|\phi_n')\\ &=\sum_{m\geqslant 1}(\phi_n(h)\psi_m(h)-(\phi_n'|\psi_m))V_m^s,\\ ((u^s)''|\phi_n)&=[(u^s)'\phi_n-u^s\phi_n']_0^h+(u^s|\phi_n'')\\ &=-ik\sum_{m\geqslant 1}\phi_n(h)\psi_m(h)V_m^s-\alpha_n^2U_n^s,\end{aligned} \tag{3.4}$$

and

$$\begin{aligned}((u^s)'|\psi_n)&=\sum_{m\geqslant 1}(\psi_n(h)\phi_m(h)-(\psi_n'|\phi_m))U_m^s,\\ ((v^s)''|\psi_n)&=-ik\frac{\gamma-2}{\gamma}\sum_{m\geqslant 1}\psi_n(h)\phi_m(h)U_m^s-\beta_n^2V_n^s.\end{aligned} \tag{3.5}$$

It can be noticed that both boundary conditions have been taken into account in the course of projection [Eq. (2.5a) for (3.4) and Eq. (2.5b) for (3.5)]. Eventually, a system of equations on $\mathbf{U^s}$, $\mathbf{V^s}$ is obtained:

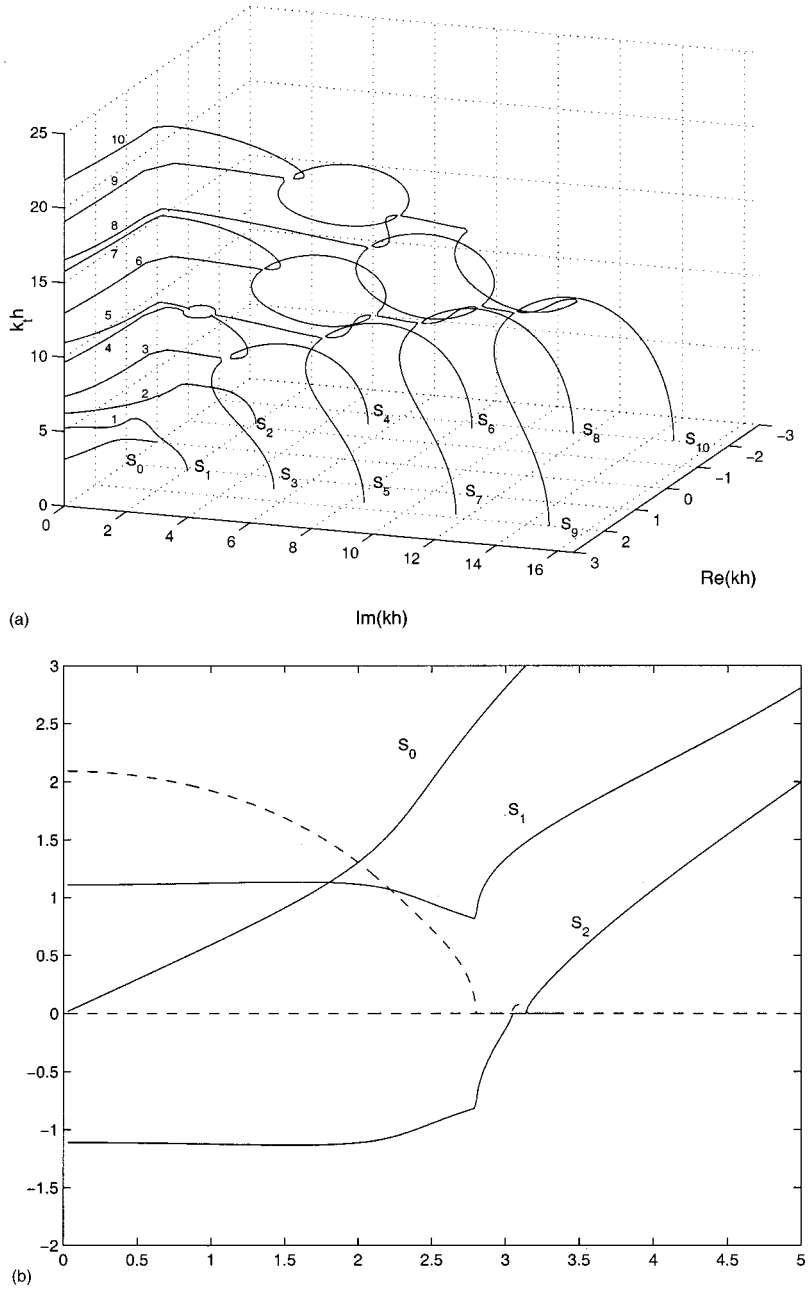


FIG. 2. (a) Tridimensional representation of the dimensionless complex wave numbers $k_n h$ for symmetric modes (S_0 to S_{10}) when the dimensionless frequency $k_t h$ varies, (b) real and imaginary parts of the dimensionless complex wave numbers $k_n h$ for symmetric modes (S_0 to S_3) as a function of the dimensionless frequency $k_t h$.

$$k^2\mathbf{U^s}+kA^s\mathbf{V^s}+B^s\mathbf{U^s}=0,$$
$$k^2\mathbf{V^s}+kC^s\mathbf{U^s}+D^s\mathbf{V^s}=0, \tag{3.6}$$

with matrices A^s, B^s, C^s, and D^s expressed by

$$A^s{}_{m,n}=i\left(\frac{\gamma-1}{\gamma}(\psi_m|\phi_n')+\frac{2-\gamma}{\gamma}\psi_m(h)\phi_n(h)\right)$$
$$=\begin{cases}\dfrac{\sqrt{2}i(2-\gamma)(-1)^m}{h\gamma}, & n=1,\\ \dfrac{2i(-1)^{m+n}(\alpha_n^2+(\gamma-2)\beta_m^2)}{h\gamma(\beta_m^2-\alpha_n^2)}, & n\geqslant 2,\end{cases}$$

$$B^s{}_{m,n}=\left(\frac{\alpha_m^2}{\gamma}-k_l^2\right)\delta_{mn}, \tag{3.7}$$

$$C^s{}_{m,n}=-\gamma A^s_{n,m}, \quad D^s{}_{m,n}=(\gamma\beta_m^2-k_t^2)\delta_{mn}.$$

B. Antisymmetric modes

Similar calculations are performed for antisymmetric modes. In this case, the scalar products of (2.4a) by ψ_n and (2.4b) by ϕ_n are performed and a system of equations on $\mathbf{U^a}$,$\mathbf{V^a}$ is obtained:

$$k^2\mathbf{U^a}+kA^a\mathbf{V^a}+B^a\mathbf{U^a}=0,$$
$$k^2\mathbf{V^a}+kC^a\mathbf{U^a}+D^a\mathbf{V^a}=0, \tag{3.8}$$

with matrices A^a, B^a, C^a, and D^a expressed by

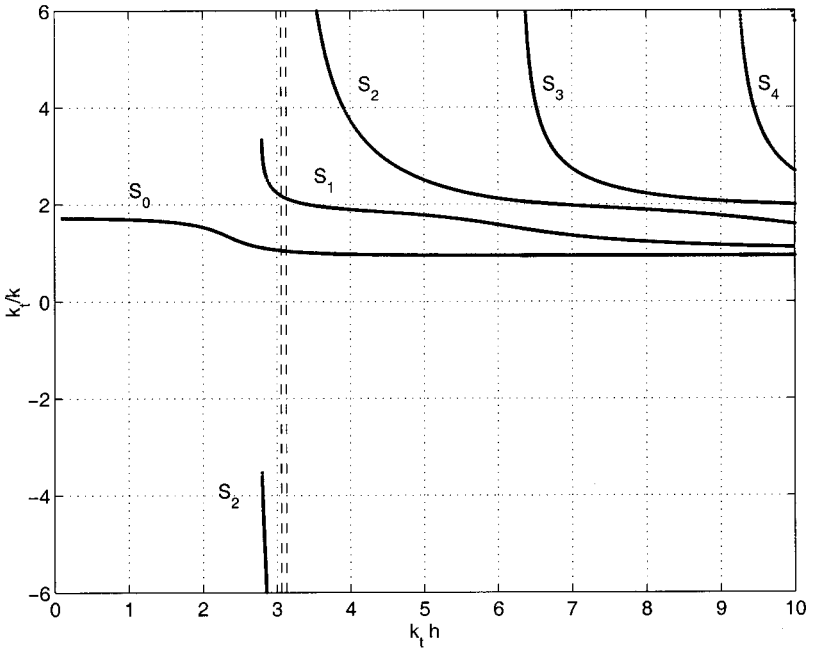


FIG. 3. Dimensionless phase velocities of symmetric Lamb waves k_t/k_n (for real k_n) as a function of the dimensionless frequency $k_t h$.

$$A^a{}_{m,n} = i\left(\frac{\gamma-1}{\gamma}(\phi_m|\psi_n') + \frac{2-\gamma}{\gamma}\phi_m(h)\psi_n(h)\right)$$

$$= \begin{cases} \dfrac{\sqrt{2}i(-1)^n}{h\gamma}, & m=1 \\ \dfrac{2i(-1)^{m+n}((\gamma-2)\alpha_m^2+\beta_n^2)}{h\gamma(\alpha_m^2-\beta_n^2)}, & m\geqslant 2 \end{cases}$$

$$B^a{}_{m,n} = \left(\frac{\beta_m^2}{\gamma} - k_l^2\right)\delta_{mn}, \tag{3.9}$$

$$C^a{}_{m,n} = -\gamma A^a_{n,m}, \quad D^a{}_{m,n} = (\gamma\alpha_m^2 - k_t^2)\delta_{mn}.$$

IV. RESOLUTION

The system (2.4) with boundary conditions (2.5) is an eigenvalue problem with differential operators. Owing to the spectral decomposition presented in Sec. III, it has become a discretized eigenvalue problem with matricial operators.

The discretized systems (3.6) and (3.8) are in the form of a nonlinear eigenvalue problem:

$$k^2\mathbf{U} + kA\mathbf{V} + B\mathbf{U} = 0, \tag{4.1a}$$

$$k^2\mathbf{V} + kC\mathbf{U} + D\mathbf{V} = 0, \tag{4.1b}$$

where matrices A, B, C, D result from projections of the original differential equations and also take into account the boundary conditions. System (4.1) can be easily expressed as a classical eigenvalue problem $(M-kI)\mathbf{X}=0$, as presented in the following section, Sec. IV A. In this case, for a given truncation corresponding to the first N basis functions, a $4N\times 4N$ system has to be solved to obtain $4N$ eigenvalues k. In Sec. IV B, it is shown that an alternative system can be derived, benefiting from the symmetry properties of the k spectrum; in this latter case, the system is only $2N\times 2N$, to also obtain $4N$ eigenvalues.

A. Eigenvalue problem

System (4.1) can be rewritten as

$$k^2\mathbf{X_1} + kF_1\mathbf{X_1} + G_1\mathbf{X_1} = 0, \tag{4.2}$$

with

$$\mathbf{X_1} = \begin{pmatrix}\mathbf{U}\\ \mathbf{V}\end{pmatrix}, \quad F_1 = \begin{pmatrix}0 & A\\ C & 0\end{pmatrix}, \quad \text{and} \quad G_1 = \begin{pmatrix}B & 0\\ 0 & D\end{pmatrix}. \tag{4.3}$$

Then, following Ref. 11, with $\mathbf{Y_1} = k\mathbf{X_1}$ and

$$\mathbf{Z_1} = \begin{pmatrix}\mathbf{X_1}\\ \mathbf{Y_1}\end{pmatrix}, \quad M_1 = \begin{pmatrix}0 & I_{2N}\\ -G_1 & -F_1\end{pmatrix}, \tag{4.4}$$

where I_{2N} is the $2N\times 2N$ identity matrix; the system (4.2) is rewritten as

$$M_1\mathbf{Z_1} - k\mathbf{Z_1} = 0. \tag{4.5}$$

It corresponds to a classical eigenvalue problem for the $4N\times 4N$ matrix M_1, in which the eigenvalue k appears linearly.

B. Reduction of the matrix dimension

By inspection, it can be noticed that system (4.1) possesses (fortunately!) the usual symmetries of the Lamb modes $k\rightarrow -k$ and $k\rightarrow k^*$. In order to reduce the dimension of the involved matrices, and, consequently, to increase numerical efficiency, it is possible to take advantage of the symmetry $k\rightarrow -k$. This can be done by casting (4.1) in the form of a nonlinear eigenvalue problem, where only the even powers of the eigenvalue appear.

Expressing $\mathbf{V}$ as a function of $\mathbf{U}$ in (4.1b), (4.1a) can be written as

$$(k^2 I_N - k^2 A(k^2+D)^{-1}C + B)\mathbf{U} = 0. \tag{4.6}$$

In (4.6), $\mathbf{U}\in \mathrm{Ker}(k^2 I_N - k^2 A(k^2+D)^{-1}C + B)$ and a solution $\mathbf{U}$ corresponds to k such that $\det(k^2 I_N - k^2 A(k^2+D)^{-1}C$

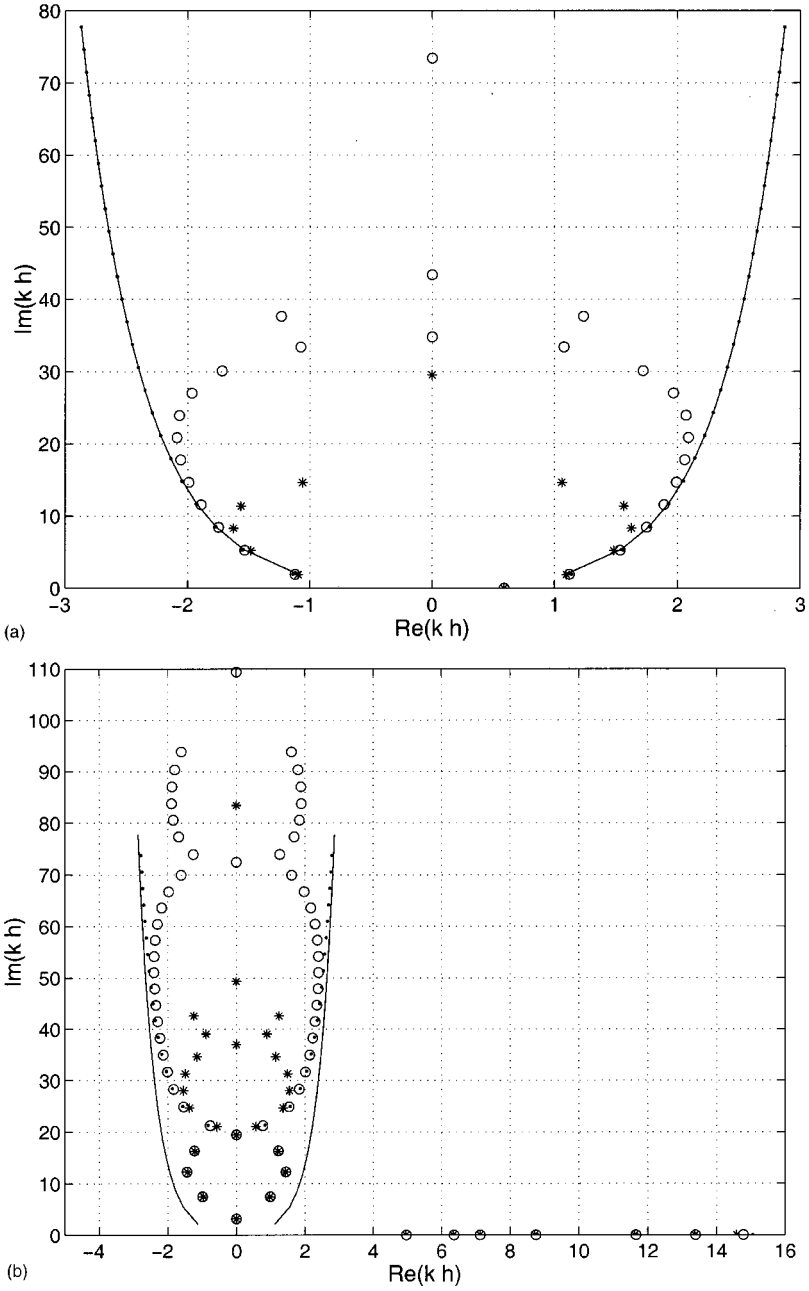


FIG. 4. Dimensionless complex Lamb wave spectrum at (a) $k_t h=1$: (*): $N=6$, (○): $N=14$, (·): exact values k^e and (—) asymptotic values k^a, (b) $k_t h=14$: (*): $N=16$, (○): $N=32$, (·): exact values k^e and (—) asymptotic values k^a, and (c) $k_t h=28$: (*): $N=24$, (○): $N=44$, (·): exact values k^e and (—) asymptotic values k^a.

$+B)=0$. If this equation is multiplied by $\det((k^2+D)A^{-1})$ on the left-hand side and $\det(A)$ on the right-hand side, we obtain $\det(K^2+(D-CA+A^{-1}BA)K+DA^{-1}BA)=0$, with $K=k^2$. The reduced eigenvalue problem on $K=k^2$ is

$$(K^2+F_2K+G_2)\mathbf{X_2}=0, \tag{4.7}$$

with $F_2=(D-CA+A^{-1}BA)$, $G_2=DA^{-1}BA$.

In the course of the derivation of the reduced system, it has been assumed that $\det(A)\neq 0$ and $\det(k^2+D)\neq 0$. The former assumption has been numerically verified. The latter assumption is verified as long as $k^2\neq k_t^2-\gamma\beta_n^2$.

As previously, we now introduce

$$\mathbf{Y_2}=K\mathbf{X_2}, \quad \mathbf{Z_2}=\begin{pmatrix}\mathbf{X_2}\\ \mathbf{Y_2}\end{pmatrix},$$

and

$$M_2=\begin{pmatrix}0 & I_N\\ -G_2 & -F_2\end{pmatrix},$$

where I_N is the $N\times N$ identity matrix. An eigenproblem for $K=k^2$ with the $2N\times 2N$ matrix M_2 is obtained:

$$M_2\mathbf{Z_2}-K\mathbf{Z_2}=0. \tag{4.9}$$

V. RESULTS

In order to check the validity and the efficiency of our technique, results obtained using the spectral method are presented. Without loss of generality, we will focus on the symmetric Lamb modes, but similar results can be obtained for antisymmetric modes. The material properties are those of

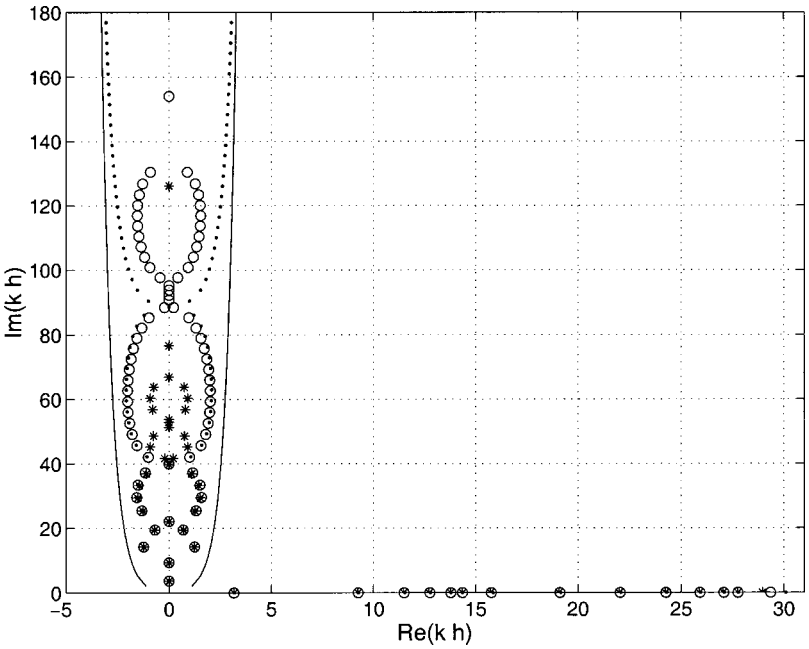


FIG. 4. (*Continued.*)

copper: $v_s=\sqrt{\mu/\rho}=2150\,\mathrm{m\,s^{-1}}$ and $v_l=\sqrt{(\lambda+2\mu)/\rho}=4170\,\mathrm{m\,s^{-1}}$ ($\gamma=(v_l/v_s)^2$) and the plate thickness is taken to be $2h=0.02\,\mathrm{m}$.

For u and v projected on N spectral basis functions, $4N$ eigenvalues are calculated from the $2N\times 2N$ system (4.9). For clarity, only the $2N$ rightgoing modes will be presented; the other $2N$ modes, which are leftgoing, are simply obtained by the symmetry $k\rightarrow -k$. On the other hand, as presented in the forthcoming section Sec. V C, a part of the calculated spectrum corresponds to spurious eigenvalues. For this reason, a qualitative criterium is used and only a subset of the determined spectrum is selected.

In the following, in order to assess the obtained values, we refer to "exact" values k^e, obtained from a Newton convergence method with a tolerance of 10^{-10}. Here, the wave numbers obtained from the spectral method are used as initial guess values and it has been verified that these "exact" values correspond to actual zeros of the dispersion relation.

A. Eigenvalues in the complex plane

Figure 2 shows the k spectrum obtained in the complex plane varying ω. In the computation, $N=16$ leads to 32 rightgoing eigenvalues k; as discussed in Sec. V C, only about N values are identified as correct values (11 eigenvalues are shown in the figure). The usual behavior of Lamb modes is recovered: for low frequency, only S_0 is propagating and, to increase the frequency leads to more and more propagating modes. We recover also the particular behavior of S_2: it becomes propagating with negative phase velocity and recovers a positive phase velocity at higher frequency.

B. Phase velocities of Lamb modes

To recover the usual representation of propagating modes,[5,14] we have plotted in Fig. 3 the dimensionless phase velocities of symmetric Lamb waves, k_t/k_n, for real k_n, as a function of the dimensionless pulsation k_th. This is obtained in Fig. 2 at constant $\mathrm{Im}(k)=0$. The expected form is obtained but we underline the fact that the branch $k<0$ of the S_2 mode corresponds to a rightgoing mode with negative phase velocity but positive group velocity. Incidently, it can be noticed that, in some papers, this negative phase velocity branch is erroneously identified as a part of an S_1 mode[5] or is not represented.[14] The S_2 mode turns purely imaginary in a frequency band corresponding to k_th between around 3.06 and 3.13, in agreement with previous studies (see, for instance, Rokhlin *et al.*[15]).

C. Representation of the spectrum for a given ω

Figures 4 show the evolution of the rightgoing complex wave spectrum derived from the spectral method when N increases. These $2N$ values are compared with the exact values k^e and the asymptotic values k^a derived by Merkulov *et al.*[6] for large k:

$$k_n^a h=\frac{1}{2}\ln\left[2\pi\left(n+\frac{1}{2}\right)\right]-\frac{i}{2}\left[\pi\left(n+\frac{1}{2}\right)-\frac{\ln[2\pi(n+1/2)]}{\pi(n+1/2)}\right]. \tag{5.1}$$

It can be seen from these figures that the k spectrum found with the spectral method coincides with the exact one (k^e) for the N or so first values, and this, independently of the complexity of the spectral structure when the frequency increases. A qualitative criterion to select the useful part of the calculated spectrum can be to restrict the complex wave spectrum to the first N values.

D. Convergence

The evolution of the relative error is shown in Fig. 5 as a function of N at two different frequencies. The relative error is defined as $|\Delta(k)/k|$, where $\Delta(k)=k-k^e$. In both cases, it appears that the method converges as $1/N$ for large N.

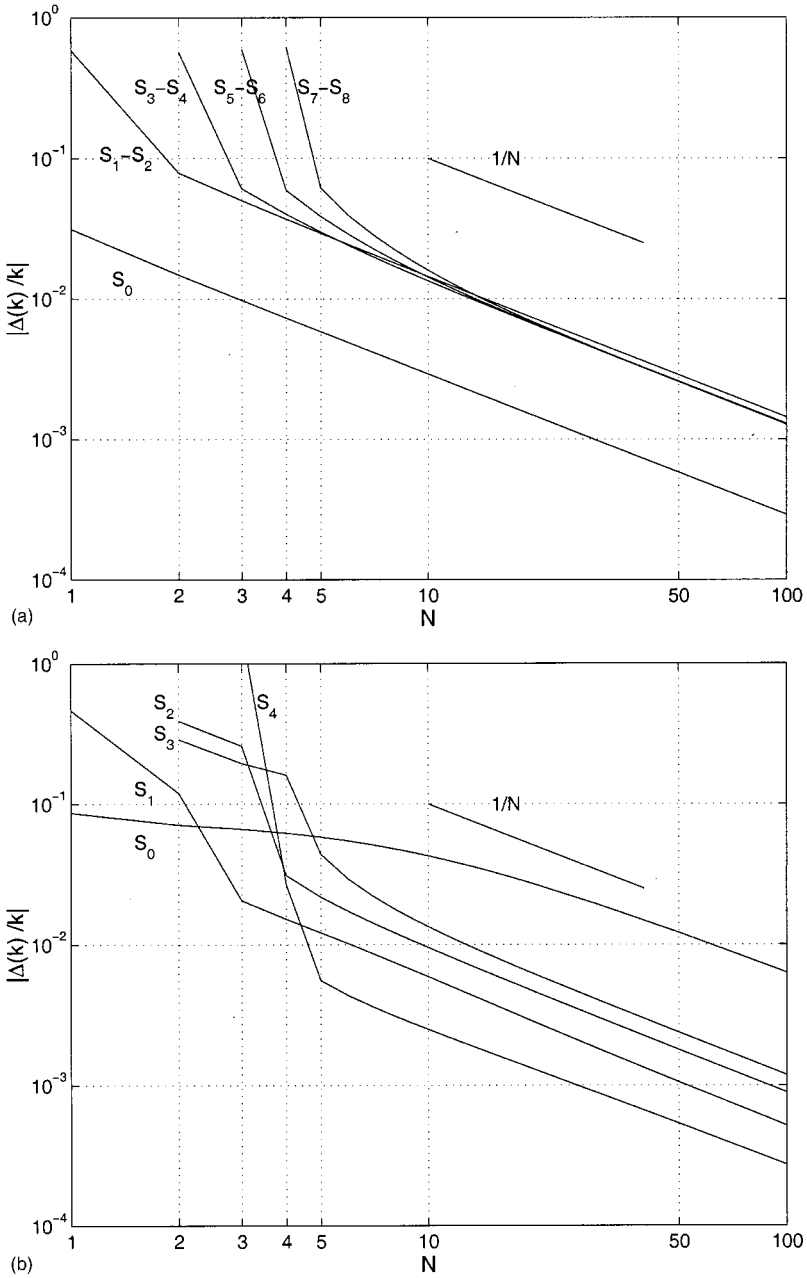


FIG. 5. Convergence for the first modes as a function of the order N of the truncation for a given frequency, (a) $k_t h=1$ (modes S_0 to S_8) and (b) $k_t h=14$ (modes S_0 to S_4).

For $k_t h=1$, only the mode S_0 is propagating; the symmetry properties of the spectrum for evanescent modes implies $k_{2n}=-k^*_{2n-1}$ for $n\geqslant 1$, implying the same convergence for the pairs $(2n,2n-1)$.

For $k_t h=14$, 14 modes are propagating; we give the convergence for the first five modes. It can be noticed that the mode S_0 reaches a type $1/N$ convergence law only for $N>10$, but is given with a reasonable accuracy (10%) as soon as $N=1$.

VI. CONCLUSION

A new method for the determination of the Lamb wave spectrum has been presented. This method is based on a spectral projection of the equation of the elasticity, leading to a classical eigenvalue problem.

This method is an alternative to the usual method of root finding of the Rayleigh–Lamb dispersion relation. The following points make the method attractive.

(i) It is easy to implement. For a given frequency, it is very simple to take the expression of the M_2 matrix from (3.9), (4.10), and (4.11), to put in it the material (γ, k_l, and k_t) and geometrical (h) properties and then, to use any eigenvalue solver package to obtain the wave numbers.

(ii) In a step by step method, the series of spectra calculated from zero to a given frequency can cross a critical frequency, for which two wave numbers collapse, leading to $\partial_k D=0$. A particular treatment then has to be applied to go through this critical frequency since $\partial k/\partial\omega=\partial_\omega D/\partial_k D$ diverges. With our method, these critical frequency cases have bearing on the determination of the spectrum because its calculation does not depend on the history of the spectrum.

(iii) If the goal is to obtain the wave numbers with a prescribed precision, the results of the spectral method can

be used as very good initial guess values in a Newton–Raphson method. That is to say that our method can be viewed as a ''super initial guess value provider.''

(iv) The method can be easily generalized to other wave guided modes. In fact, it can be implemented as long as the transverse problem can be written in the form of a differential equation with an integer power of the eigenvalue in the coefficients. Then, this technique offers the possibility of much greater certainty in finding all the families of transverse modes. For instance, it may be applied to fluid-loaded plates, plates with damping, or transversally layered plates.

[1] D. Ensminger, *Ultrasonics, Fundamentals, Technology, Applications* (Marcel Dekker, New York, 1988).

[2] R. D. Gregory and I. Gladwell, ''The reflection of a symmetric Rayleigh–Lamb wave at a fixed or free edge of a plate,'' J. Elast. **13**, 185–206 (1983).

[3] R. D. Gregory and I. Gladwell, ''The generation of waves in a semi-infinite plate by a smooth oscillating piston,'' ASME J. Appl. Mech. **51**, 787–791 (1984).

[4] C. Scahdrett and N. Vasudevan, ''The propagation of time harmonic Rayleigh–Lamb waves in a bimaterial plate,'' J. Acoust. Soc. Am. **89**, 1606–1614 (1991).

[5] I. A. Viktorov, *Rayleigh and Lamb Waves: Physical Theory and Applications* (Plenum, New York, 1967), Chap. 2.

[6] L. G. Merkulov, S. I. Rokhlin, and O. P. Zobnin, ''Calculation of the spectrum of wave numbers for Lamb waves in a plate,'' Sov. J. Nondestruct. Test. **6**, 369–373 (1970).

[7] J. Miklowitz, *The Theory of Elastic Waves and Waveguides* (North-Holland, New York, 1978), Chap. 4.

[8] J. D. Achenbach, *Wave Propagation in Elastic Solids* (Elsevier, Amsterdam, 1984).

[9] R. D. Mindlin, ''Waves and vibrations in isotropic elastic plates,'' *Structural Mechanics*, edited by J. N. Goodier and N. J. Hoff (Pergamon, New York, 1960).

[10] P. G. Drazin and W. H. Reid, *Hydrodynamic Stability* (Cambridge University Press, Cambridge, 1982).

[11] T. J. Bridges and P. J. Morris, ''Differential eigenvalue problems in which parameter appears non-linearly,'' J. Comput. Phys. **55**, 437–460 (1984).

[12] P. M. Morse and H. Feshback, *Mathematical Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).

[13] V. Pagneux, N. Amir, and J. Kergomard, ''A study of wave propagation in varying cross-section waveguides by modal decomposition. Part I. Theory and validation,'' J. Acoust. Soc. Am. **100**, 2034–2048 (1996).

[14] Y. Cho and J. L. Rose, ''A boundary element solution for a mode conversion study on the edge reflection of Lamb waves,'' J. Acoust. Soc. Am. **99**, 2097–2109 (1996).

[15] S. I. Rokhlin, D. E. Chimenti, and A. H. Nayfeh, ''On the topology of the complex wave spectrum in a fluid-coupled elastic layer,'' J. Acoust. Soc. Am. **85**, 1074–1080 (1989).

Theory and experiments on the coupling of two Helmholtz resonators

Torbjörn A. Johansson and Mendel Kleiner
Department of Applied Acoustics, Chalmers University of Technology, Sven Hultins gata 8A, SE-412 96 Gothenburg, Sweden

(Received 11 July 2000; revised 7 June 2001; accepted 25 June 2001)

The acoustic coupling between closely located acoustic resonators can have a significant effect on the resonance properties. Examples of the influence of coupling include pipe musical instruments such as flue organs or flutes, where the coupling may affect the sound. Here, a physical model of a system of two coupled Helmholtz resonators is developed. The coupling is described by the mutual impedance between the neck openings, and an additive reflection impedance modification to the radiation impedance, which models the effect of reflections from the outer ''shell'' of the other resonator. The theoretical predictions show a good agreement with measurements also reported here. Finally, the effects of coupling on the present system are analyzed. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1394741]

PACS numbers: 43.20.Ks, 43.20.Wd, 43.20.Rz, 43.75.Np, 43.75.Qr [ANN]

I. INTRODUCTION

A. The importance of coupling

Coupling between acoustic resonators in speaking and silent mode is important in many technical systems. Such resonant systems may be found in musical instruments, loudspeakers, ventilation ducts, as well as in exhaust systems, for example.

In some musical instruments, the coupling may lead to undesirable detuning and possible beat effects. These phenomena lead to increased workload for instrument builders and tuners since an interactive building approach is then needed to correct for the coupling, which cannot be estimated beforehand, but is obvious when the instrument has been assembled. As an example, the possible importance of the coupling in flue organs is illustrated by Fig. 1, which is a picture of the interior of a large pipe organ.

It should be mentioned here that the ear is particularly sensitive to small changes in frequency, which means that even effects of very weak coupling can be audible. In this paper, the term pitch, which is really a subjectively perceived quantity related to frequency, will be used to denote measured or predicted resonance frequencies.

B. Coupling between acoustic resonators

The case of acoustic coupling between resonators has not been profoundly analyzed. Some work has been done on general coupling between resonant systems, but the coupling here is generally a simple linear spring force (see Refs. 1 and 2), although research has also been done on weak coupling of a more general kind.[3] These methods are based on lumped circuit modeling of the coupling, which implies that what happens at one resonator immediately affects the other, implying that the path time between the resonators is not included in the model. This would be a reasonable approximation at the shorter of the resonator separations used here, but not at the longer ones. Also, here the coupling is often too strong for the weak coupling assumptions to hold. It would, however, be possible to employ the weak coupling assumption for the larger resonator separations. Consequently, to develop a general coupling model that holds for all studied configurations, none of the above assumptions is made here.

The mutual acoustic impedance between radiating surfaces of different simple shapes, such as rectangular or circular pistons, has been studied extensively both analytically and numerically; see for instance Refs. 4 and 5. The mutual impedance between radiators is also important in electromagnetic engineering, for example in antenna design.[6]

Mutual acoustic impedance between resonator orifices in simple arrays, and its effect on the resonance properties, has been qualitatively studied in Refs. 7 and 8, but no detailed analysis has been performed. Ingard[7] found that the coupling decreases the maximum absorption of each resonator. A radiation impedance approach to the problem was taken by Mechel,[8] who found that an array of identical resonators absorbs in a wider frequency band than a single resonator.

The influence of reflections from infinite planes in the vicinity of a source has been modeled by the mutual acoustic impedance between the radiator and its image source (reflected in the plane);[9] otherwise, not much work has been done on the influence or reflections. Here, the influence of reflections is modeled as an equivalent additive contribution to the radiation impedance.

In this paper, investigation of the effects of two-resonator coupling for different resonator separations and uncoupled pitch differences is performed. A theory, based on previous research on Helmholtz resonators and a novel coupling model, is described and compared with measurement results. Finally, descriptions of the coupling effects are given, together with attempted physical explanations.

II. BACKGROUND

The findings reported here represent the first step in a study of the coupling between close flue organ pipes. Organ pipe coupling effects include pitch drift, output sound-pressure level change, and alteration of system losses. These

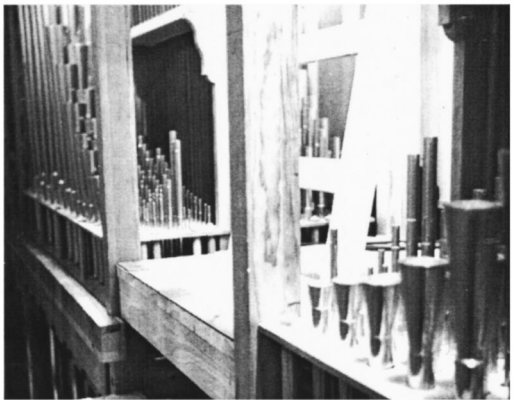


FIG. 1. A pipe organ may contain many thousands of acoustically coupled resonators.

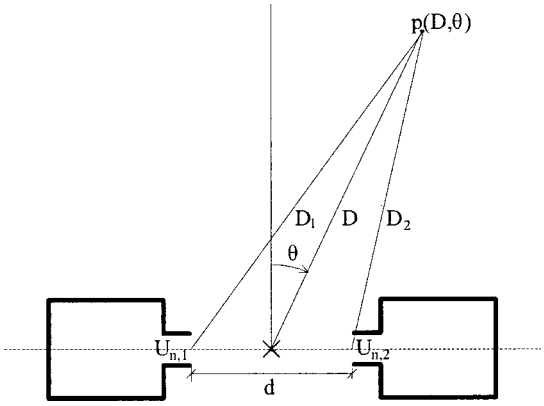


FIG. 3. The measurement setup geometry definition for the coupled resonator measurements.

effects can significantly change the audible impression of the organ sound. It is for instance believed that the differences in organ pipe coupling account for some of the differences in sound between new and older (baroque) organs.

There are many complicated issues in flue organ pipe coupling, for example the jet drive, the possible interaction of the driving air jets, and the large number of pipes that may couple, and it was therefore decided to start this study by investigating coupling between simpler, nonjet-driven resonators.

A flue organ pipe uses a quarter- or half-wavelength resonator, and thus it would seem natural to use the same kind of resonator in this initial study. However, the Q value of the main resonance is typically about an order of magnitude higher in Helmholtz resonators than in quarter-wavelength resonators, which means a much better signal-to-noise ratio around resonance and thus a facilitated measurement of resonance properties.

There are, of course, some very important differences between an organ pipe and a Helmholtz resonator, primarily in the number of openings (although there are organ pipes with covered top ends), the resonator drive, and the harmonic spectrum. However, as a first-order approximation, the mouth and top of a silent organ pipe are so much smaller than the wavelength that they will act as monopole sinks, just like the neck opening of a Helmholtz resonator at low frequencies. This means that organ pipe coupling will resemble Helmholtz resonator coupling, even though the speaking pipe/resonator does not function in the same way.

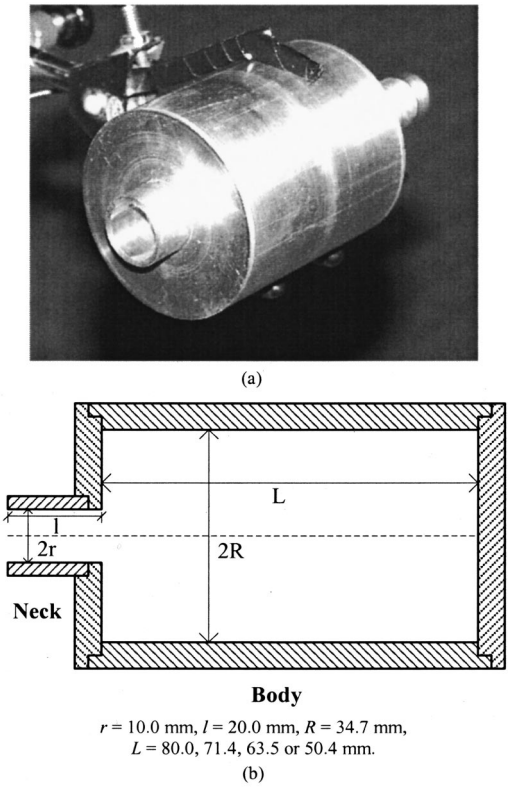


FIG. 2. Cylindrical aluminum Helmholtz resonator for measurements. (a) Photo. (b) Parameter definitions and values.

III. EQUIPMENT

A special set of detachable aluminum Helmholtz resonator ''bottles,'' see Fig. 2, was made for the experiments. Four different body parts (of different lengths but otherwise the same) were made to allow variation of the pitch. The body length L and internal radius R are defined in Fig. 2, as is the neck length l and inner radius r.

The resonators are designed to allow simple calculations of acoustic parameters, particularly the neck-end corrections and the mutual impedance. They are made from cylindrical aluminum pipes and plates, of thickness 2 mm or more. Because of the thick material, the influence of structural modes is neglected.

The body and neck are cylindrical, and the bottom and top parts are flat. For fine-tuning of the pitch, a screw of diameter 20 mm and a small rise per turn is threaded into the bottom of each resonator. A miniature electret condenser microphone is also placed in the bottom plate of each resonator. The influence of the (short) screw and the microphone is neglected in all the analytical calculations. The screw's maximum effect on the pitch is 1.5 Hz, as found in both measurements and simulations. The microphone is so small that its influence on the pitch is less than 0.1 Hz.

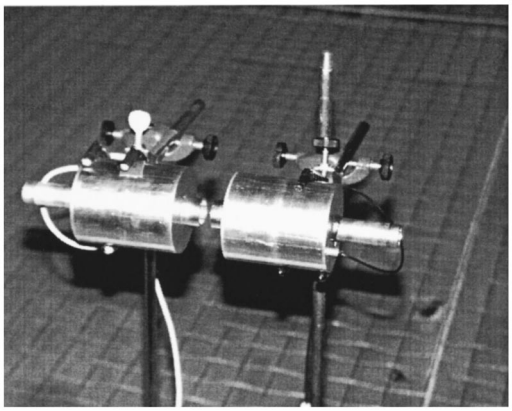


FIG. 4. Close-up of the resonator arrangement for coupled resonator measurements.

IV. MEASUREMENTS

Reciprocity measurements are used throughout this work. Pseudorandom ''white'' noise, filtered to the frequency range of interest, is sent to the Dynaudio 17W-75 EXT loudspeaker, and small microphones in the bottom of each resonator record the resulting sound pressure. The MLSSA measurement system (DRA Laboratories, Sarasota, FL) is used to record system responses. Using the MLS method, the response is then deconvolved by the output signal to form the impulse response of the system.

The measurement setup is shown in Fig. 3, where, for clarity, the resonator separation is considerably exaggerated. Here, D_1 and D_2 are the distances between the measurement point and the corresponding neck openings; d is the resonator separation; θ is the receiver angle; and $U_{n,1}$ and $U_{n,2}$ are the neck volume velocities. The resonators are mounted on tripods, standing on a computer-controlled turntable. Because the tripods are about 1 m high, the influence of reflections from the turntable is small, and it is neglected here. The acoustical center of the loudspeaker, used to excite the resonators, is 1.12 m away from the center point of the two resonators. This center point, indicated by the cross in Fig. 3, is on the axis of the turntable, and is therefore fixed in space. To record the radiation pattern, the turntable is rotated 5 deg between each measurement.

The resonators are arranged along a common symmetry axis, as shown in Fig. 3 and Fig. 4. To investigate the effects of different coupling strengths, measurements are carried out for resonator separations d of 1, 2, 4, 8, and 16 cm, and the impulse responses are recorded for both the lower- and the higher-pitch resonator. The data are saved for subsequent analysis using MATLAB®. To obtain the frequency responses, the impulse responses are Fourier transformed.

The resonant frequency of the lower-pitch resonator is kept constant, while the pitch of the other resonator is varied by changing the body parts. As previously mentioned, there are four different body parts. For easy comparison with organ pipe measurements not reported here, the length L of these is chosen so as to make the second resonator resonant at the same frequency as, a semi-tone above, a whole tone above, and a major third above the reference resonator, respectively. That is, in units of the pitch f_1 of the reference resonator, the other resonator is resonant at f_1, $2^{1/12}f_1$, $2^{1/6}f_1$, and $2^{1/3}f_1$, respectively, with an error less than 0.029% (corresponding to less than 0.5 musical cents). Throughout this work, the modal circle fit algorithm is employed to find resonance parameters.

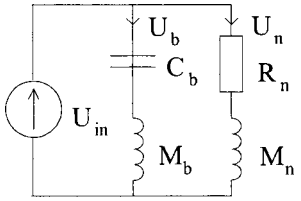


FIG. 5. Acoustic impedance model of a single resonator.

To compensate for the frequency dependence of the loudspeaker volume velocity, the sound pressure at the center point of the two resonators is measured in free field, i.e., the resonators are removed before the measurement. Assuming monopole radiation, the volume velocity U_{sp} of the loudspeaker is calculated from these measurement data. As the monopole sound pressure is proportional to U_{sp}, the measured frequency responses can then easily be normalized to a constant U_{sp}, which allows easy comparison to theory. This normalization technically yields a quantity with impedance units. However, impedance refers to the relation between sound pressure and volume velocity at the same measurement point, and therefore our quantity is not an impedance. It should be regarded as the resulting sound pressure, given a certain constant source volume velocity at a certain point in space. Here, the source volume velocity U_{sp} is (arbitrarily) set to $10^{-4}\,\mathrm{m}^3/\mathrm{s}$.

V. THEORY

A. Reciprocity

A loudspeaker on the outside and microphones on the insides of the resonators are used in the measurements. When deriving the equivalent acoustical impedance analogy circuit for the coupled resonators, it simplifies matters to put the sources inside the resonators and the microphone on the outside. By reciprocity this gives the same results. Note that reciprocity only allows a source in *one* of the resonators; therefore, using this approach, we can only study the coupling of one speaking and one silent resonator, and not the coupling between two speaking resonators. The terms ''speaking'' and ''silent,'' borrowed from organ pipe terminology, are used throughout this paper to distinguish between a resonator excited by a source on the inside and a nonexcited one, respectively.

B. Single resonator with losses

First, the model for the neck volume velocity of a single resonator in free field is derived. In an acoustical impedance analogy, the Helmholtz resonators used here can, near the fundamental mode, be described by the lumped circuit model in Fig. 5. In this model, the volume velocity source U_{in} is, as mentioned, inside the resonator. Alternatively, the resonator can be modeled using a transfer matrix approach. This could possibly result in a need for fewer approximations and an

improved accuracy, but it would also result in a model that is far less intuitive. As priority here is given to intuitivity, and because predictions derived from the present model are later shown to agree reasonably well with measurements, the lumped circuit model is used in what follows.

In Fig. 5, the parameters used are the volume velocities U_b in the resonator body and U_n in the neck; body compliance and mass C_b and M_b; and finally, neck resistance and mass R_n and M_n. The radiation impedances at the inner and outer end of the neck are included as end corrections in the neck impedances.

The lumped circuit body impedances are, to the first approximation

$$C_b=\frac{\pi R^2 L}{\rho_0 c^2}, \quad M_b=\frac{1}{3}\frac{\rho_0 L}{\pi R^2}, \tag{1}$$

where C_b is the acoustical compliance of the air in the body, and M_b is a cylindrical resonator body air mass correction derived by Panton and Miller.[10] M_b is equal to one-third of the acoustical mass of the body air. In using M_b, the influence of the adjusting screw and microphone is neglected. In Eq. (1), c is the speed of sound in air; ρ_0 is the density of air; and R and L are the body length and radius, respectively, as defined in Fig. 2. Numerical values for ρ_0 and c are given in Table I. The losses inside the body are neglected since they can be expected to be much smaller than the losses at the neck.[11]

M_n is the acoustic mass of the air in the neck, including the end correction l'. It is given by

$$M_n=\frac{\rho_0(l+l')}{\pi r^2}. \tag{2}$$

The end correction l' is the sum of the end corrections at the inner and outer neck ends. An approximate equation for the inner-end correction, derived from modal theory, is given by Kergomard and Garcia.[12] Using this equation, for the present configuration, the numerical value of the inner-end correction is

$$l'_{\text{inner}}=0.509r. \tag{3}$$

The outer-end correction of

$$l'_{\text{outer}}=0.68r, \tag{4}$$

is obtained from end-correction data measured by Peters *et al.*[13] for an open pipe end. The numerical value of Eq. (4) is arrived at by linearly interpolating between the end correction values for inner- to outer neck radius ratios of 0.70 and 0.85, thus arriving at the end correction for the present ratio of 0.80. Strictly, the influence of the body part of the resonator will distort the flow from that found around an open pipe. The body can be seen as an expanded part of the pipe, restricting the flow to a narrower region in space, which implies that its presence must increase the equivalent amount of moving air outside the neck, i.e., the mass end correction. The amount of increase is difficult to estimate without the use of numerical flow modeling. However, the body radius is much smaller than the wavelength at the frequencies of interest ($kR<0.3$), which means that its effect on the flow is small. To at least estimate the order of a pitch shift due to possible errors in Eq. (4), this pitch shift is certainly much smaller than in the extreme case of a baffled pipe, which leads to an end correction of $0.8127r$,[13] and a pitch shift of up to 2% compared to when Eq. (4) is employed.

TABLE I. Physical properties of air at 20 °C and 760-mm Hg.

Density	ρ_0	kg/m^3	1.205
Speed of sound	c	m/s	343.3
Dynamical viscosity	μ	Ns/m^2	1.88×10^{-5}
Ratio of specific heats	γ		1.4
Heat capacitivity at constant pressure	c_p	J/kgK	1010
Thermal conductivity	v	W/mK	0.026

The neck resistance R_n represents the resonator losses

$$R_n=\frac{l}{r}\frac{\sqrt{2\mu_{\text{eff}}\rho_0\omega}}{\pi r^2}+2\frac{\sqrt{2\mu_0\rho_0\omega}}{\pi r^2}+\frac{\pi\rho_0}{c}f^2, \tag{5}$$

where

$$\mu_{\text{eff}}=\mu_0\left(1+(\gamma-1)\sqrt{\frac{\nu}{\mu_0 c_p}}\right)^2. \tag{6}$$

In Eq. (5), μ_0 is the dynamical viscosity of air (numerical value given in Table I); f is the frequency; and ω is the angular frequency. In Eq. (6), μ_{eff} is an effective coefficient of viscosity, taking the heat conductivity of the neck walls and the resulting energy loss into account;[14,15] γ is the ratio of specific heats; ν is the coefficient of thermal conductivity; and c_p is the specific heat capacity at constant pressure.

R_n includes several contributions. The first right-hand term in Eq. (5), adapted from Ingard,[7] represents the viscous losses in the neck wall boundary layer. R_n's second term, also adapted from Ingard,[7] represents the viscous losses at the neck ends. This approximate expression can be used both for flanged (in infinite baffle) and unflanged (free) neck ends.[16] As neither the inner- nor the outer neck end is strictly neither flanged nor unflanged, the use of it here is another approximation. Conditions at the outer end can be expected to be somewhere between the flanged and the unflanged case. At the inner end the influence of the finite body radius should be taken into account, but this is probably not possible by analytical methods, and it is not done here.

It should be noted here that the expression for the second term in Eq. (5) was originally arrived at by empirical methods, and its validity is not supported by analysis. New measurements are needed to validate it, or arrive at a better one. However, this is out of the scope of the present paper. The present expression, being the only available approximation, is therefore used. To illustrate the sensitivity of the predicted results to this second term of Eq. (5), a 10% change in this term gives a Q value change of up to 4%. As will be seen in the Results section, this is on the order of the discrepancy between measured and predicted Q values; therefore, it can be stated that the discrepancies likely may be caused mostly by errors in the uncertain estimation of the viscous losses at the neck ends.

The (monopole) radiation losses at the outer neck end, calculated for the low-frequency limit of a piston in a long

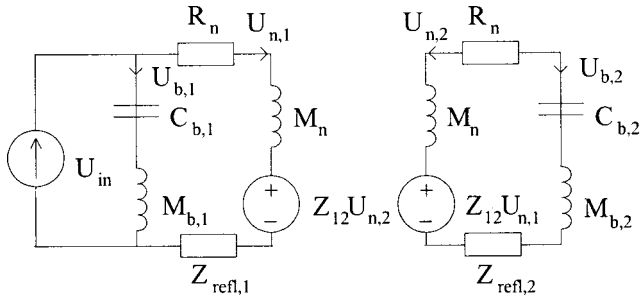


FIG. 6. Acoustic impedance model of two coupled Helmholtz resonators.

tube,[16] are represented by the final term of R_n. The dominant loss mechanisms for a low-frequency Helmholtz resonator driven at low amplitudes are the viscous and radiation losses. In the present system and at the low frequencies of interest (approximately 240 to 440 Hz), the viscous losses are 2–6 times bigger than the radiation losses. Losses in the body part of the resonator, primarily thermal losses to the walls and viscous losses in the boundary layers, were found to be several orders of magnitude smaller than those in the neck. Consequently, they are not included in the model.

Returning to Fig. 5, an equation for the neck volume velocity U_n is desired. For simplicity, the total neck and body impedances are defined as in Eq. (7)

$$\begin{aligned} Z_n &= R_n + j\omega M_n \\ Z_b &= j\omega M_b + 1/j\omega C_b. \end{aligned} \tag{7}$$

The frequency response of a single resonator is now obtained from Fig. 5 as

$$U_n(\omega) = \frac{Z_b}{Z_b + Z_n} U_{in}, \tag{8}$$

using volume velocity division between the body and the neck.

C. Dual resonator with losses and coupling

For the system of two coupled resonators, Fig. 6 shows the equivalent acoustic impedance analogy for calculation of the neck velocities. In this figure, the speaking resonator and all corresponding impedances are labeled ''1,'' and the silent resonator is labeled ''2.''

There are two different coupling effects in the present system. The sound field at the neck of one resonator is affected by, on one hand, the radiation from a source at the neck opening of the other resonator, and, on the other hand, the altered radiation conditions due to reflections off the outer shell of the other resonator.

The influence of the source at the other resonator is modeled by calculating the equivalent sound-pressure contribution at the present resonator's neck opening. The mutual impedance Z_{12} is used to model how a volume velocity in the neck of one resonator causes a sound pressure on the neck opening of the other resonator. Consequently, at the neck opening of resonator 1, the equivalent source modeling the influence of the radiation from resonator 2 is $Z_{12}U_2$, and vice versa.

The influence of sound waves reflecting off the outer shell of the other resonator is modeled by modifying the radiation impedance at the outer neck end. For clarity, this additive modification is shown as a separate impedance in Fig. 6.

Because the only difference between the two resonators is the different body length L, the neck impedance Z_n is the same in both resonators. If the body lengths L differ, the body impedances $Z_{b,1}$ and $Z_{b,2}$ are different [see Eq. (1)], as are the reflection impedances $Z_{\text{refl},1}$ and $Z_{\text{refl},2}$ [the equivalent sphere radius a of Eq. (22) depends on L]. In Eq. (9), a total neck impedance is defined as the sum of the ''uncoupled'' neck impedance from Eq. (7) and the reflection impedance

$$\begin{aligned} Z_{n,1} &= Z_n + Z_{\text{refl},1} \\ Z_{n,2} &= Z_n + Z_{\text{refl},2}. \end{aligned} \tag{9}$$

Using Fig. 6 and Eq. (9), equations for the volume velocities are given in Eq. (10)

$$U_{n,1}Z_{n,1} = U_{b,1}Z_{b,1} - U_{n,2}Z_{12}, \tag{10a}$$

$$U_{n,2}Z_{n,2} = U_{b,2}Z_{b,2} - U_{n,1}Z_{12}, \tag{10b}$$

$$U_{b,1} + U_{n,1} = U_{in}, \tag{10c}$$

$$U_{b,2} + U_{n,2} = 0. \tag{10d}$$

Equations (10a) and (10b) express the equality of sound pressures on the parallel neck and body branches in the Fig. 6 model. Equations (10c) and (10d) express the conservation of mass in resonators 1 and 2, respectively.

Defining

$$\begin{aligned} Z_1 &= Z_{n,1} + Z_{b,1} \\ Z_2 &= Z_{n,2} + Z_{b,2}, \end{aligned} \tag{11}$$

the solution of Eq. (10) is obtained as

$$\begin{aligned} U_{n,1} &= U_{in} Z_{b,1} \frac{Z_2}{Z_1 Z_2 - Z_{12}^2} \\ U_{n,2} &= U_{in} Z_{b,1} \frac{-Z_{12}}{Z_1 Z_2 - Z_{12}^2}. \end{aligned} \tag{12}$$

The mutual impedance between the resonators is defined as the average sound pressure on one of the neck openings, generated by a unit volume velocity on the other.

To be able to deduce analytical approximations of the mutual impedance and the reflection impedance, and also to derive an equation for the receiver sound pressure, a monopole source action is assumed. This is a plausible assumption at the low frequencies of interest, but at the smallest resonator separation of 1 cm its validity needs to be investigated further. Measurements of the distance dependence of the sound-pressure level show that at the frequencies of interest, the neck opening of a resonator acts as a monopole at distances larger than 2 cm. The influence of the source size can be observed closer to the neck opening. In order to keep things simple, this is neglected, and throughout this work it is assumed that the resonator neck openings act as perfect monopole sources. The body and the neck opening of the other resonator both act as low-pass filters in space, decreas-

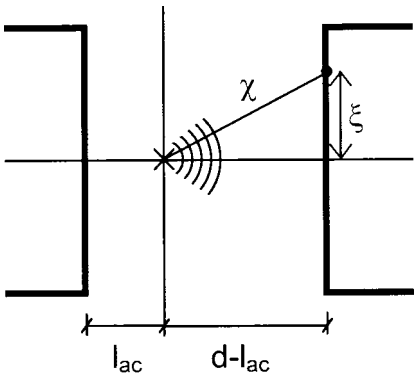


FIG. 7. Setup for calculation of the mutual impedance.

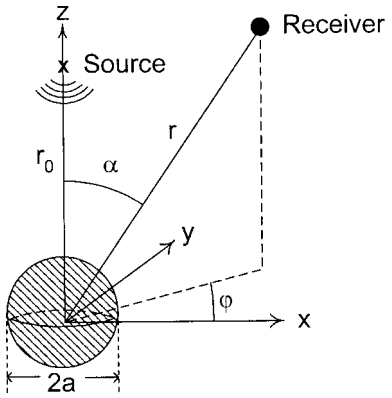


FIG. 8. Setup for calculation of the reflection impedance.

ing the influence of the near-field sound-pressure fluctuations, which makes the approximation reasonable, even at the smallest resonator separation.

According to Ando,[17] the radiation from a circular pipe end can, in the far field, be modeled as that of an imagined monopole source at the acoustical center of the radiating pipe end. Ando[17] gives graphs of the distance l_{ac} between the neck opening and the acoustical center for certain ratios of inner to outer pipe radius. Technically, the far field does not extend down to the smallest resonator separation, but this is neglected here, and a monopole source action is, as stated, assumed. To find the location of the acoustical center in the present case, Ando's data are linearly interpolated between inner- to outer-neck radius ratios of 0.70 and 0.85, giving the numerical value of

$$l_{ac}=0.85r. \tag{13}$$

The geometry for calculation of the mutual impedance is shown in Fig. 7. The sound pressure radiated from a monopole source of volume velocity U_0 at the acoustical center (the cross in Fig. 7), evaluated at a distance χ, is

$$p(\mathbf{r})=\frac{j\omega\rho_0}{4\pi}U_0\frac{e^{-jk\chi}}{\chi}, \tag{14}$$

where k is the wave number, $k=2\pi/\lambda$, λ being the acoustical wavelength. In Fig. 7, l_{ac} is the distance between the neck opening of the speaking resonator and the acoustical center; and ξ is the radial distance of the measurement point from the symmetry axis. From Fig. 7, $\chi=\sqrt{(d-l_{ac})^2+\xi^2}$. To calculate the mutual impedance, the sound pressure is averaged over the neck opening S of the other resonator and then divided by the source strength, as in Eq. (15)

$$\begin{aligned}Z_{12}=\frac{\langle p\rangle}{U_0}&=\frac{j\omega\rho_0}{4\pi}\frac{1}{\pi r^2}\int_S\int\frac{e^{-jk\chi}}{\chi}dS\\&=\frac{j\omega\rho_0}{2}\frac{1}{\pi r^2}\int_0^r\frac{e^{-jk\sqrt{(d-l_{ac})^2+\xi^2}}}{\sqrt{(d-l_{ac})^2+\xi^2}}d\xi\\&=\frac{c\rho_0}{2\pi r^2}(e^{-jk(d-l_{ac})}-e^{-jk\sqrt{(d-l_{ac})^2+r^2}}).\end{aligned} \tag{15}$$

In order to deduce an analytical approximation of the reflection impedance, the speaking resonator neck opening is assumed to act as a monopole radiator. There is an analytical solution for the sound pressure radiated from a piston in a sphere, a case that could be applicable here and might closer approximate the actual conditions, but fulfilling the boundary conditions on the reflecting resonator using this approach would require numerical solution. At the low frequencies of interest, the comparative gain of this approach would probably be small, and it is not pursued.

Because the dimensions of the resonators are much smaller than the wavelengths of interest, the individual shape of a resonator is only of slight importance to the scattering. Therefore, the scattering off a resonator does not differ much from the scattering off a rigid sphere of the same volume and the same center of gravity as an imagined homogeneously filled resonator. This rigid-sphere model of the other resonator's scattering properties is used here.

The scattering of a spherical wave incident on a rigid sphere of radius a is calculated in Ref. 18, using the setup in Fig. 8. The source is at $\mathbf{r}_0=(r_0,\alpha_0,\varphi_0)$; the measurement point is at $\mathbf{r}=(r,\alpha,\varphi)$; and the sphere is centered at the origin.

The sound pressure incident on the sphere is thus given by

$$p^{\text{inc}}(\mathbf{r})=\frac{j\omega\rho_0}{4\pi}U_0\frac{e^{-jk|\mathbf{r}-\mathbf{r}_0|}}{|\mathbf{r}-\mathbf{r}_0|}, \tag{16}$$

where U_0 is the source volume velocity. The scattered sound pressure is obtained as[18]

$$\begin{aligned}p^{sc}(\mathbf{r})=&\frac{\rho_0\omega^2}{\pi c}U_0\sum_{l=0}^{\infty}h_l(kr)\frac{j_l'(ka)}{h_l'(ka)}h_l(kr_0)\\&\times\sum_{m=0}^{l}\varepsilon_m S_l^m(\cos\alpha)S_l^m(\cos\alpha_0)\cos m(\varphi-\varphi_0),\end{aligned} \tag{17}$$

where $h_l(x)$ and $j_l(x)$ are spherical Hankel functions of the first kind and spherical Bessel functions, respectively. The primes indicate derivation with respect to the entire argument. In the summation over m, ε_m is the Neumann symbol given by Eq. (18)

$$\varepsilon_m=\begin{cases}1, & m=0\\ 2, & m\neq 0\end{cases}. \tag{18}$$

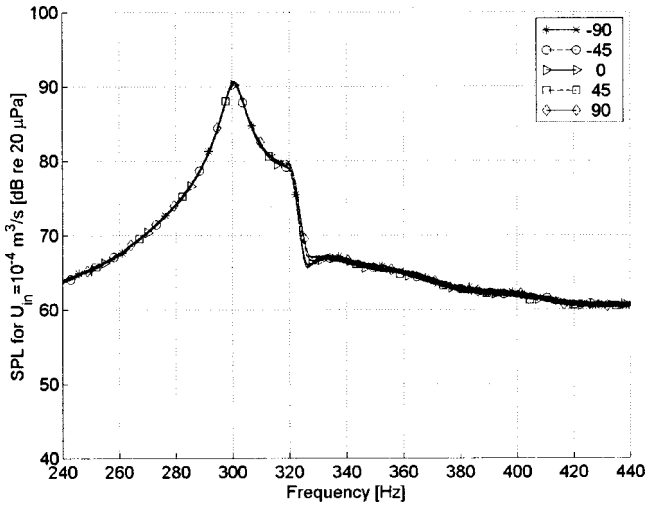


FIG. 9. Measured frequency responses of coupled resonators for different receiver angles. Lower-pitch resonator active, other resonator a semi-tone above. Separation 4 cm.

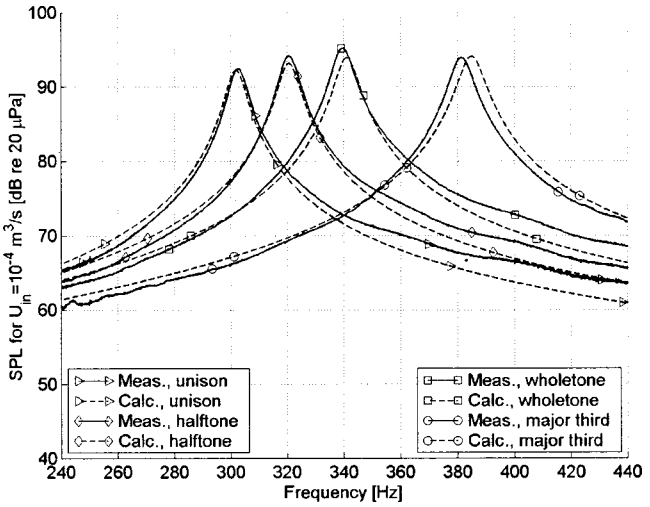


FIG. 10. Measured (full line) and calculated (dashed line) frequency responses of all four resonators in free-field conditions.

$S_l^m(x)$ is given by Eq. (19)

$$S_l^m(x)=(-1)^m\sqrt{\frac{2l+1}{2}\frac{(l-m)!}{(l+m)!}}P_l^m(x), \tag{19}$$

where $P_l^m(x)$ is an associated Legendre function.

The measurement point of interest is the same as the source point, $\mathbf{r}=\mathbf{r}_0$. The source is (arbitrarily) chosen to be at $\alpha=0$, as shown in Fig. 8. Also using

$$P_l^m(1)=\delta(m), \tag{20}$$

Eq. (17) can be greatly simplified:

$$p^{sc}(\mathbf{r}=\mathbf{r}_0)=\frac{\rho_0\omega^2}{\pi c}\sum_{l=0}^{\infty}\frac{j_l'(ka)}{h_l'(ka)}h_l(kr_0)^2\frac{2l+1}{2}. \tag{21}$$

The equivalent reflection impedance is obtained as

$$Z_{\text{refl}}=\frac{p^{sc}(\mathbf{r}=\mathbf{r}_0)}{U_0}=\frac{\rho_0\omega^2}{\pi c}\sum_{l=0}^{\infty}\frac{j_l'(ka)}{h_l'(ka)}h_l(kr_0)^2\frac{2l+1}{2}, \tag{22}$$

where, because the magnitude of the summation terms decrease rapidly with l at the low frequencies of interest, the summation is truncated at $l=10$.

Because the resonator separation is much smaller than the wavelength (below 500 Hz), the sound pressure of the reflected wave is approximately in phase with the transmitted wave at a neck opening. The phase difference increases with frequency, resulting first in a more inefficient radiation the higher the frequency. Therefore, in the present system the influence of reflections should move the resonances toward lower frequencies. This effect is observed in the predicted receiver sound pressure by doing the calculations both with and without the reflection impedance.

Using Fig. 3 and again assuming monopole sources at the neck openings, the receiver sound pressure can be deduced. This sound pressure should also include the scattering off the resonators. The first-order scattering can be evaluated using Eq. (17) with α as the actual receiver angle and all other parameter values the same as in the reflection impedance analysis. The scattering contribution evaluated in this way was found to be completely negligible for the presently analyzed system, and consequently is not included in the receiver sound-pressure equations.

The resulting receiver sound-pressure equation is

$$p(D,\theta)=\frac{j\omega\rho_0}{4\pi}\left(U_{n,1}\frac{e^{-jkD_1}}{D_1}+U_{n,2}\frac{e^{-jkD_2}}{D_2}\right). \tag{23}$$

The ratio of the receiver distance D and the resonator separation d is $(d/D)\leqslant(0.16/1.12)=0.14$, which justifies the use of the standard dipole approximation

$$D_1\approx D+\Delta D,\ \ D_2\approx D-\Delta D, \tag{24}$$

where

$$\Delta D=\frac{d}{2}\sin\theta, \tag{25}$$

and D_1 and D_2 are defined in Fig. 3. This gives a maximum distance error of 0.3%. Since $\Delta D/D\leqslant 0.16/2.24=0.071\ll 1$, it is possible to simplify Eq. (23) further, using

$$\frac{1}{D\pm\Delta D}\approx\frac{1}{D}\left(1\mp\frac{\Delta D}{D}\right). \tag{26}$$

The relative error in the approximation of Eq. (26) is less than 0.6%. Equation (23) can now be rewritten

$$p(D,\theta)=\frac{j\omega\rho_0}{4\pi}\frac{e^{-jkD}}{D}\left[(U_{n,1}e^{jk\Delta D}+U_{n,2}e^{-jk\Delta D})+\frac{\Delta D}{D}(U_{n,1}e^{jk\Delta D}-U_{n,2}e^{-jk\Delta D})\right]. \tag{27}$$

The first and second terms in the above expression correspond to the in-phase and the out-of-phase radiation component, respectively.

TABLE II. Measured and calculated resonance frequencies for resonators in free field.

Frequency difference	Measured f_0 [Hz]	Calculated f_0 [Hz]	Relative discrepancy in f_0 [%]
Unison	302.64	301.85	−0.26
Semi-tone	320.59	320.62	0.01
Whole tone	339.60	341.09	0.44
Major third	381.38	384.95	0.94

TABLE IV. Measured and calculated resonance amplitudes for resonators in free field.

Frequency difference	Measured A [dB]	Calculated A [dB]	Discrepancy in A [dB]
Unison	92.45	92.23	−0.23
Semi-tone	94.16	93.16	−1.00
Whole tone	95.23	93.90	−1.33
Major third	96.29	94.10	−2.18

VI. RESULTS

A. Frequency response

1. Introduction

In this section, the measured frequency responses for free-field and coupled resonators are compared to theoretical predictions from the equations presented here, especially Eq. (27), for predicting the receiver sound pressure. In all subsequent data plots, the frequency range of 240 to 440 Hz is used. This is where all resonances and all relevant coupling effects occur.

As stated, reciprocity implies that in the present system the same results are measured if the source and receiver positions are swapped. In order not to confuse the reader, the source-on-the-inside approach, which was used to develop the theory, is used in what follows. Therefore, the problem can be thought of as sound radiation from coupled resonators, only one of which is speaking.

The variation of the frequency responses with receiver angle is consistently very small, and no ''new'' features are found in the responses for angles other than 0°. Figure 9, showing the measured responses for coupled resonators a semi-tone apart and separated by 4 cm at receiver angles of −90 to 90°, in steps of 45°, illustrates this. To reduce the amount of data to present, only the responses for the 0° angle are presented throughout this section. This is no major limitation; it merely implies that the radiation model of Eq. (23), which is almost trivial, cannot be validated. Table I lists the relevant physical properties of air at 20 °C and 760-mm Hg.

2. Single resonators

A comparison between the measured and the calculated frequency spectra of all four resonators in the free field is shown in Fig. 10. The four resonators are the reference one, and three of higher pitch (a semi-tone above, a whole tone above, and a major third above the reference pitch). The agreement is good, which shows that our model of the resonator is accurate. Possible causes of the discrepancies between predictions and measurements include the use of lumped circuit elements and the approximate expressions for the losses [Eq. (5)].

TABLE III. Measured and calculated resonance Q-values for resonators in free field.

Frequency difference	Measured Q	Calculated Q	Relative discrepancy in Q [%]
Unison	46.13	44.16	−4.27
Semi-tone	46.69	44.52	−4.65
Whole tone	46.56	44.89	−3.59
Major third	47.52	45.21	−4.85

In Tables II, III, and IV, the resonant frequencies, resonance amplitudes, and Q values for all four resonators in the free field are displayed. The tables also show the relative discrepancies between measured and predicted values. The relative discrepancies in resonant frequency and Q value are less than 0.94% and 4.85%, respectively. The Q value is underestimated, i.e., the losses are overestimated, which again shows that better expressions for the losses are needed. Finally, the resonance peak amplitudes are slightly underestimated, the absolute discrepancy staying below −2.2 dB.

3. Coupled resonators

Measurement results and calculated predictions of frequency responses for coupled resonators are shown in Figs. 11, 12, and 13. Resonator separations of 1, 4, and 16 cm [subfigures (a), (b), and (c), respectively], and pitch differences of zero and a whole tone, corresponding to silent resonator lengths of 80.0 and 63.5 mm, respectively, give a representative view of the performance of the theory, and are therefore used to validate it. Both the magnitudes and the phases of the frequency responses are shown for a receiver angle θ of 0° (cf. Fig. 3).

As the figures show, the predicted responses agree well with the measured ones, a substantial part of the discrepancies probably due to the inaccuracies in the resonator model. The differences between the measurements and the predictions are approximately the same regardless of the specific parameter configuration, which further strengthens the model.

Figure 11 shows the receiver sound-pressure frequency response for resonators of the same pitch. As both the measurements and the predictions show, only one resonance, at a frequency quite far below the uncoupled pitch (measured at 302.64 Hz, predicted at 301.85 Hz), is found in the response. This phenomenon is specific to the 0° receiver angle; two resonances are actually developed in the neck velocities, but one is canceled in the 0° direction, i.e., at the canceled resonance, the sound-pressure contributions from resonator 1 and 2 are completely out of phase at the receiver. This topic is further discussed later.

As expected, the errors are largest for the smallest resonator separation, where the monopole approximation is least valid. Also, the influence of small absolute errors in the resonator separation is larger at small separations. For d larger than 1 cm, the prediction stays within ±3 dB of the mea-

sured response magnitude, and the phases never differ by more than 0.2π.

A resonator pitch difference of a whole tone and excitation of the lower-pitch resonator yields the responses of Fig. 12. The deviations between the measurements and the predictions stay within the tolerances stated above in this case as well.

Figure 13 shows the corresponding responses when the higher-pitch resonator is speaking. The responses are quite different from when the lower-pitch resonator is speaking; a discussion of this will follow later. The theory still performs well, the errors fulfilling the criteria for the lower-pitch resonator.

Concluding, Figs. 11, 12, and 13 show that the coupling theory developed in this paper is able to accurately predict the frequency responses of coupled Helmholtz resonators for

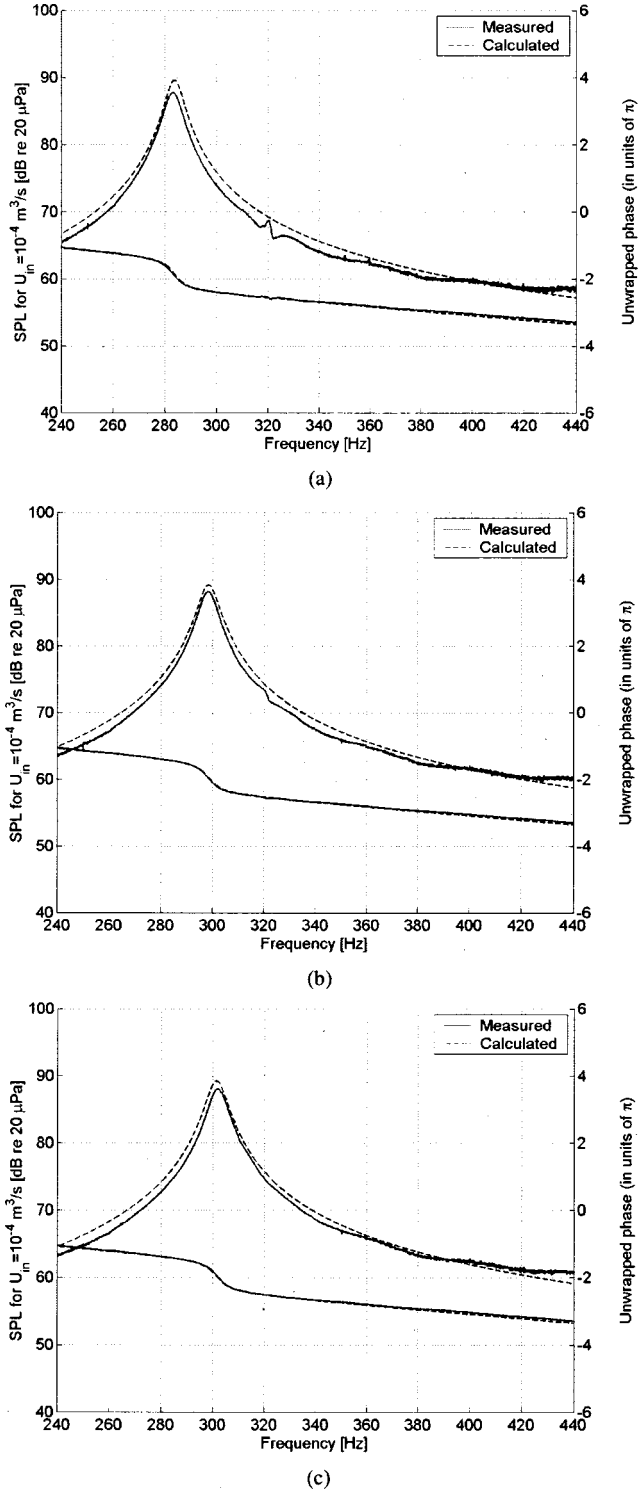


FIG. 11. Measured and calculated frequency responses (magnitude and phase) of coupled resonators. Resonators tuned to the same frequency. (a) Separation 1 cm. (b) Separation 4 cm. (c) Separation 16 cm.

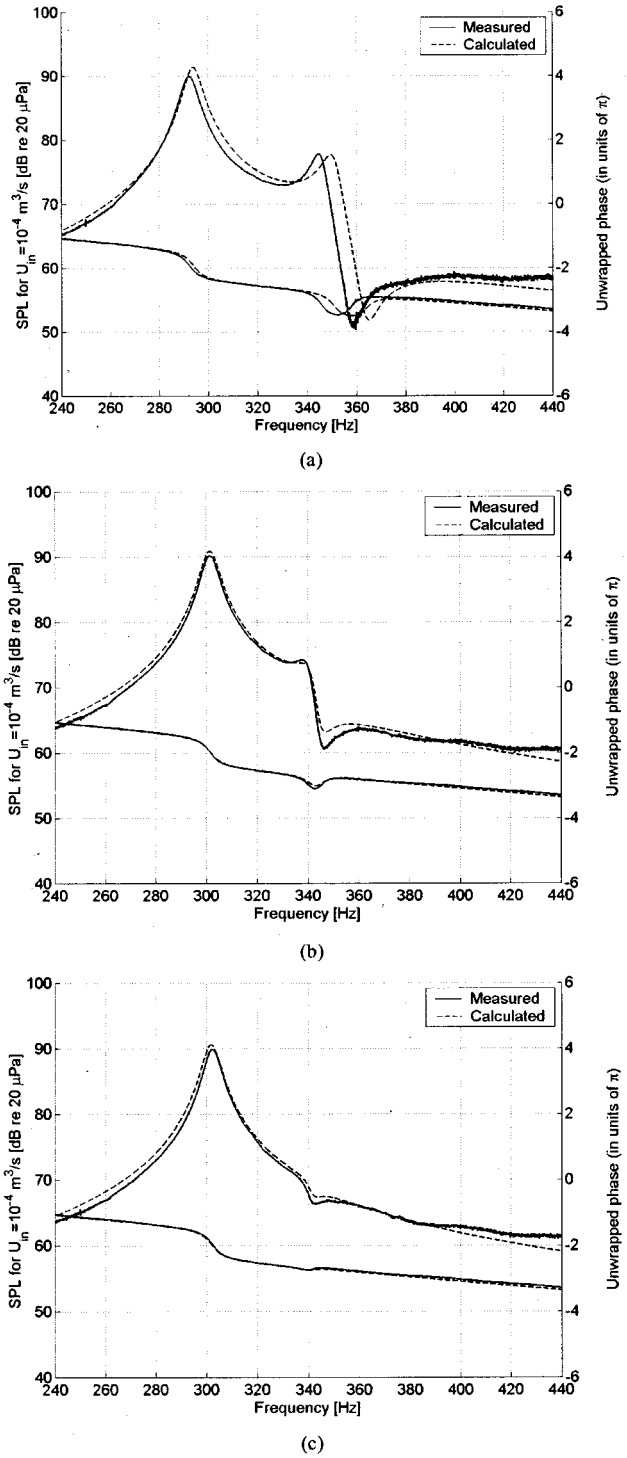


FIG. 12. Measured and calculated frequency responses (magnitude and phase) of coupled resonators. Lower-pitch resonator active, other resonator a whole tone above. (a) Separation 1 cm. (b) Separation 4 cm. (c) Separation 16 cm.

different resonator separations and pitch differences.

Since the effects of acoustic coupling between resonators have not been examined much in literature, attention is now turned to illustrating the various coupling effects that occur in the present system. For this purpose, and because the model has been successfully validated, Figs. 14, 15, 16, and 17 show the *predicted* receiver sound-pressure frequency response magnitudes for resonators of varying pitch differ-

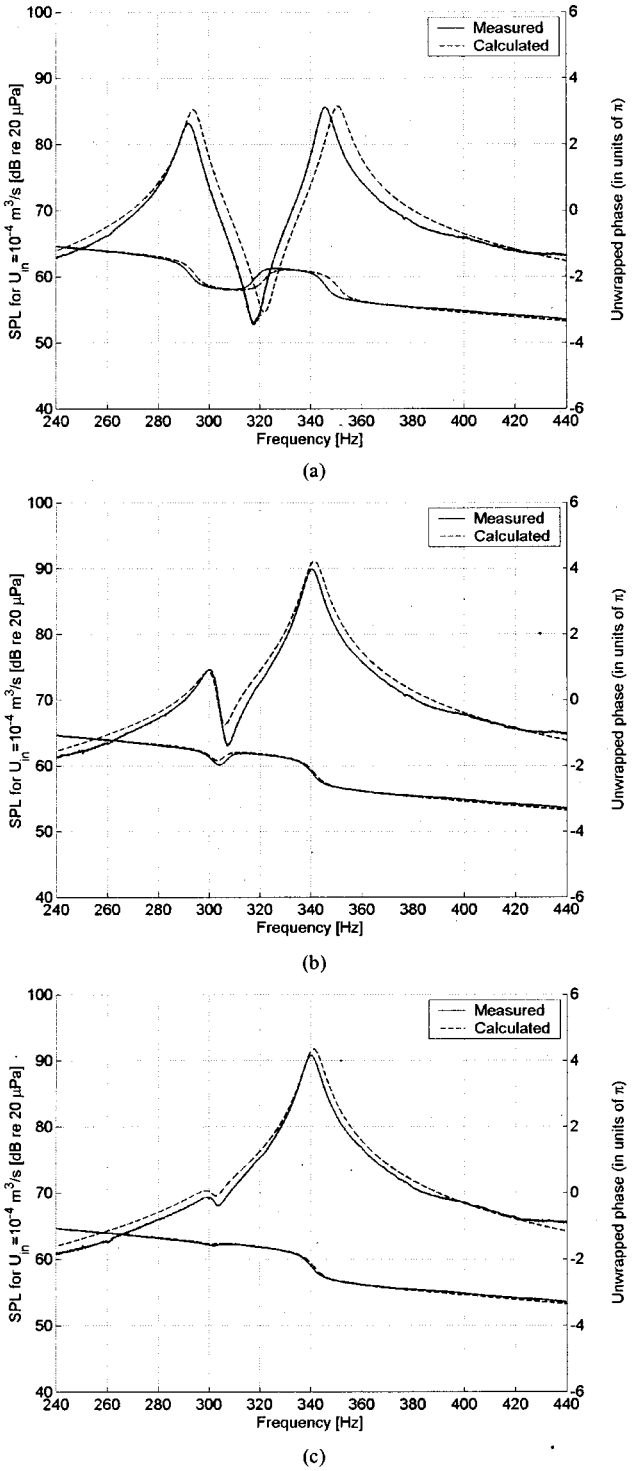


FIG. 13. Measured and calculated frequency responses (magnitude and phase) of coupled resonators. Higher-pitch resonator active, other resonator a wholetone below. (a) Separation 1 cm. (b) Separation 4 cm. (c) Separation 16 cm.

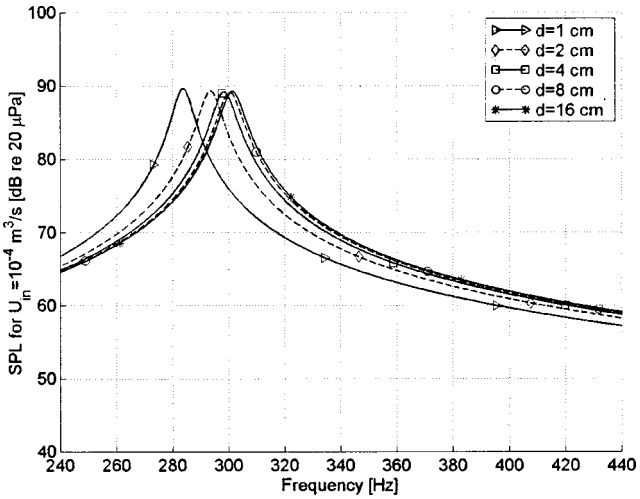


FIG. 14. Calculated predictions of frequency response magnitudes for coupled resonators. Resonators at unison resonant frequency, one of them active.

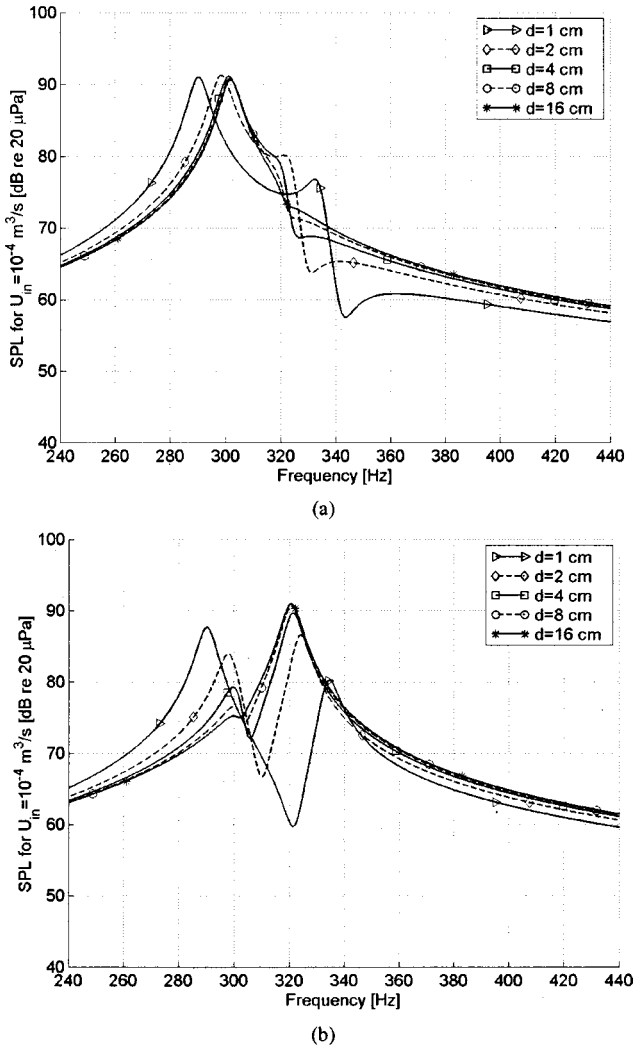


FIG. 15. Calculated predictions of frequency response magnitudes for coupled resonators. Resonator pitch difference of a semi-tone. (a) Lower-pitch resonator active. (b) Higher-pitch resonator active.

ence. The subfigures (a) and (b) show the responses for systems excited by the lower- and higher-pitch resonators, respectively. In each subfigure, the separation d is varied from 1 to 16 cm.

In the following, the resonances associated with the speaking resonator and the silent resonator will be called the primary and the secondary resonance, respectively. It is important to note that these primary and secondary resonances do not directly originate in any one resonator, but are a result of the interaction of sound waves originating in both resonators.

Throughout the measurements, no modifications were made to the lower-pitch resonator, but the body sections of the other resonator were changed to alter its pitch, as stated in Sec. II. This precludes drawing conclusions on properties of higher-pitch resonator-driven systems from the measurements. It is of course possible to make predictions from the model, but this is out of the scope of the present paper. For the small resonator separations used here (kd is below 1), the strength of the coupling increases when the resonators are moved closer together and when the uncoupled pitch differ-

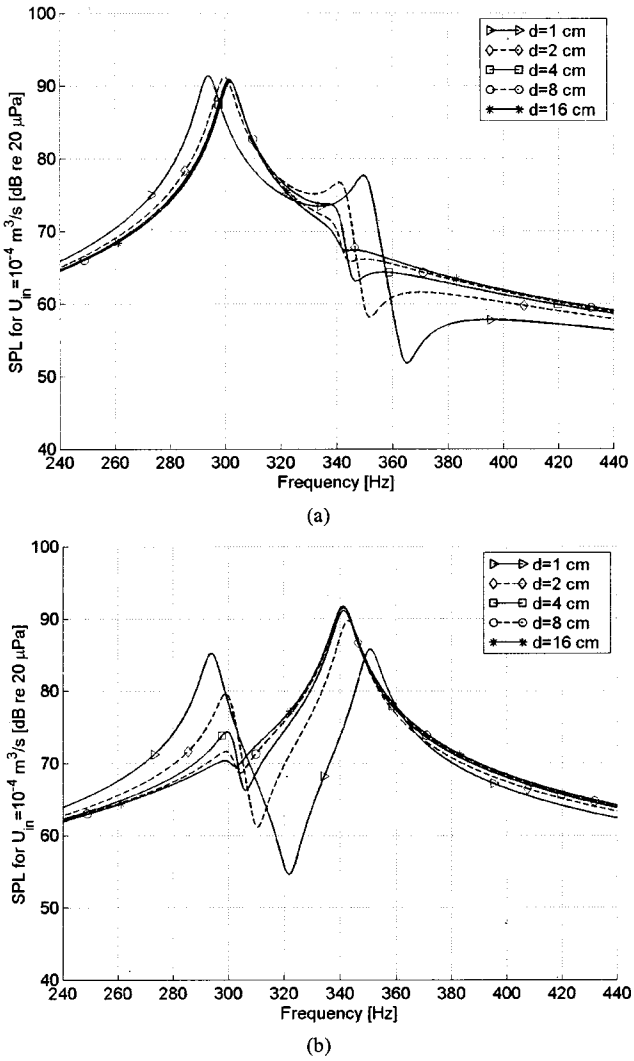


FIG. 16. Calculated predictions of frequency response magnitudes for coupled resonators. Resonator pitch difference of a wholetone. (a) Lower-pitch resonator active. (b) Higher-pitch resonator active.

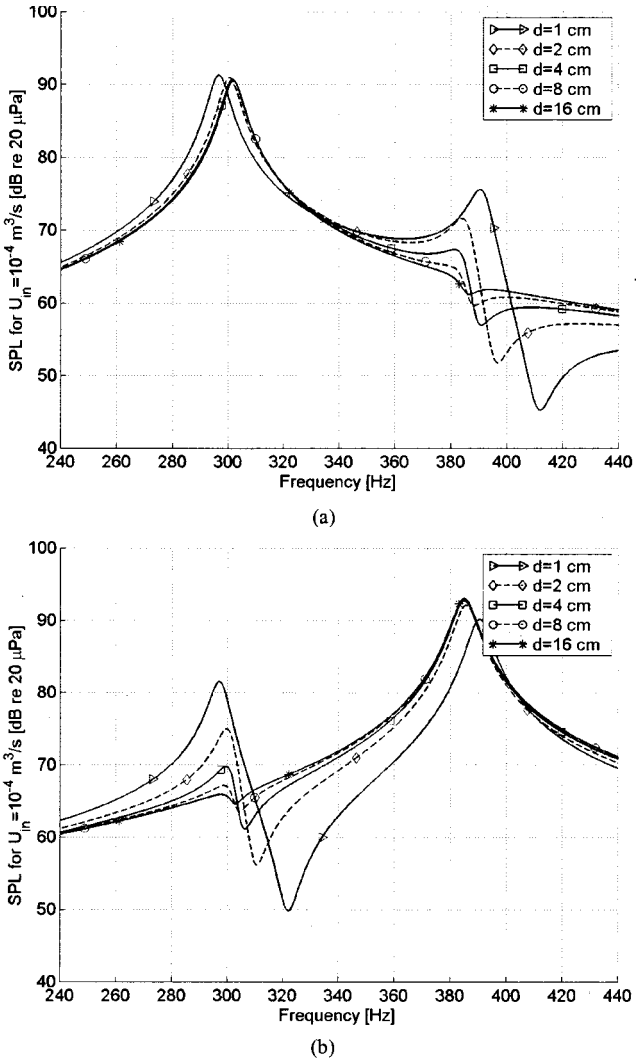


FIG. 17. Calculated predictions of frequency response magnitudes for coupled resonators. Resonator pitch difference of a major third. (a) Lower-pitch resonator active. (b) Higher-pitch resonator active.

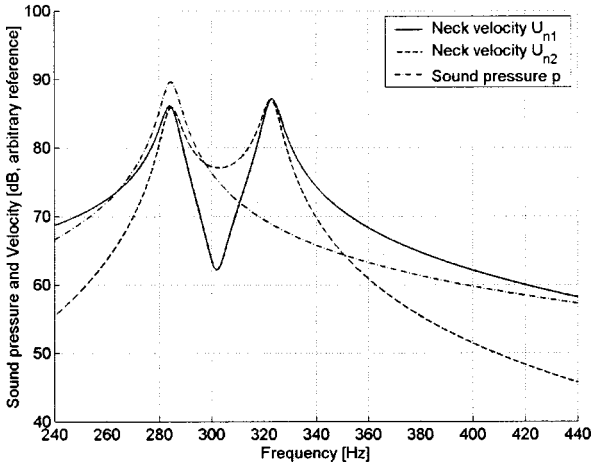


FIG. 18. Neck velocity frequency responses and corresponding receiver sound pressure in the 0° direction for resonators tuned to the same frequency and spaced by 1 cm.

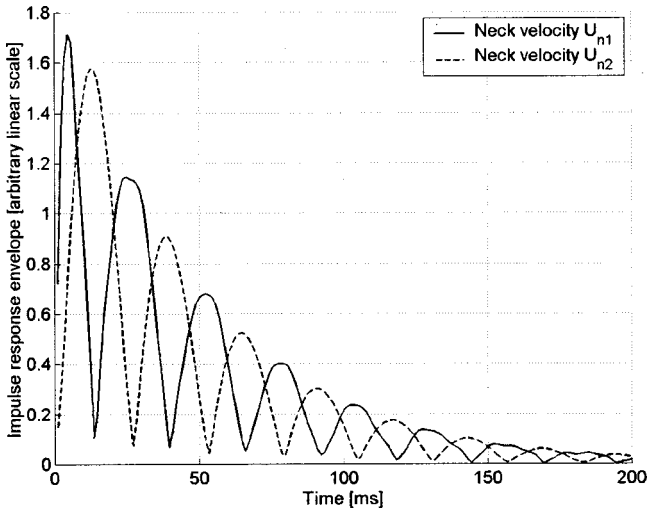


FIG. 19. Absolute values of neck velocity impulse response envelopes for coupled resonators tuned to the same frequency and separated by 1 cm.

ence is decreased. If the separation is on the order of the wavelength, resonant effects occur in the mutual impedance, causing coupling effects that are probably different from those reported here.

A stronger coupling implies a stronger excitation of the secondary resonance, and a decreased primary resonance amplitude. However, one cannot directly state that the energy lost in the primary resonance goes into the secondary resonance, because, due to the changed radiation conditions, the total system energy may change under the influence of coupling.

The coupled resonant frequencies (primary and secondary) are further apart than the uncoupled ones; the coupled higher pitch is higher and the lower one is lower than the uncoupled ones. The shift is larger the closer the two uncoupled pitches, i.e., the stronger the coupling. Figure 14 shows that if the uncoupled pitches are equal, only one resonance is present, at a pitch lower than any other pitch found here for the same resonator separation. Further investigation of the neck velocities for this special case reveal that two resonances *are* present in each neck velocity frequency response. The resonant frequencies for the smallest separation are around 285 and 323 Hz, respectively. Figure 18, a plot of the predicted neck velocity and receiver sound-pressure magnitudes, clearly shows the two resonances in the neck velocities, but only one resonance in the receiver sound pressure. The higher-pitch resonances in U_{n1} and U_{n2} are out of phase, and therefore cancel each other completely in the 0° direction. Because of the inefficiency of dipole radiation, the higher-pitch resonance results in a small frequency response peak at approximately 323 Hz only at large receiver angles.

The amplitude of the primary resonance decreases when coupling is introduced into the system (compared to the free-field results of Table III). However, the decrease is much smaller, and practically independent of resonator separation, when the lower-pitch resonator is speaking than when the higher-pitch resonator is speaking. In the former case, the amplitude shows only little variation with the uncoupled pitch difference, except when the uncoupled pitches are the same, in which case the amplitude of the common resonance is approximately 3 dB lower than the minimum primary resonance amplitude for different pitches. In summary, a system excited by the lower-pitch resonator is less influenced by coupling.

The Q value of the primary resonance exhibits very different behavior depending on which resonator is speaking. When the lower-pitch resonator is speaking, the primary resonance Q value decreases when coupling is introduced into the system. The decrease is larger the stronger the coupling, except for very small separations. However, when the higher-pitch resonator is speaking, the behavior is the opposite. The primary resonance Q value now increases when coupling is introduced, the more the stronger the coupling.

The Q values of the secondary resonances can only be evaluated for small separations, and they are not discussed here.

The losses of the system as a whole increase when the coupling grows stronger. Possible causes for this include the less efficient radiation resulting from the raised pressure at the neck openings, which in turn is due to the reflections from the other resonator.

It is important to realize that the specific coupling effects reported here might not be the same in another coupled sys-

TABLE V. Frequencies of energy oscillation between the resonators, compared to the difference between the primary resonant frequencies.

Resonant frequency difference	Resonator separation, d [cm]	Active resonator, lower pitch (L) or higher pitch (H)	Primary resonant frequency of lower-pitch resonator, f_L [Hz]	Primary resonant frequency of higher-pitch resonator, f_H [Hz]	$\frac{1}{2}(f_H - f_L)$, [Hz]	Energy oscillation frequency, f_E [Hz]
Unison	1		284.05	322.65	19.30	19.03
	4		298.05	304.15	3.05	⋯
	16		No oscillations can be detected			
Whole tone	1	L	294.80	351.80	28.50	28.70
	1	H	295.10	352.00	28.45	29.16
	4	L	301.55	341.15	19.80	19.89
	4	H	301.90	341.45	19.78	19.13
	16	L	301.80	340.75	19.48	19.34
	16	H	302.15	341.05	19.45	19.23

tem. For example, the phase of the mutual impedance is crucial to the coupling effects. In the studied system, the phase of the mutual impedance is approximately $\pi/2$ for all studied configurations and all frequencies of interest. This is equal to the phase of the mutual impedance between two point monopole sources spaced by a distance d if $d \ll \lambda$, which implies that the effect of the nonzero neck opening area is small.

For qualitative reasoning, one can thus approximate the mutual impedance at small separations by the mutual impedance between two point monopole sources. The phase of this mutual impedance varies as the phase of je^{-jkd}, which shows that when the separation is increased the phase of the mutual impedance decreases. In the present system, when the phase is zero, the pitch shift due to the mutual impedance is nearly zero. When the phase decreases just below zero, many of the coupling effects are the opposite of what is reported here. For example, in this case the coupling decreases the pitch differences, and the *higher*-pitch resonator-driven system becomes less influenced by coupling.

The conclusion of this reasoning is that the coupling effects presented here should be seen as the effects that occur in the present system when the resonator separation is much smaller than the wavelength. If the separation is comparable to the wavelength, the coupling effects will be very different. However, the theory developed here can handle this case as well.

In this paper, no equations for predicting the coupling effects have been derived. However, such equations can be derived using the coupling theory developed here.

B. Power flow

When the coupled system is excited, energy moves back and forth between the resonators. This is a fundamental property of coupled systems; analysis of the subject can be found in, e.g., Ref. 1, p. 67 for the case of a lossless system of two linear oscillators coupled by a linear restoring force. It is found that the energy surges back and forth at a frequency of

$$\omega_E = \tfrac{1}{2}|\omega_1 - \omega_2|, \tag{28}$$

where ω_1 and ω_2 are the coupled primary resonant frequencies of oscillator 1 and 2, respectively.

Considering the relative complexity of the studied system, as compared to the system studied in Ref. 1, it is of interest to find out whether the above equation still holds. Therefore, the predicted resonator neck velocities for a few system configurations are inverse-Fourier transformed to find the neck velocity impulse responses. The kinetic energy of the air in a resonator at any given time should be proportional to the square of the neck velocity, plus the small contribution from the much smaller velocity in the resonator body, which is neglected here.

The predicted neck velocity impulse response envelope magnitudes for resonators tuned to the same pitch and separated by 1 cm are shown in Fig. 19. The corresponding neck velocity frequency responses, shown in Fig. 18, have been previously discussed. The envelopes were obtained by Hilbert transformation. In Fig. 19, the energy oscillations back and forth between the resonators are evident. The frequencies of energy oscillation for several system configurations, including those in Figs. 18 and 19, are tabulated in Table V. The table also shows the primary resonant frequencies of the neck velocities and the differences between these, multiplied by 1/2, as suggested by Eq. (28).

As Table V shows, Eq. (28) still holds approximately true for the present system. Note, however, that Eq. (28) was derived in Ref. 1 assuming that $(f_H - f_L) \ll f_H$, which is not a valid assumption here.

VII. CONCLUSIONS

It has been shown to be possible, using analytical methods, to accurately model the sound radiation from a system of two coupled Helmholtz resonators, of which only one is speaking. For more complicated systems and setups, where one can no longer assume that all dimensions and relevant distances are much smaller than the wavelength, numerical methods are required to model the system.

The mutual impedance approach used in this work can of course also be used for more complicated geometries, where numerical methods must probably be used to deduce sound field parameters.

The ''reflection'' impedance approach taken here can also be used to model the influence of walls and other rigid objects in the vicinity of a resonator, as well as the influence of reflections from other resonators.

The theory developed here can easily be extended to handle a system with several silent resonators. Also, it would not be difficult to extend it to handling several speaking resonators; however, the reciprocity measurement approach then cannot be used.

ACKNOWLEDGMENTS

The work presented here was done within the framework of the research project ''Changing Processes in North European Organ Art 1600–1970,'' conducted at the Göteborg Organ Art Center, Göteborg University, and at Chalmers University of Technology, Göteborg, with financial support from the Bank of Sweden Tercentenary Foundation and Chalmers University of Technology. The authors also wish to extend their thanks to Munetaka Yokota both for stating the problem and for discussions and demonstrations of various coupling phenomena between organ pipes, as well as to Vincent Rioux for help in doing measurements and for valuable discussions.

[1] P. Morse and U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), Chap. 3.
[2] D. E. Newland, ''Calculation of power flow between coupled oscillators,'' J. Sound Vib. **3**, 262–276 (1966).
[3] R. H. Lyon and G. Maidanik, ''Power flow between linearly coupled oscillators,'' J. Acoust. Soc. Am. **34**, 623–639 (1962).
[4] R. L. Pritchard, ''Mutual acoustic impedance between radiators in an infinite plane,'' J. Acoust. Soc. Am. **32**, 730–737 (1966).
[5] H. P. Neff, Jr., ''Mutual Impedance of Circular Pistons,'' J. Audio Eng. Soc. **43**, 695–699 (1995).
[6] J. D. Kraus, *Antennas* (McGraw-Hill, New York, 1950), Chap. 10.
[7] U. Ingard, ''On the theory and design of acoustic resonators,'' J. Acoust. Soc. Am. **25**, 1037–1061 (1953).
[8] F. P. Mechel, ''Notes on the radiation impedance, especially of piston-like radiators,'' J. Sound Vib. **123**, 537–572 (1988).

[9] T. L. Rhyne, ''Radiation coupling of a disk to a plane and back or a disk to a disk: An exact solution,'' J. Acoust. Soc. Am. **61**, 318–324 (1977).
[10] R. L. Panton and J. M. Miller, ''Resonant frequencies of cylindrical Helmholtz resonators,'' J. Acoust. Soc. Am. **57**, 1533–1535 (1975).
[11] A. Cummings, ''Acoustics of a cider bottle,'' Appl. Acoust. **5**, 161–170 (1972).
[12] J. Kergomard and A. Garcia, ''Simple discontinuities in acoustic waveguides at low frequencies: Critical analysis and formulae,'' J. Sound Vib. **114**, 465–479 (1987), Table 2A, Eq. (1).
[13] M. C. A. M. Peters, A. Hirschberg, A. J. Reijnen, and A. P. J. Wijnands, ''Damping and reflection coefficient measurements for an open pipe at low Mach and low Helmholtz numbers,'' J. Fluid Mech. **256**, 499–534 (1993).
[14] S. N. Rschevkin, *The Theory of Sound* (Pergamon, London, 1963), p. 101.
[15] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed. (Pergamon, Oxford, 1987), p. 301.
[16] L. L. Beranek, *Acoustics* (American Institute of Physics, Woodbury, NY, 1993).
[17] Y. Ando, ''On the sound radiation from semi-infinite circular pipe of certain wall thickness,'' Acustica **22**, 219–225 (1969/70).
[18] G. Natsiopoulos, *Scattering—An Analytical Study* (Report E00-05, Chalmers University of Technology, Gothenburg, Sweden, 2000), Chap. 3.2.3.

Sound propagation in rigid bends: A multimodal approach

S. Félix[a] and V. Pagneux
Laboratoire d'Acoustique de l'Université du Maine, UMR-CNRS 6613, Université du Maine, Avenue Olivier Messiaen, 72085 Le Mans Cedex 09, France

(Received 1 February 2001; revised 11 June 2001; accepted 13 June 2001)

The sound propagation in a waveguide with bend of finite constant curvature is analyzed using multimodal decomposition. Two infinite first-order differential equations are constructed for the pressure and velocity in the bend, projected on the local transverse modes. A Riccati equation for the impedance matrix is then derived, which can be numerically integrated after truncation at a sufficient number of modes. An example of validation is considered and results show the accuracy of the method and its suitability for the formulation of radiation conditions. Reflection and transmission coefficients are also computed, showing the importance of higher order mode generation at the junction between the bend and the straight ducts. The case of varying cross-section curved ducts is also considered using multimodal decomposition. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1391249]

PACS numbers: 43.20.Mv [ANN]

I. INTRODUCTION

Because the study of the propagation in curved ducts appears as a logical extension to the theory of the propagation of acoustics and electromagnetic fields in straight waveguides, it has aroused a number of works and papers (for a review in acoustics, see Rostafinski[1]).

The most classical way of solving the problem is the method of separation of variables,[2–6] leading to the calculation of modes in the bend. In the sequel we will call this approach the ''bend mode method'' (BMM). Indeed, for a circular bend of rectangular cross section, and in particular for a two-dimensional circular bend, the equations of motion are known to be separable. However, for a bend of finite length this solution must be joined to the wave solutions of the adjoining straight ducts. To avoid the problem, Krasnushkin,[2] Grigor'yan,[3] and others restricted themselves to consideration of infinitely long bends. The case of a bend joining two straight ducts was treated by Rostafinski,[4] Cummings,[5] and Osborne,[6] by using the BMM.

The dispersion relation for bend modes involves Bessel functions of noninteger order. In view of these difficulties due to the implicit dispersion relation, approximations were proposed. Krasnushkin[2] adopted a perturbation method and treated only the case of slightly bent tubes. Grigor'yan[3] developed a method of power-series expansion of the expressions involving the Bessel functions. Rostafinski[4] restricted his investigation to very low frequencies, and also expanded the Bessel functions in terms of increasing powers of their argument. Cummings[5] considered only the propagation of the plane wave mode, and Osborne[6] used a simplified theory, notably in considering only the propagating modes.

Methods different from the BMM have been used, such as the Galerkin method (Tam[7]), finite differences method (Cabelli[8]) or more recently a method based on parabolic approximation (Dougherty[9]). Such approaches permit to investigate more general cases than those, restricted, presented previously, and are powerful to yield information. Nevertheless, generally, they do not allow a direct physical interpretation, and are not always convenient for formulating radiation conditions.

In this context we propose to formulate a multimodal method (MMM) to determine the wave propagation in two-dimensional rigid circular bends of any dimension and at any frequency. Such study for ducts of straight longitudinal axis was done by Pagneux *et al.*,[10,11] who proposed and used successfully a multimodal method for calculating the wave propagation in varying cross-section waveguides. Following works done by Stevenson,[12] Albertson,[13] Roure,[14] and Kergomard,[15] they constructed two infinite first-order differential equations for the components of the pressure and of the axial velocity projected onto the normal modes. Then they derived a new equation for the impedance matrix which is numerically workable. Tam[7] adopted a quite similar approach, using the local transverse modes to construct a set of basis functions for calculating the solution of the Euler equations in a circular bend of constant cross section by the Galerkin method. He gave a very interesting algebraic calculation of the reflection and transmission matrices. Results were, on the other hand, limited to tables of values, without particular applications. A similar algebraic approach is presented in the appendix.

In Sec. II of this paper, the modal method described previously is formulated for circular bends of constant cross section, pointing out the suitability of this approach for the formulation of radiation conditions, and the possibility to access rapidly to the impedance, and consequently to the input impedance and reflection coefficient. A particular example is treated in Sec. III, for which an analytical solution can be found, allowing the proposed method to be validated. Section IV presents several applications, showing the accuracy and advantages of the present method. Finally, in Sec. V, the multimodal approach is extended to varying cross-section curved ducts, and results are presented.

II. EQUATIONS AND BOUNDARY CONDITIONS

The general duct system to be analyzed here is shown in Fig. 1; it consists of a two-dimensional circular bend which

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

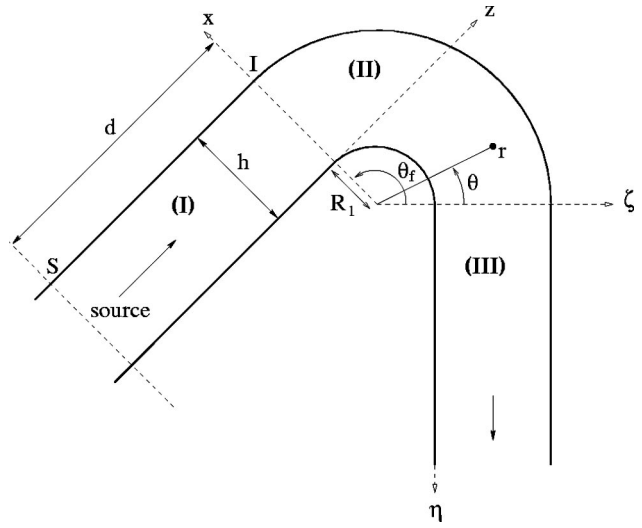


FIG. 1. Geometry of the duct system with bend and systems of coordinates.

joins rigid straight ducts. The parameters of the bend are its overall angle θ_f, inner radius R_1 and width h. A source generates sound upstream from the bend, and a radiation condition is imposed at the outlet.

A. Formulation

Throughout this paper adiabatic lossless linear media will be assumed. The acoustic velocity $\hat{\mathbf{v}}$ and the acoustic pressure $\hat{p}$ satisfy the equation of mass conservation

$$\boldsymbol{\nabla}\cdot\hat{\mathbf{v}} = -\frac{j\omega}{\rho_0 c_0^2}\hat{p}, \tag{1}$$

and the equation of momentum conservation

$$j\omega\hat{\mathbf{v}} = -\frac{1}{\rho_0}\boldsymbol{\nabla}\hat{p}, \tag{2}$$

the caret " ˆ " denoting dimensional quantities and time dependence $\exp(j\omega t)$ being omitted. $\hat{\mathbf{v}}$ has components $(\hat{v}_r, \hat{v}_\theta)$, ρ_0 is the density of air, and c_0 is the speed of sound.

The reference pressure $\rho_0 c_0^2$ and velocity c_0 are used to reduce Eqs. (1) and (2) to dimensionless terms, and we introduce the frequency $k=\omega/c_0$, so that these equations become

$$\boldsymbol{\nabla}\cdot\mathbf{v} = -jkp, \tag{3}$$

$$-jk\mathbf{v} = \boldsymbol{\nabla}p. \tag{4}$$

By eliminating v_r these equations give in the bend

$$-jkrp = \frac{\partial v_\theta}{\partial\theta} - \frac{1}{jk}\frac{\partial}{\partial r}\left(r\frac{\partial p}{\partial r}\right), \tag{5}$$

$$-jkrv_\theta = \frac{\partial p}{\partial\theta}. \tag{6}$$

Pressure p and axial velocity v_θ are now expressed using infinite series

$$p(r,\theta) = \sum_n \psi_n(r)P_n(\theta), \tag{7}$$

$$v_\theta(r,\theta) = \sum_n \psi_n(r)U_n(\theta), \tag{8}$$

where P_n and U_n are scalar coefficients, and ψ_n are the eigenfunctions obeying the transverse eigenproblem $\partial_{r^2}^2\psi_n = -\alpha_n^2\psi_n$, with the homogeneous Neumann boundary condition on the walls:

$$\left(\frac{\partial\psi_n}{\partial r}\right)_{r=R_1,R_1+h} = 0, \tag{9}$$

and the orthogonality relation

$$\int_{R_1}^{R_1+h}\psi_m(r)\psi_n(r)\,dr = \delta_{mn}. \tag{10}$$

These functions ψ_n are

$$\psi_n(r) = A_n\cos\left(\frac{n\pi}{h}(r-R_1)\right), \tag{11}$$

where

$$A_n = \sqrt{\frac{2-\delta_{n0}}{h}},$$

with the associated eigenvalues $\alpha_n = n\pi/h$. They are the classical transverse modes in a straight duct.

Following the matricial terminology, modal decompositions (7) and (8) will be written as

$$p = {}^t\boldsymbol{\psi}\mathbf{P}, \tag{12}$$

$$v_\theta = {}^t\boldsymbol{\psi}\mathbf{U}, \tag{13}$$

where $\mathbf{P}=(P_n)_{n\geqslant 0}$, $\mathbf{U}=(U_n)_{n\geqslant 0}$, and $\boldsymbol{\psi}=(\psi_n)_{n\geqslant 0}$ are columns vectors.

Then, by projecting Eqs. (5) and (6) on the functions ψ_n (see Appendix A), the following equations are obtained:

$$\mathbf{U}' = \frac{1}{jk}(C+KB)\mathbf{P}, \tag{14}$$

$$\mathbf{P}' = -jkB\mathbf{U}, \tag{15}$$

with matrices K, B, and C given by (see Appendix B)

$$K_{mn} = (k^2-\alpha_m^2)\,\delta_{mn}, \tag{16}$$

$$B_{mn} = \int_{R_1}^{R_1+h} r\psi_n(r)\psi_m(r)\,dr, \tag{17}$$

$$C_{mn} = \int_{R_1}^{R_1+h}\psi_n(r)\psi_m'(r)\,dr. \tag{18}$$

Because of the presence of evanescent modes (corresponding to $k^2-\alpha_m^2<0$), Eqs. (14) and (15) are numerically unstable,[10,11] and cannot be integrated directly. Therefore, following Ref. 10, an impedance matrix is defined, fulfilling

$$\mathbf{P} = Z\mathbf{U}. \tag{19}$$

Thus, by substituting $\mathbf{P}'$ and $\mathbf{U}'$ from Eqs. (14) and (15) in the derivation of Eq. (19), $\mathbf{P}'=Z'\mathbf{U}+Z\mathbf{U}'$, a Riccati equation for Z can be obtained:

$$Z'=-jkB-\frac{1}{jk}Z(C+KB)Z. \tag{20}$$

This first-order differential equation allows us to obtain the impedance matrix at the inlet, once the radiation condition (i.e., impedance matrix) is given at the outlet.

B. Boundary conditions

A radiation condition downstream from the bend is defined, which gives an initial condition for the Riccati equation (20). In the case of a bend terminated by a semi-infinite straight duct, as shown in Fig. 1, the radiation impedance is the characteristic impedance matrix, which is diagonal and given by

$$Z_{c_n}=-\frac{k}{k_n}, \tag{21}$$

where $k_n=\sqrt{k^2-\alpha_n^2}$ for propagative modes, $k_n=-j\sqrt{\alpha_n^2-k^2}$ for evanescent modes.

With this radiation condition Z can be calculated everywhere in the duct, since it is governed by a first-order differential equation [Eq. (20)]. To know the acoustic field, it is then possible to apply a boundary condition upstream, characterized as a known velocity or pressure, so once the impedance is found the pressure or velocity can be calculated as follows: we substitute Eq. (19) in Eq. (15) [respectively, (14)] to obtain

$$\mathbf{P}'=-jkBY\mathbf{P}, \tag{22}$$

respectively,

$$\mathbf{U}'=\frac{1}{jk}(C+KB)Z\mathbf{U}, \tag{23}$$

where Y is the admittance matrix ($Y=Z^{-1}$). Equation (22) [respectively, (23)] can be integrated from the inlet to the radiating end, giving $\mathbf{P}$ (respectively, $\mathbf{U}$). Afterwards the velocity (respectively, pressure) can be calculated, once $\mathbf{P}$ (respectively, $\mathbf{U}$) and Z are known. In contrast to direct numerical integration of Eq. (14) or Eq. (15), the integration of Eq. (22) or Eq. (23) is numerically stable because of the presence of the impedance matrix (see Ref. 10 for more details).

C. Numerical procedures

The numerical procedure adopted here is the same as used by Pagneux *et al.*[10] After truncation at a sufficient number of modes, the Riccati equation for Z is integrated down to the inlet of the bend using Runge–Kutta algorithms, with an adaptive step size. Z being stored on a grid of regularly spaced points along the axis, the integration of Eq. (23) in the opposite direction allows the calculation of the pressure and velocity fields.

Just as any radiation condition can be easily considered (except a tube closed by a rigid wall that necessitates putting attenuation in the model), the integration of Z, $\mathbf{U}$, and $\mathbf{P}$ can be extended to any type of waveguide upstream from the bend (see Appendix C). In contrast, when using the BMM, the formulation and computation of the continuity equations at the discontinuity between the bend and a joined duct may lead to mathematical and numerical difficulties (notably in term of convergence, see Ref. 16). More generally, joining solutions developed on different basis of functions may be a source of numerical difficulties. The iterative calculation of the impedance matrix along a system of joined ducts avoids such difficulties.

III. VALIDATION

To validate our method, we consider in this section a particular case of propagation in a bend for which an analytical solution can be found: the propagation of the first bend mode. The analytical and the multimodal approaches are compared and discussed.

As mentioned above, the Helmholtz equation is separable in the cylindrical coordinate system, and consequently there exists solutions of the type $p(r,\theta)=R(r)\Theta(\theta)$. These modes are characterized in the bend by a dimensionless wave number called (following Krasnushkin[2]) ''angular wave number.'' For a hard-wall bend these angular wave numbers may be real or pure imaginary numbers, corresponding, respectively, to propagative, and evanescent modes. We can thus sort them in decreasing order of their square values and call the first ν_0. In this paragraph, we are interested in calculating the first bend mode, characterized by ν_0, in order to validate our multimodal method with this solution.

A. Analytical solution

For any geometry of infinite bend, whatever its curvature, the Helmholtz equation for the acoustic pressure can be solved by separation of variables, to obtain the general solution

$$p(r,\theta)=\sum_{n\geqslant 0}(\alpha_n e^{j\nu_n\theta}+\beta_n e^{-j\nu_n\theta})[Y'_{\nu_n}(kR_1)J_{\nu_n}(kr) - J'_{\nu_n}(kR_1)Y_{\nu_n}(kr)], \tag{24}$$

obeying the homogeneous Neumann conditions ($\partial_r p=0$) on the walls ($r=R_1$ and R_1+h). Angular wave numbers ν_n are the solutions of the dispersion relation (see Krasnushkin[2])

$$J'_\nu(kR_1)Y'_\nu(k(R_1+h))-J'_\nu(k(R_1+h))Y'_\nu(kR_1)=0 \tag{25}$$

for a given value of $k=\omega/c_0$, and where J_ν and Y_ν are the Bessel functions of first and second kinds.

For a finite bend, expression (24) is also an exact analytical solution, but with boundary conditions at the inlet and the outlet that have to be determined. This point is studied in the following paragraph. In the sequel we will be interested in the first bend mode

$$p(r,\theta)=\alpha_0 e^{j\nu_0\theta}[Y'_{\nu_0}(kR_1)J_{\nu_0}(kr) - J'_{\nu_0}(kR_1)Y_{\nu_0}(kr)], \tag{26}$$

and we want to validate the multimodal approach with this known exact solution.

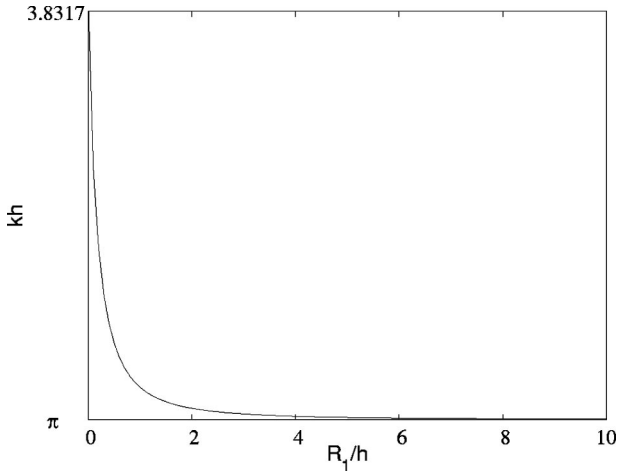


FIG. 2. First cutoff frequency in an infinite bend, as a function of the ratio R_1/h.

B. Multimodal approach

As mentioned previously (Sec. II B), we need an initial condition for the Riccati equation (20). We project the relation obtained from the exact solution (26)

$$\frac{\partial p}{\partial \theta} = j\nu_0 p \tag{27}$$

over the modes ψ_n to calculate the equivalent boundary condition for the impedance at the outlet ($\theta = 0$):

$$Z = -\frac{k}{\nu_0} B. \tag{28}$$

At the inlet of the bend, $\mathbf{P}(\theta_f)$ is given by projection of Eq. (26) over the ψ_n. This yields:

$$P_n(\theta_f) = A_n \alpha_0 e^{j\nu_0\theta_f} \int_{R_1}^{R_1+h} [Y'_{\nu_0}(kR_1) J_{\nu_0}(kr) - J'_{\nu_0}(kR_1) Y_{\nu_0}(kr)] \cos\left[\frac{n\pi}{h}(r-R_1)\right] dr. \tag{29}$$

C. Results

The source frequency is chosen under the first cutoff frequency in the bend; this cutoff frequency k is the first nonzero solution of

$$J'_0(kR_1) Y'_0(k(R_1+h)) - J'_0(k(R_1+h)) Y'_0(kR_1) = 0 \tag{30}$$

(see Osborne[6]). Figure 2 shows this cutoff frequency adimensioned with the duct width h, as a function of R_1/h. It is always larger than π—the first cutoff in a straight duct—and tends toward this value for extremely large ratio R_1/h, as expected. In our case, the geometry of the bend will be given by $R_1/h = 0.5$. This gives for the solution kh of Eq. (30) a value of 3.27. We then take an overall angle $\theta_f = \pi/2$ and a frequency such that $kh = \pi/4$, for which the angular wave number ν_0 equals 0.76.

Figure 3 shows the isopressure contours calculated using respectively the exact analytical expression (26) and the MMM, the pressure being in this case computed as described

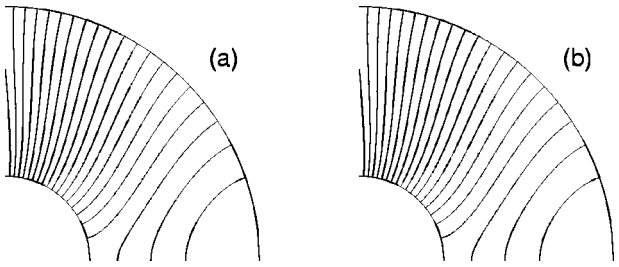


FIG. 3. Contour of the real part of pressure in a bend for the first angular mode, calculated (a) analytically, (b) using the MMM ($\theta_f = \pi/2$, $h/R_1 = 2$, $kh = \pi/4$).

in Sec. II C, with 17 modes. The results obtained by the modal decomposition agree perfectly with those calculated analytically [Eq. (26)]. As an example of check, the multimodal approach gives a pressure field with modulus independent of the axial coordinate θ and with a phase independent of r, as expected with expression (26). The maximum range compared to the mean value in both cases is less than 0.001%.

In order to measure the rate of convergence of the multimodal method, the analytical result (26) is chosen as the reference, and we define the error as follows:

$$\epsilon = \left(\frac{\int_0^{\theta_f}\int_{R_1}^{R_1+h} \|p - p_{\text{ref}}\|^2 r\, dr\, d\theta}{\int_0^{\theta_f}\int_{R_1}^{R_1+h} \|p_{\text{ref}}\|^2 r\, dr\, d\theta} \right)^{1/2}. \tag{31}$$

The error ϵ is shown in Fig. 4: it decreases monotically as more modes are introduced in the calculation. The convergence follows a $1/N^{3.23\pm0.1}$ law, N being the number of modes taken into account in the calculation. Thus the solution of the modal approach rapidly converges on the exact solution.

D. Conclusion of the validation

The multimodal method is validated, giving results in almost prefect agreement with the analytical solution in the example developed previously. Because this calculation is not limited to particular values of the geometrical and fre-

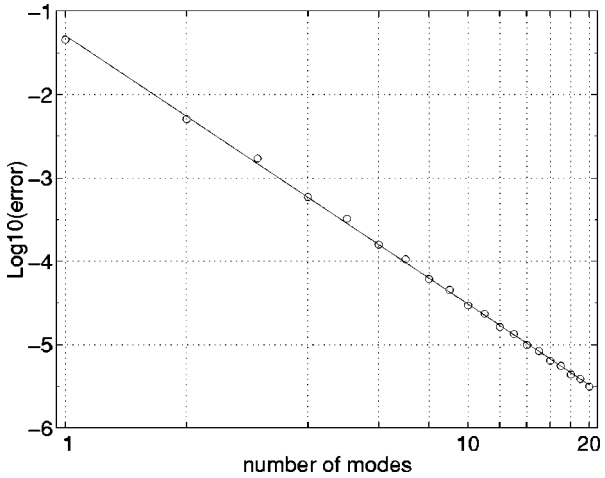


FIG. 4. Convergence of the MMM. The logarithm of the error, $\log(\epsilon)$, is plotted as a function of the number of modes for the same case as Fig. 3.

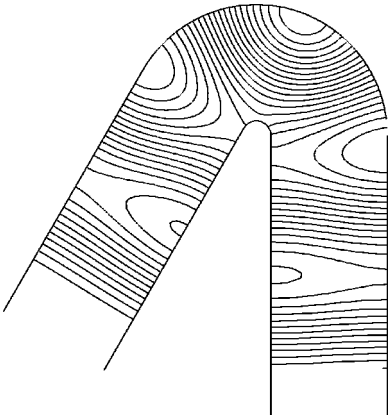


FIG. 5. Contour of the real part of pressure in a duct with $\theta_f = 2.62$ and $h/R_1 = 8$. A plane piston source of frequency such that $kh = 3$ is placed a distance $2h$ upstream from the bend.

quency parameters, it constitutes an interesting validation of the multimodal method for bends of constant cross section. Moreover the desired solution is a nontrivial solution considering our method, and has been calculated with very good accuracy.

IV. APPLICATION

In this section we show an example of applications of the MMM for the characterization of a particular bend, which has been already characterized by an experiment and a finite differences method calculation in the literature.[8]

The geometry consists in a bend joining two semi-infinite straight ducts, as shown in Fig. 1, with $\theta_f = 2.62$ and $h/R_1 = 8$. The pressure field represented in Fig. 5 with contours of its real part is calculated with 15 modes, using the procedure described in Sec. II C. A condition $P_n = \delta_{n0}$ (piston source) is imposed at a distance $2h$ upstream from the bend, and the frequency of analysis is such that $kh = 3$. This frequency being under the first cutoff frequency, the plane wave is the only propagating mode in the straight sections of the duct system considered here. Thus we can see the distortion of the pressure distribution profiles, and the influence of the evanescent modes in the straight sections, even relatively far from the junction with the bend.

The amplitude reflection coefficient for the plane wave mode (see Pagneux *et al.*[10] for details) is calculated at the inlet of the bend in the duct system previously described and is given in Fig. 6 as a function of the frequency parameter kh, varying between 0 and π. The algebraic method presented in Appendix D gives similar results to our method, as expected. Experimental and numerical results of Cabelli[8] for the same geometry are also reported. The multimodal method gives results in good agreement with the experimental measurements. While the finite difference calculation of Cabelli seems to give satisfying results for low frequency, values near the cutoff are underestimated. As anechoic termination, Cabelli applied a boundary condition a distance $0.8h$ downstream from the bend, ensuring a lack of reflection of the propagative modes and the first nonpropagative mode only. Near the cutoff frequency, this condition may be too limited. Furthermore, the calculation of the standing wave

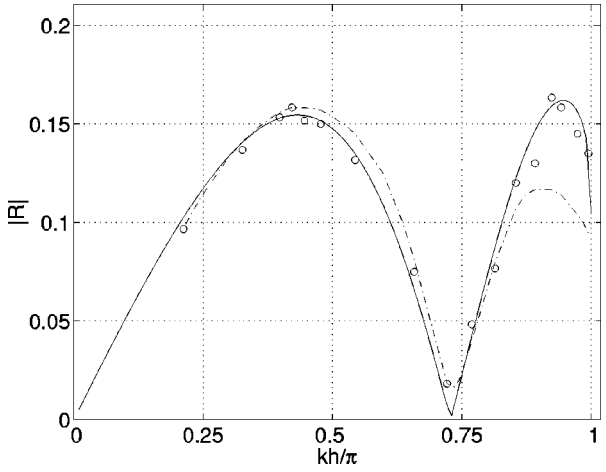


FIG. 6. Variations of the amplitude reflection coefficient with the frequency kh/π. MMM (plain line), finite differences (dashed dotted line, from Cabelli—Ref. 8), experimental results (circles, from Cabelli—Ref. 8) ($\theta_f = 2.62$, $h/R_1 = 8$).

ratio upstream from the bend implies that the wave is plane, without significant evanescent modes, and this hypothesis may cause considerable errors near the cutoff. Such errors are avoided with the multimodal method, since a sufficiently large number of evanescent modes can be taken into account, in order to get accurate results.

V. VARYING CROSS SECTION CURVED DUCTS

In this section, the study initiated by Pagneux *et al.*[10,11] for two-dimensional varying cross section waveguides of straight longitudinal axis is extended to curved ducts with varying cross section, as shown in Fig. 7, where notations are defined. A new Riccati equation is derived and we present results obtained with this generalized formulation.

A. Formulation

The basis functions ψ_n are now functions of both r and θ :

$$\psi_n(r,\theta) = \sqrt{\frac{2-\delta_{n0}}{R_2(\theta)-R_1(\theta)}}\cos\left(n\pi\frac{r-R_1(\theta)}{R_2(\theta)-R_1(\theta)}\right). \tag{32}$$

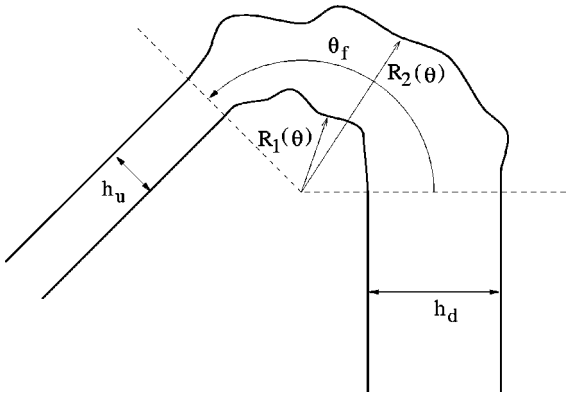


FIG. 7. Geometry of the duct system with varying cross section curved duct and notations.

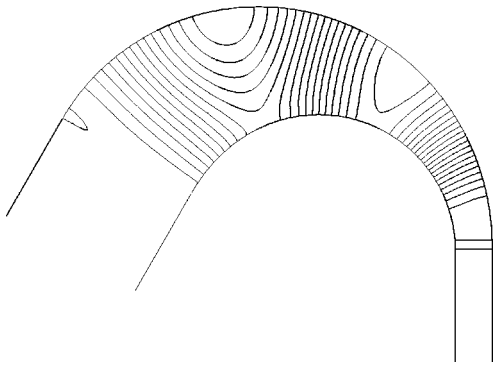


FIG. 8. Contour of the real part of pressure in a trunk with $\theta_f=2.62$, $h_u/h_d=4$ and $R_1(\theta)$ and $R_2(\theta)$ given, respectively, by Eqs. (38) and (39) (see Fig. 7). A plane piston source of frequency such that $kh_e=3$ is placed at the inlet of the curved part.

Furthermore, the boundary condition is now $\partial_n p=0$ on the hard walls. Paying attention to these remarks, we project Eqs. (5) and (6) on the set of functions ψ_n to find (see Appendix E)

$$\mathbf{U}'=\frac{1}{jk}(C+KB)\mathbf{P}-D\mathbf{U}, \tag{33}$$

$$\mathbf{P}'=-jkB\mathbf{U}+(E-D)\mathbf{P}, \tag{34}$$

where

$$D_{mn}=\begin{cases}\dfrac{R_2'-R_1'}{R_2-R_1}\left(1-\dfrac{\delta m0}{2}\right) & \text{for } m=n\\[2ex] A_mA_n((-1)^{m+n}R_2'-R_1')\dfrac{m^2}{m^2-n^2} & \text{for } m\neq n\end{cases} \tag{35}$$

and

$$E_{mn}=A_mA_n((-1)^{m+n}R_2'-R_1'). \tag{36}$$

Equations (33) and (34), combined with Eq. (19), lead to the new Riccati equation:

$$Z'=-jkB-\frac{1}{jk}Z(C+KB)Z+ZD-DZ+EZ. \tag{37}$$

We thus recognize Eqs. (14), (15), and (20) with additional terms. These terms are functions of R_1' and R_2', i.e., the variations of the cross section.

The same numerical procedures as described in Sec. II C can then be used to calculate the acoustic field or the reflection coefficient matrix of such geometries.

B. Results

The duct system considered here is the ''elephant's trunk'' represented in Fig. 8. The widths of the two semi-infinite straight ducts are such that $h_u/h_d=4$ and R_1 and R_2 are given by

$$R_1(\theta)=(h_u-h_d)\left(\frac{\theta}{\theta_f}\right)^3-\frac{3}{2}(h_u-h_d)\left(\frac{\theta}{\theta_f}\right)^2+R_m-\frac{h_d}{2} \tag{38}$$

and

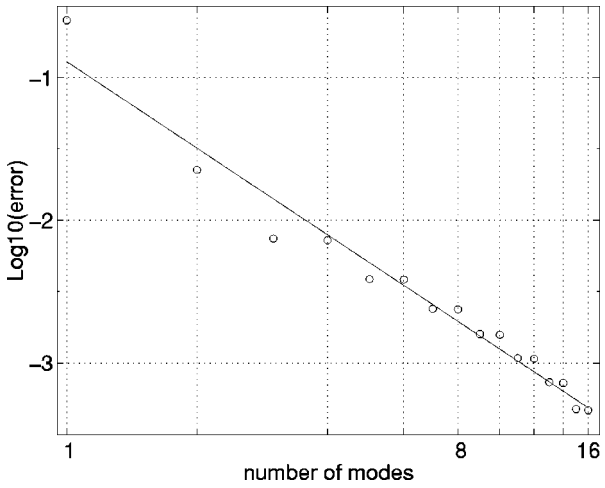


FIG. 9. Convergence of the MMM, in the case of the elephant's trunk (see Fig. 8). The logarithm of the error, $\log(\epsilon)$, is plotted as a function of the number of modes.

$$R_2(\theta)=(h_d-h_u)\left(\frac{\theta}{\theta_f}\right)^3-\frac{3}{2}(h_d-h_u)\left(\frac{\theta}{\theta_f}\right)^2+R_m+\frac{h_d}{2}, \tag{39}$$

where $R_m=1.25h_u$ and $\theta_f=2.62$. The pressure field (real part) shown in Fig. 8 has been calculated using 20 modes, and with a frequency such that $kh_u=3$. A plane piston source is imposed at the inlet of the trunk.

In order to measure the rate of convergence of the reformulated method, the result obtained with 20 modes is chosen as the reference, and we calculate the error (31) as a function of the number of modes taken into account. The error is shown in Fig. 9: The convergence follows a $1/N^{2\pm0.5}$ law, N being the number of modes introduced in the calculation. The convergence rate is lower than in the case of a constant cross section, although it remains high, because the basis functions ψ_n do not satisfy the boundary condition of vanishing normal derivative. Furthermore, one should mention that the convergence rate of an error based on derivatives of the pressure would also decrease. Work is in progress to improve the convergence by taking mixed basis of functions.[17]

Moreover, we can see in Fig. 9 an unexpected behavior: The addition of antisymmetric transverse modes (N even) in the calculation seems to have no effect on the error, as if their contribution to the solution was negligible, with the exception of the first antisymmetric mode. This is probably due to the nature of the coupling between the modes. We recall that two ways of coupling are to be considered in such geometry: One due to the curvature of the duct and the other due to the varying cross section. The first may contribute to generation of the first higher order modes only, the second being then preponderant for the remaining higher order modes. This latter is responsible for generation of symmetric modes only, because of the nature of the source (a plane piston) in our example. It may explain the weak influence of the antisymmetric modes.

Reflection coefficient. For the trunk described previously, the amplitude reflection coefficient is calculated at the

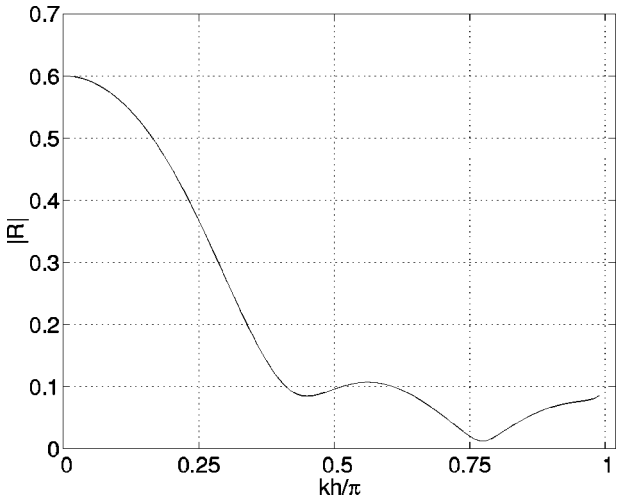


FIG. 10. Variations of the amplitude reflection coefficient with the frequency kh_u/π, calculated at the inlet of a trunk, as described in Fig. 8.

inlet of the curved part ($\theta=\theta_f$) and given in Fig. 10 as a function of the frequency parameter kh_u, varying between 0 and π.

As could be expected, the reflection coefficient tends toward

$$\left|\frac{h_e-h_s}{h_e+h_s}\right|=0.6 \tag{40}$$

for $kh_u=0$. This limit value corresponds to the low frequency reflection coefficient of a section discontinuity in a straight duct. As in the case of a circular bend, the effects of curved ducts vanish at low frequency.

VI. CONCLUSION

A multimodal method for the calculation of the sound propagation in bends has been formulated and validated. It is suitable for any dimension of bend and in any domain of frequency. In contrast to the bend mode method, the difficulties related to the necessary calculation, for each frequency, of the angular wave numbers are avoided. This approach is also convenient both for the formulation of radiation conditions and when considering junctions with ducts of various geometries, such as straight waveguides with varying cross section that have already been treated by modal decomposition.[10,11]

The influence of the evanescent modes generated at the discontinuities on both sides of the bend has been studied, showing the importance of a method that takes into account these evanescent modes.

The multimodal method has been successfully generalized to curved ducts with varying cross section, showing the large field of applications of this method.

Work is in progress to formulate the multimodal method to three-dimensional bends of circular cross section, in which the wave equation is not separable.

APPENDIX A

We want to project the equation of momentum conservation $-jkrv_\theta=\partial p/\partial\theta$ on the basis of functions ψ_n. We have, for any $m\geqslant 0$,

$$-jk\int_{R_1}^{R_1+h} rv_\theta\psi_m\,dr=\int_{R_1}^{R_1+h}\frac{\partial p}{\partial\theta}\psi_m\,dr. \tag{A1}$$

Using the series decomposition (7) and (8) of p and v_θ, Eq. (A1) gives

$$-jk\sum_{n\geqslant 0}\left(\int_{R_1}^{R_1+h} r\psi_m\psi_n\,dr\right)U_n=\frac{\partial}{\partial\theta}P_m, \tag{A2}$$

thus

$$-jkB\mathbf{U}=\mathbf{P}', \tag{A3}$$

with B given by Eq. (17).

The procedure is the same for the equation of mass conservation (5).

APPENDIX B

We give in this appendix the expressions for the matrices B and C defined in Sec. II A [Eqs. (17) and (18)]:

$$B_{mn}=\begin{cases} R_1+\dfrac{h}{2} & \text{for } m=n \\ A_mA_n\left(\dfrac{h}{\pi}\right)^2((-1)^{m+n}-1)\dfrac{m^2+n^2}{(m^2-n^2)^2} & \text{for } m\neq n \end{cases} \tag{B1}$$

$$C_{mn}=\begin{cases} 0 & \text{for } m=n \\ A_mA_n((-1)^{m+n}-1)\dfrac{m^2}{m^2-n^2} & \text{for } m\neq n. \end{cases} \tag{B2}$$

APPENDIX C

Considering the duct system of Fig. 1, Z, $\mathbf{U}$, and $\mathbf{P}$ can be calculated in the straight region upstream from the bend, where P and U satisfy[10]

$$\mathbf{P}^{(S)}=D_1\mathbf{P}^{(I)}+D_2Z_c\mathbf{U}^{(I)} \tag{C1}$$

and

$$\mathbf{U}^{(S)}=D_2Z_c^{-1}\mathbf{P}^{(I)}+D_1\mathbf{U}^{(I)}, \tag{C2}$$

where the superscripts (I) and (S) indicate, respectively, the inlet of the bend and a point upstream, the two being distant from d, as shown in Fig. 1, D_1 and D_2 are diagonal and defined by $D_1=\cos(k_nd)$ and $D_2=j\sin(k_nd)$. Substituting Eq. (19) into Eq. (C1) and defining $D_3=j\tan(k_nd)$, we can obtain a relation between the impedance at the point S upstream from the bend and its value $Z^{(I)}$ at the inlet:

$$Z^{(S)}=D_3(I+D_2^{-1}Z^{(I)}(Z^{(I)}+D_3^{-1}Z_c)^{-1}D_2^{-1})Z_c, \tag{C3}$$

where I is the identity matrix. It is now possible to propagate the pressure or velocity down the straight duct by using

$$\mathbf{U}^{(I)}=(-D_2Z_c^{-1}(Z^{(S)}-Z_c)+e^{-jk_nd})\mathbf{U}^{(S)} \tag{C4}$$

derived from Eqs. (C1) and (C2).

Equation (C4) is greatly simplified in the semi-infinite straight duct downstream from the bend, since the impedance at each point is the characteristic impedance Z_c.

APPENDIX D

An algebraic method for the calculation of the reflection and transmission matrices, in the same manner as Tam,[7] is developed in this appendix.

The duct system studied is represented in Fig. 1, consisting in a bend joining two semi-infinite straight ducts. The incident and reflected waves in region (I) and the transmitted wave in region (III) can be written as follows:

$$p^{(i)}=\sum_n P_n^{(i)}\psi_n(x)e^{j(\omega t-k_n z)}, \tag{D1}$$

$$p^{(r)}=\sum_n P_n^{(r)}\psi_n(x)e^{j(\omega t+k_n z)}, \tag{D2}$$

and

$$p^{(t)}=\sum_n P_n^{(t)}\psi_n(\zeta)e^{j(\omega t-k_n \eta)}. \tag{D3}$$

In region (II), we deduce from Eqs. (14) and (15) a second-order differential equation for $\mathbf{P}$:

$$\mathbf{P}''+B(C+KB)\mathbf{P}=0. \tag{D4}$$

A general solution of Eq. (D4) can be constructed in terms of the eigenvalues $\nu_1^2,\nu_2^2,\ldots$ and eigenvectors $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\ldots$ of the matrix $B(C+KB)$:

$$\mathbf{P}=XD(\theta)\mathbf{C}_1+XD^{-1}(\theta)\mathbf{C}_2, \tag{D5}$$

where $X=[\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\ldots]$, $D(\theta)$ is diagonal and given by

$$D_n(\theta)=e^{j\nu_n\theta} \tag{D6}$$

with

$$\nu_n=\begin{cases}\sqrt{\nu_n^2} & \text{for propagative modes,}\\ -j\sqrt{-\nu_n^2} & \text{for evanescent modes,}\end{cases} \tag{D7}$$

$\mathbf{C}_1$ and $\mathbf{C}_2$ are arbitrary constant column vectors. The term $XD(\theta)\mathbf{C}_1$ in Eq. (D5) corresponds thus to the right going waves [toward region (III)], while the second term corresponds to the left going waves [toward the region (I)].

We now match the solution for the pressure in the bend and its normal derivative to the external solutions (D1), (D2), and (D3).

At $\theta=0$,

$$\mathbf{P}^{(t)}=X(\mathbf{C}_1+\mathbf{C}_2), \tag{D8}$$

$$jkE\mathbf{P}^{(t)}=B^{-1}X\nu^{-1}(\mathbf{C}_1-\mathbf{C}_2), \tag{D9}$$

and similarly at $\theta=\theta_f$,

$$\mathbf{P}^{(i)}+\mathbf{P}^{(r)}=XD\mathbf{C}_1+XD^{-1}\mathbf{C}_2, \tag{D10}$$

$$jkE(\mathbf{P}^{(i)}-\mathbf{P}^{(r)})=B^{-1}X\nu^{-1}(D\mathbf{C}_1-D^{-1}\mathbf{C}_2), \tag{D11}$$

with $D=D(\theta_f)$, E is diagonal and given by $E_n=k_n/k$, and where the n^{th} term of the diagonal matrix ν is $1/j\nu_n$. We have $B^{-1}X\nu^{-1}=-jkHY$, with $H=(1/jk)(C+KB)$ and $Y=X\nu$. Equations (D8)–(D11) thus become

$$\mathbf{P}^{(t)}=X(\mathbf{C}_1+\mathbf{C}_2), \tag{D12}$$

$$-E\mathbf{P}^{(t)}=HY(\mathbf{C}_1-\mathbf{C}_2), \tag{D13}$$

$$\mathbf{P}^{(i)}+\mathbf{P}^{(r)}=XD\mathbf{C}_1+XD^{-1}\mathbf{C}_2, \tag{D14}$$

$$-E(\mathbf{P}^{(i)}-\mathbf{P}^{(r)})=HY(D\mathbf{C}_1-D^{-1}\mathbf{C}_2). \tag{D15}$$

The reflection matrix R and the transmission matrix T are defined, respectively, by $\mathbf{P}^{(r)}=R\mathbf{P}^{(i)}$ and $\mathbf{P}^{(t)}=T\mathbf{P}^{(i)}$. We find them by solving the set of equations above, taking care to keep only the matrix D^{-1} in the expressions of R and T. D is indeed a source of numerical problems, since terms such as $\exp(\sqrt{-\nu_n^2}\theta_f)$ can be exceedingly large when ν_n^2 is large and negative. We thus find

$$R=-\Delta^{-1}\tilde{\Delta}, \tag{D16}$$

$$T=4\Delta^{-1}D^{-1}(EX-HY)^{-1}E, \tag{D17}$$

with

$$\Delta=D^{-1}(EX-HY)^{-1}(EX+HY)\times D^{-1}(X^{-1}+Y^{-1}H^{-1}E)+(X^{-1}-Y^{-1}H^{-1}E), \tag{D18}$$

$$\tilde{\Delta}=(X^{-1}+Y^{-1}H^{-1}E)+D^{-1}(EX-HY)^{-1}\times(EX+HY)D^{-1}(X^{-1}-Y^{-1}H^{-1}E). \tag{D19}$$

APPENDIX E

We give in this appendix useful elements to project the two Euler equations (5) and (6), paying attention to the dependence of the basis functions ψ_n with θ and to the correct boundary condition

$$\frac{\partial p}{\partial n}=\frac{\partial p}{\partial r}-\frac{R'}{R^2}\frac{\partial p}{\partial \theta}=0. \tag{E1}$$

We use Leibniz's rule to get the projection for $\partial p/\partial\theta$ and $\partial v_\theta/\partial\theta$. From

$$\int_{R_1}^{R_2}\frac{\partial p}{\partial\theta}\psi_m\,dr=\int_{R_1}^{R_2}\frac{\partial(p\psi_m)}{\partial\theta}dr-\int_{R_1}^{R_2}p\frac{\partial\psi_m}{\partial\theta}dr, \tag{E2}$$

we deduce

$$\int_{R_1}^{R_2}\frac{\partial p}{\partial\theta}\psi_m\,dr=\frac{\partial}{\partial\theta}\left(\int_{R_1}^{R_2}p\psi_m\,dr\right)-\int_{R_1}^{R_2}p\frac{\partial\psi_m}{\partial\theta}dr-R_2'[p\psi_m](R_2)+R_1'[p\psi_m](R_1), \tag{E3}$$

which implies

$$\int_{R_1}^{R_2}\frac{\partial p}{\partial\theta}\psi_m\,dr=P_m'+D_{mn}P_n-E_{mn}P_n, \tag{E4}$$

with the matrices D and E given by Eqs. (35) and (36).

Moreover, when projecting $\partial_r(r\partial_r p)$ in Eq. (5), a term $[r(\partial_r p)\psi_m]_{R_1}^{R_2}$ appears. We calculate this term by substituting Eq. (6) into Eq. (E1), to obtain

$$\frac{\partial p}{\partial r}=-jk\frac{R'}{R}u. \tag{E5}$$

Finally, we have

$$\left[r \frac{\partial p}{\partial r} \psi_m \right]_{R_1}^{R_2} = -jkE_{mn}U_{mn}. \tag{E6}$$

[1] W. Rostafinski, ''Monograph on propagation of sound waves in curved ducts,'' NASA Reference Publ. 1248 (1991).
[2] P. E. Krasnushkin, ''On waves in curved tubes,'' Uch. Zap. Mosk. Gos. Univ. **75**, 9–27 (1945).
[3] F. E. Grigor'yan, ''Theory of sound wave propagation in curvilinear waveguides,'' Akust. Zh. **14**, 376–384 (1968) [English translation: Sov. Phys. Acoust. **14**, 315–321 (1969)].
[4] W. Rostafinski, ''On propagation of long waves in curved ducts,'' J. Acoust. Soc. Am. **52**, 1411–1420 (1972).
[5] A. Cummings, ''Sound transmission in curved duct bends,'' J. Sound Vib. **35**, 451–477 (1974).
[6] W. C. Osborne, ''Higher mode propagation of sound in short curved bends of rectangular cross section,'' J. Sound Vib. **45**, 39–52 (1976).
[7] C. K. W. Tam, ''A study of sound transmission in curved duct bends by the Galerkin method,'' J. Sound Vib. **45**, 91–104 (1976).
[8] A. Cabelli, ''The acoustic characteristics of duct bends,'' J. Sound Vib. **68**, 369–388 (1980).
[9] R. P. Dougherty, ''A Wave-splitting technique for nacelle acoustic propagation,'' AIAA Pap. **97**, 1652 (1997).
[10] V. Pagneux, N. Amir, and J. Kergomard, ''A study of wave propagation in varying cross section waveguides by modal decomposition. Part I. Theory and validation,'' J. Acoust. Soc. Am. **100**, 2034–2048 (1996).
[11] N. Amir, V. Pagneux, and J. Kergomard, ''A study of wave propagation in varying cross section waveguides by modal decomposition. Part II. Results,'' J. Acoust. Soc. Am. **101**, 2504–2517 (1997).
[12] A. F. Stevenson, ''Exact and approximate equations for wave propagation in acoustic horns,'' J. Appl. Phys. **22**, 1461–1463 (1951).
[13] R. J. Albertson, ''The propagation of sound in a circular duct of continuously varying cross-sectional area,'' J. Sound Vib. **23**, 433–442 (1972).
[14] A. Roure, ''Propagation guidée, étude des discontinuités,'' Doctoral thesis, Université d'Aix-Marseille, 1976.
[15] J. Kergomard, ''Calculation of discontinuities in waveguides using mode-matching method: An alternative to the scattering matrix approach,'' J. Acoust. **4**, 111–137 (1991).
[16] K. Lin and R. L. Jaffe, ''Bound state and threshold resonances in quantum wires with circular bends,'' Phys. Rev. B **54**, 5750–5762 (1996).
[17] C. Hazard and V. Pagneux, ''Improved multimodal approach in waveguides with varying cross-section,'' accepted in 17th I.C.A., Rome (2001).

Tomographic inversion of geoacoustic properties in a range-dependent shallow-water environment

Patrick Pignot and N. Ross Chapman[a)]
School of Earth and Ocean Sciences, University of Victoria, P.O. Box 3055, Victoria, British Columbia V8W 3P6, Canada

(Received 13 November 2000; revised 30 April 2001; accepted 29 May 2001)

This paper presents a matched-field tomographic method to estimate the geoacoustic properties of the ocean bottom for a range-dependent medium in shallow water. The inversion method has been developed in order to interpret experimental data from the Haro Strait PRIMER sea trial. This experiment was carried out in June '96 and used low-frequency broadband signals that were received on three vertical line arrays. Inversion of the data is particularly difficult because of the complex bathymetry of the Haro Strait experimental site. For this inversion, a range-dependent ray code was developed to solve the forward problem, allowing an arbitrarily layered bottom environment. The inversion scheme is based on modeling the propagation time and the amplitude of the recorded data, and a simple new cost function is proposed. The signal ray paths are identified automatically using a simple process that compares calculated and measured travel times. Data from multiple source positions are used to invert the range dependence of the geoacoustic model. The environment is separated into segments, and within each segment the inversion is carried out layer by layer for a multilayer geoacoustic model. Starting with the topmost layer, the range-dependent thickness and sound speed are estimated via a Monte Carlo method. Inversion results are presented for synthetic and experimental data from the Haro Strait sea trial. [DOI: 10.1121/1.1388000]

PACS numbers: 43.30.Cq, 43.30.Gv, 43.30.Pc [DLB]

I. INTRODUCTION

Matched-field (MF) inversion methods have been applied to estimate geoacoustic properties of the ocean bottom from data obtained in various shallow-water environments.[1–6] The experiments have generally been carried out with single vertical line arrays (VLAs), and the inversions have used data from a specific source position to estimate the local bottom properties near the array. The model parameter estimates obtained from this type of experimental geometry represent averaged values of the geoacoustic properties over the propagation range between the source and array. In this paper, we describe a new broadband MF inversion technique to estimate the geoacoustic properties of range-dependent environments. The inversion method was developed to process data from an experiment that was designed to investigate geoacoustic MF tomography in shallow water.

The MF tomography concept was first suggested by Tolstoy[7] for estimating geoacoustic profiles in range and cross range. The experiment envisaged by Tolstoy required multiple sources and arrays, with relatively large propagation ranges, much greater than the water depth. For this condition, the signals received at the arrays are modal, and assuming that the range dependence is adiabatic, the inversion can be linearized.[8] In our case, the experimental site imposed significant constraints on the design of the experiments. The propagation ranges were much shorter, less than two or three times the water depth, and therefore the signals consisted of multipath arrivals for specific acoustic paths. Consequently, we have developed a nonlinear range-dependent inversion method that is appropriate for ray-like data. The MF tomography experiments constituted part of the Haro Strait PRIMER Experiment, a multipurpose sea trial carried out in June 1996 to study coastal ocean processes.[9] The experiments made use of a multiple vertical line array system that was deployed at a site east of Vancouver Island at the northern end of Haro Strait, as indicated in Fig. 1. Impulsive sound sources generated by implosion of light bulbs were deployed from a small research vessel around the perimeter of the array system and within the area enclosed by the arrays in order to provide the multiplicity of acoustic paths to sample the bottom and subbottom in range, cross-range and depth. A preliminary inversion was previously reported that used the conventional single-source/single VLA geometry to obtain estimates of the average geoacoustic properties between the source and the array.[10] In this paper, the data from a series of light bulbs deployed on a radial track from one of the arrays are inverted to estimate the range dependence of the geoacoustic properties along the radial. The data are inverted using a multistage, coherent broadband method that matches measured and calculated travel times and magnitudes of the signals received at the array. A range-dependent ray theory code is used to calculate the modeled signal times and magnitudes. The method depends on the correct identification of specific acoustic travel paths, and an automatic identification process was developed for this purpose.

The paper is organized in the following way. The experiment is described in Sec. II. The multistage inversion method, in-

[a)]Author to whom correspondence should be addressed; electronic mail: chapman@uvic.ca

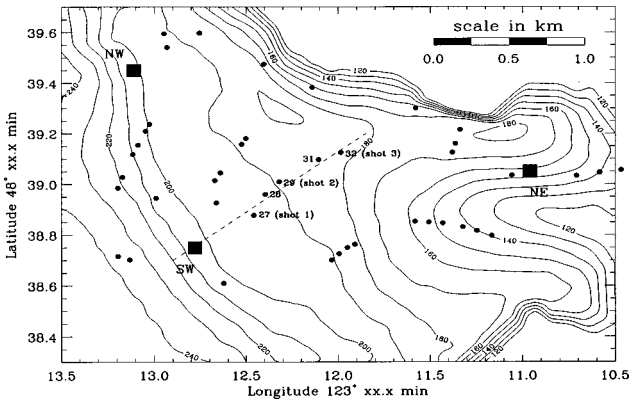


FIG. 1. Local bathymetry of the experimental site. Location of the light bulbs is indicated by the circles; squares indicate the position of the vertical arrays.

cluding the ray-path identification process, is presented in Sec. III. It is applied to synthetic data simulated for the Haro Strait environment, and then to the experimental data in Sec. IV. The estimated geoacoustic profile for the experimental site is discussed at the end of Sec. IV. The HARORAY ray propagation code suited for the particularly complicated bathymetry encountered in the Haro Strait sea trial is described in the Appendix.

II. EXPERIMENT

The geoacoustic MF tomography experiment was part of a collaborative sea trial that took advantage of a multiple vertical line array data acquisition system that was deployed south of Stuart Island in Haro Strait[9] (Fig. 1). Each array consisted of 16 hydrophones spaced at 6.25 m and a 1.5-kHz tomography source, with a total aperture of 100 m as shown in Fig. 2. The top hydrophone was approximately 30 m from the sea surface, so that the VLA spanned only the top half of the water column. The acoustic data were recorded at a sampling rate of 1750 Hz, and stored on hard drives at the surface float. The arrays were configured remotely for data acquisition by an rf ethernet command link.

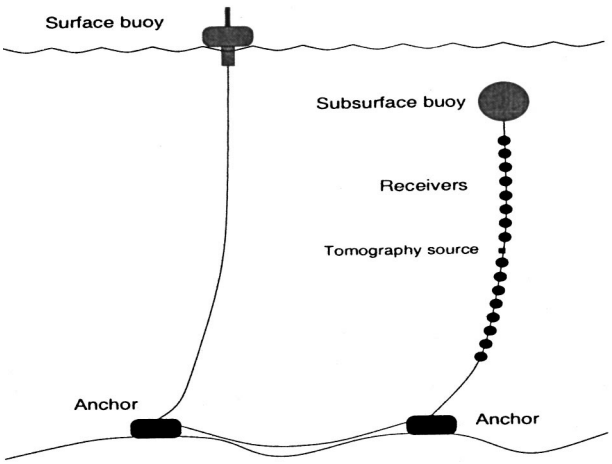


FIG. 2. Schematic diagram of the bottom-moored array. The separation of the two bottom weights was about 150 m, and the top hydrophone was about 30 m deep.

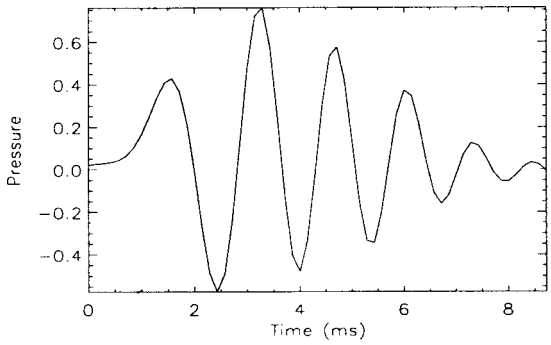


FIG. 3. Light-bulb wavelet $s(t)$ for a 70-m source depth.

Low-frequency broadband sound signals were provided by implosions of household light bulbs that were deployed from the Canadian research ship CSS R. B. YOUNG using a downrigger fishing line apparatus.[9] The light-bulb signals satisfied the two basic requirements for low-frequency sound sources in this experiment: (1) a short, reproducible impulse with adequate signal level to operate in the extremely high ambient noise environment in Haro Strait (in excess of 110 dB); and (2) a broadband signal within the 100–800-Hz band defined by the data acquisition system. The waveform for a 70-m light-bulb implosion is shown in Fig. 3. The center frequency is about 600 Hz, with a 3-dB bandwidth of about 300 Hz.[9,11] Forty-five light bulbs were deployed at nominal depths of 30, 50, and 70 m at the sites indicated in Fig. 1. The implosion data were recorded in 20-s segments, following a procedure to coordinate the shipboard operations with the data acquisition process that was controlled in a shore station on Vancouver Island at Sidney, B. C. Source depths were estimated using the length of line deployed, and the source position was measured using differential GPS. This information provided approximate estimates (to within 4–5 m for the depth); more accurate estimates were subsequently obtained from analysis of the acoustic travel time data from the light-bulb shot. The signal received at the southwest (SW) array from one of the bulbs is shown in Fig. 4. Several coherent signals from distinct acoustic paths are identified in the figure.

In addition to the light bulb deployments, a number of other tasks were also carried out on the CSS R. B. YOUNG to provide ground truth information for comparison with the acoustic inversions.

(i) The bathymetry of the region was surveyed using 38- and 200-kHz echosounders and differential GPS navigation over a detailed grid pattern; the bathymetry (corrected for tidal variation) shown in Fig. 1 is accurate within $\pm$ 0.5 m.

(ii) Samples of the bottom type (at the sea floor) were obtained using a Shipek grab sampling device and a trip corer at 13 sites within the area defined by the arrays.

(iii) The sound-speed profile in the water was measured using a velocimeter immediately before and after the light-bulb deployment period. The profile was essentially independent of depth with an average value of 1482.5 m/s during the light-bulb experiments. The

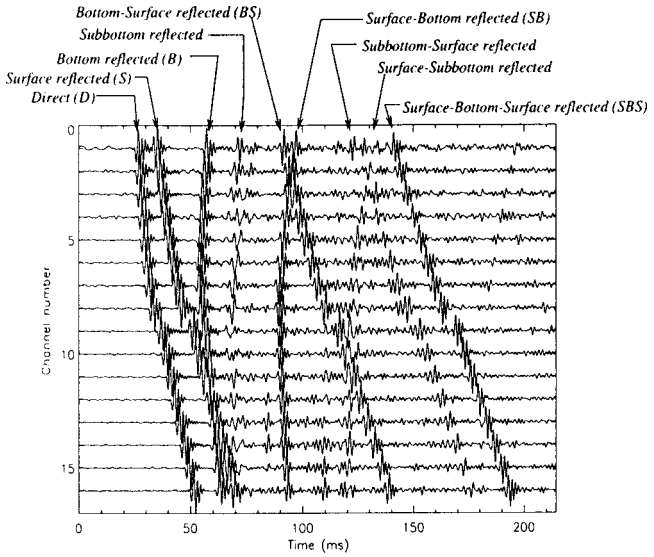


FIG. 4. Pressure signal recorded at the SW array (shot 29). Hydrophone 1 represents the shallowest receiver and hydrophone 16 the deepest. The acoustic paths are labeled at the top of the figure.

difference between the sound speeds at the top and bottom of the water column was less than 0.5 m/s.

III. INVERSION METHOD

A. Principle

The objective of this inversion is the estimation of geoacoustic parameters such as sound speeds and depths of subbottom layers in a range-dependent environment. In designing an efficient and effective approach for MF inversion of the Haro Strait data, there are several important considerations: (1) the bathymetry is strongly range dependent in portions of the experimental site, particularly near the northeast (NE) array; (2) the data are broadband; and (3) there is imprecise knowledge of the experimental geometry. In addition to these constraints imposed by the experiment there are other significant factors that affect the design: the number of model parameters to estimate in a range-dependent inversion is very large, and the sensitivity of specific geoacoustic parameters is known to vary over several orders of magnitude. In order to reduce the calculation cost, only the most sensitive parameters are estimated in the inversion described here; all others are fixed using *a priori* information.

The first two constraints strongly affect the choice of the propagation model. Since the data consist of coherent signals from specific acoustic paths, our approach is based on ray theory for calculation of acoustic field quantities. Ray theory provides rapid computation of travel times and amplitudes for constructing the broadband signal, and is relatively straightforward to apply in range-dependent environments. This approach has also been used by Porter *et al.*[12] for the analysis of INTIMATE96 data. The third constraint is significant for the relatively high-frequency signal from the light-bulb implosions, since the hydrophone and source positions must be known to within a fraction of a wavelength (about 1 meter).

The inversion makes use of data from multiple sources deployed along a radial track from one of the VLAs to estimate the range dependence of the geoacoustic properties. Other investigations reported previously have suggested the simultaneous inversion of geometric parameters and geoacoustic properties.[4,13] However, our experience with the Haro Strait data has shown that it was not effective to invert for all the different types of parameters at the same time. Instead, our approach is to establish the experimental geometry first in a separate inversion, and then invert for the geoacoustic properties of the waveguide. The geoacoustic inversion itself is decomposed into a sequence of steps. We assume that the environment is separated into segments and within each segment the medium consists of homogeneous sediment layers overlying a halfspace, as indicated in Fig. 5. In this figure, the number of layers in the geoacoustic model is arbitrarily set at two. The segment boundaries are defined by the locations of the light-bulb shots. In each segment, the sound

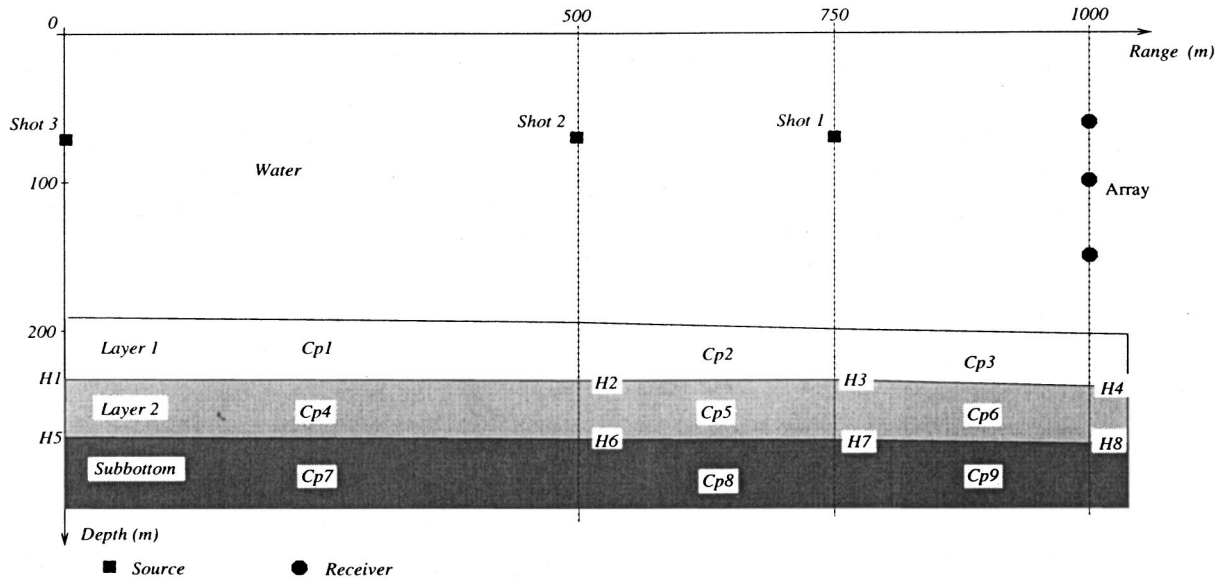


FIG. 5. Stratified geoacoustic ocean model. In this figure the range is measured from shot 3, the farthest shot from the array.

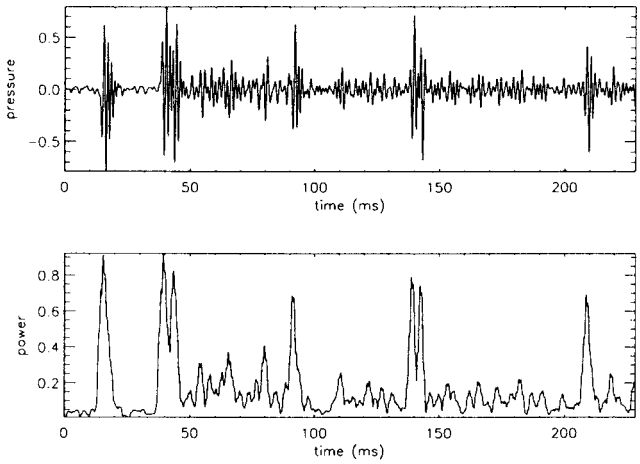


FIG. 6. Estimation of the energy of a light-bulb implosion recorded on the SW array. The top panel shows the measured time series, and the bottom panel shows the integrated data over the time window.

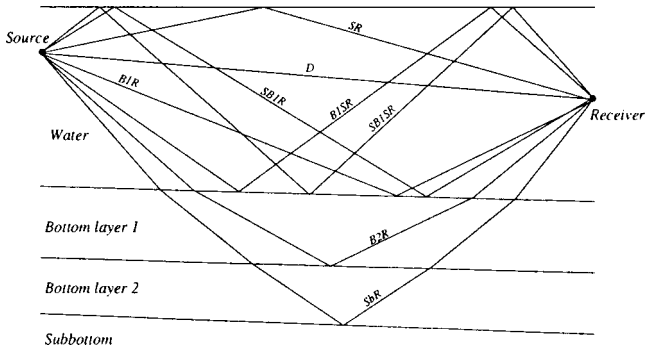


FIG. 7. Travel paths between the source and the receiver. Additional paths corresponding to the surface reflected paths for the subbottom reflections are not shown.

speed and the interface slopes are constant, but can be different for each layer.

The sound speed in the water and the bathymetry of the site were measured during the experiment and are assumed to be known. Starting with the first segment (from the array to shot 1 in Fig. 5), the inversion estimates the experimental geometry (ranges and source/array depths). When the geometry is known, the geoacoustic parameters are then estimated, layer by layer, starting with the topmost layer and continuing to the deeper layers, one at a time. In each step, the estimated parameter values in the upper layers are used to obtain the estimates for the values in the deeper layers. When the geoacoustic parameters in the first segment are known, the inversion proceeds to estimate the parameters in the next segment (between shots 1 and 2). The process is carried out in the same way: ranges and source depth are estimated first and then the geoacoustic parameters, layer by layer. The range-dependent inversion is thus obtained segment by segment, assuming known values for the properties in all preceeding segments.

B. Travel time and amplitude estimation

The experimental data recorded during the Haro Strait experiment consist of multipath signals that are generally well-resolved in time (see Fig. 4). The inversion method has been designed to make use of the travel time and magnitude information in the multipath signals. Both quantities are estimated from the data by a simple processing scheme based on the calculation of the averaged signal power, defined by

$$P=\frac{1}{T}\int_T |s(t)|^2 dt, \tag{1}$$

where T is a small duration of the recorded signal $s(t)$ (a few milliseconds). Figure 6 presents the result of processing the data for one of the sensors, using $T=1.3$ ms. The upper part of the figure shows the acoustic pressure that was first upsampled by a factor of 4, and the lower part the result of the averaged power estimation. The peak positions could be estimated to within 0.2 ms by this method.

C. Travel time and amplitude inversion

The inversion process is based on the correct identification of the different acoustic paths between the source and the array. The inversion compares the travel time (relative to the travel time for the direct path) and magnitude measurements for the different multipath signals with values calculated by ray theory for specific environmental models. The positions of the sources and arrays were monitored during the experiment, but not with sufficient accuracy. Consequently, the first step in the inversion process is the estimation of the experimental geometry: the depth $R_{d,i}$ and distance L_i to the source for the ith sensor, and the source depth S_d. This first inversion is based on the travel time of the signals that traveled only in the water. The recorded data generally show six different signals that traveled via water column paths: the direct (D), the surface reflected (SR), the bottom reflected (B1R), the bottom-surface reflected (B1SR), the surface-bottom reflected (SB1R), and the surface-bottom-surface (SB1SR) reflected path. The different ray paths are illustrated in Fig. 7.

The inversion of the travel time data is realized by minimizing a simple cost function

$$\mathrm{CF}_i(R_{d,i},L_i,S_d)=\overline{|\Delta T|}, \tag{2}$$

where $\overline{|\Delta T|}$ is the average of the modulus of the difference between the measured and the calculated travel times for the ith sensor. Since the implosion instants were not known, the travel times were determined relative to the direct path arrival time. The comparison between the synthetic and recorded travel time was done automatically using a simple method to identify the acoustic paths. We define T_k^{exp} to be the kth measured travel time for the kth multipath arrival and N the number of measured arrivals, and similarly, T_l^{th} the lth calculated travel time for the lth ray path and M the number of calculated travel times. In order to identify the type of acoustic path corresponding to each multipath arrival, we first assume a geometry for the source and array, calculate the travel times, and form the matrix $\overline{\overline{Q_{kl}}}$ of the differences between the measured and calculated travel times

$$\overline{\overline{Q_{kl}}} = \{|T_k^{\exp} - T_l^{\text{th}}|\}.$$

The position (kth row and lth column) of the minimum value of the $\overline{\overline{Q_{kl}}}$ matrix provides the connection between one experimental multipath arrival and one calculated travel time (or ray path): $T_k^{\exp}$ and T_l^{th}. In order to determine the connections between the other measured and calculated travel times, we remove the kth row and lth column of the $\overline{\overline{Q_{kl}}}$ matrix and repeat the minimization process described above until all the different connections are established. The maximum number of connections is $\min\{N,M\}$. For some sensors, more than one arrival may have nearly the same travel times, as for instance the BS and SB paths for channel 2 in Fig. 4. This condition does not present a problem, since the method identifies multiple elements in the same row with nearly equal travel times.

This method provides the advantage of being totally automatic without requiring any external decisions. The cost function $\text{CF}_i(R_{d,i}, L_i, S_d)$ has to be minimized for each sensor of the array.

When the experimental geometry has been determined, the inversion process continues to the second stage to estimate the range-dependent thickness and the speed of sound in the first bottom layer. The inversion used two types of rays, the bottom-reflected rays and the rays reflected at the interface between the first and second layers, e.g., the B2R path in Fig. 7. The travel time of the B2R path is sensitive to both the thickness and sound speed of layer 1. However, the magnitude of this path depends on the reflection coefficient at the interface between layers 1 and 2. Since the geoacoustic properties of layer 2 are unknown, the magnitude of the B2R path cannot be taken into account at this stage of the inversion process. On the other hand, although the travel times of the rays reflected at the water bottom are insensitive to the properties of layer 1, the magnitudes of the signals corresponding to these paths (B1R, SB1R, B1SR, and Sb1SR) depend on the reflection coefficient at the water/bottom interface and are thus sensitive to the geoacoustic properties of layer 1. The inversion is then realized by minimizing the cost function

$$\text{CF}(H3, H4, Cp3) = \overline{|\Delta T_{B1}|} \cdot \overline{|\Delta A_W|}, \tag{3}$$

where T_{B1} is the propagation time of the waves traveled in layer 1 and A_W the magnitude of the waves traveled in the water only. The identification of the different propagation paths is determined by the same method as described previously for comparing measured and calculated travel times.

The estimation of the sound speed and the range-dependent depth of the second layer is realized by the same method, assuming that the geoacoustic properties of the first layer are known. The cost function in this case is

$$\text{CF}(H7, H8, Cp6) = \overline{|\Delta T_{B2}|} \cdot \overline{|\Delta A_{B1}|}, \tag{4}$$

where T_{B2} is the propagation time of the waves traveled in layer 2 and A_{B1} the magnitude of the waves traveled in layer 1.

The last step is the estimation of the sound speed in the subbottom. Since the sound speed in the subbottom has no effect on the propagation time of the different waves, the cost function to minimize is based only on the magnitude of the waves reflected at the subbottom, and is given by

$$\text{CF}(Cp9) = \overline{|\Delta A_{B2}|}, \tag{5}$$

where A_{B2} is the magnitude of the waves traveled in layer 2.

The inversion of the data for the second shot provides estimates of the characteristics of the medium in the second segment between shot 2 and shot 1 (see Fig. 5). The characteristics of the medium between shot 1 and the array are now assumed to be known (i.e., the layer depths H3, H4, H7, H8 and the sound speed Cp3, Cp6, Cp9 that have been estimated previously). The inversion technique is the same as that for the first segment. The location of the sensors has to be determined first and then the inversion continues with the estimation of the sound speed and the thickness of each layer. For each layer, the number of parameters to estimate is reduced to two (compared to three for shot 1) because the layer depths (H3 and H7) at the location of shot 1 are now known.

The inversion can be carried out for several different shots. However, the distance between each shot has to be sufficient to provide adequate sampling in each segment of the reflection coefficient at the water bottom by the seafloor reflecting rays, and of the subbottom properties by the rays transmitted into the sediment. In practice, sufficient sampling can be obtained if the distance between the nth shot and the array is about twice the distance between the $(n-1)$th shot and the array.

D. Minimization of the cost function

The minimization process is a hybrid method based on a random process and the downhill simplex[14,15] method. For each inversion, the number of parameters to estimate is small (three parameters maximum) but the inversion has to be as efficient as possible because the process is repeated several times, depending on the number of segments and layers in the geoacoustic model. For each segment, one inversion has to be done to determine the geometry for each sensor of the array and three inversions for the estimation of the geoacoustic parameters in the synthetic example in Fig. 5.

The first step consists of globally sampling the model parameter space by choosing 100 models at random, and calculating the cost function for each set of parameters. Then, the simplex method is applied to converge quickly to a local minimum (over a maximum of 50 iterations), using a simplex defined by the best solutions from the global search. The process is repeated up to 10 times for a total number of between 1000 and 1500 models investigated.

IV. RANGE-DEPENDENT INVERSIONS

A. Synthetic data

The inversion process has been tested with synthetic data generated by a range-dependent ray code (HARORAY) for homogeneous layered environments as shown in Fig. 5. For the simulation, the water depth was 200 m, and the sound speed was 1482.5 m/s. Three different shots have been used: shot 1 at a range of 250 m from the vertical array, shot 2 at a range of 500 m, and shot 3 at a range of 1000 m. The vertical array consists of three sensors at depths of 60, 100,

TABLE I. Inversion results of synthetic data.

		No noise		20-dB SNR		10-dB SNR	
Parameter	True value	Estimated value	s.d.	Estimated value	s.d.	Estimated value	s.d.
Sd (m)	70.	70.01	0.04	69.88	0.07	70.35	0.49
Rd (m)	60.	59.94	0.04	59.91	0.10	58.84	0.38
Distance (m)	250.	249.99	0.07	249.75	0.19	248.42	0.85
Cp1 (m/s)	1600.	1599.6	1.3	1598.4	1.4	1600.4	3.8
Cp2 (m/s)	1610.	1610.0	0.4	1602.5	3.9	1610.3	13.5
Cp3 (m/s)	1620.	1620.0	0.2	1621.9	0.3	1632.6	17.2
Cp4 (m/s)	1750.	1751.5	1.9	1761.8	1.4	1754.3	0.9
Cp5 (m/s)	1765.	1763.6	1.3	1754.2	6.2	1767.5	19.3
Cp6 (m/s)	1780.	1779.4	0.1	1785.4	7.9	1784.1	42.4
Cp7 (m/s)	2200.	2199.5	3.8	2192.3	39.1	2330.7	120.2
Cp8 (m/s)	2225.	2225.3	1.3	2198.4	8.3	2304.6	121.9
Cp9 (m/s)	2250.	2248.8	0.6	2282.4	23.7	2308.5	95.9
H1 (m)	232.	230.3	1.3	233.1	1.1	238.4	5.6
H2 (m)	236.	236.3	0.2	235.1	0.4	229.7	2.8
H3 (m)	236.	235.8	0.1	235.7	0.1	240.6	2.0
H4 (m)	240.	240.2	0.1	240.3	0.1	239.9	1.8
H5 (m)	270.	269.2	0.8	267.8	1.3	268.1	2.1
H6 (m)	273.	273.3	0.4	271.1	1.1	272.2	2.7
H7 (m)	274.	274.0	0.1	273.8	0.3	272.7	2.7
H8 (m)	278.	278.0	0.1	278.5	0.5	283.9	2.6

and 150 m. The source depth is 70 m. The signal has the shape of a light-bulb implosion wavelet shown in Fig. 3. The synthetic waveforms $q_i(t)$ at the different sensors of the array are obtained by convolving the complex source waveform $S(t)$ with the complex impulse response $g_i(t)$ of the ith sensor

$$q_i(t) = S(t) * g_i(t). \tag{6}$$

The imaginary part of the source waveform is estimated from the Hilbert transform of the real part $s(t)$ of the source (measured signal). The complex signal $S(t)$ is defined by

$$S(t) = s(t) + iH[s(t)]. \tag{7}$$

The impulse response is calculated with the HARORAY code. The inversion estimates the nine sound speeds (Cp1 to Cp9) and the eight layer depths (H1 to H8) shown in Fig. 5. Each arrival was classified according to its propagation path (direct path, bottom reflecting path, etc.). The different travel times and amplitudes were estimated with the method described in Sec. III B. The inversion was realized by assuming that the location of the different sensors was known with an error of ± 25 m for the range and ± 10 m for the source or receiver depth.

The results for the three different shots are presented in Table I. In order to compare the different solutions, the cost functions were normalized, and the solution was expressed in terms of an energy parameter E, defined by

$$E = 1 - \frac{\mathrm{CF}}{\min(\mathrm{CF})}, \tag{8}$$

where min(CF) is the minimum value of the cost function. The energy parameter is always negative and the value $E = 0$ corresponds to the best estimation. However, the solution is generally nonunique since several different solutions may have an energy close to zero. In that case, the solution is given in terms of statistics defined by the mean value and the standard deviation of the different solutions with energy close to zero. It is then necessary to fix an energy limit above which the solutions are acceptable. The limit depends on the sensitivity of the parameters: in general, for less sensitive parameters, the energy limit needs to be closer to zero. However, it is very difficult to determine an appropriate value in practice, and some experimentation is necessary for each problem. For this simulation environment, the limit $E_l = -0.3$ is adequate for the geophysical parameters and $E_l = -4.0$ for the location estimation of the sensors, since the sensitivity of the geometrical parameters is much larger than the sensitivity of the geophysical parameters. The solutions for the estimated parameters are presented in Figs. 8, 9, and 10. The solid lines represent the mean value of the possible solutions, and the dotted lines represent the true value in each panel in the figures. The estimated value (mean value) and the standard deviation is given above each graph. The same inversion has been carried out with different signal-to-noise ratios (SNR) of 20- and 10-dB, obtained by adding zero mean white noise to the different waveforms $q_i(t)$. The main effect of the noise is to modify the value of the travel time and the energy estimated for each signal using the estimation process described in Sec. III B.

The inversion results presented in Table I show that the estimated parameter values are excellent for the noise-free case, and still very accurate with a 20-dB SNR. The inversion algorithm is able to reproduce the range variation of the sound speed with an accuracy that strongly depends on the noise level. The sediment layer depths are estimated within 1 or 2 m for low noise and within 5 m for a 10-dB SNR. The accuracy of the geometric parameters is always good, even with a 10-dB SNR. However, the precision of the estimates of subbottom sound speeds (Cp7, Cp8 and Cp9) is generally poor. The main characteristic of this inversion is that the

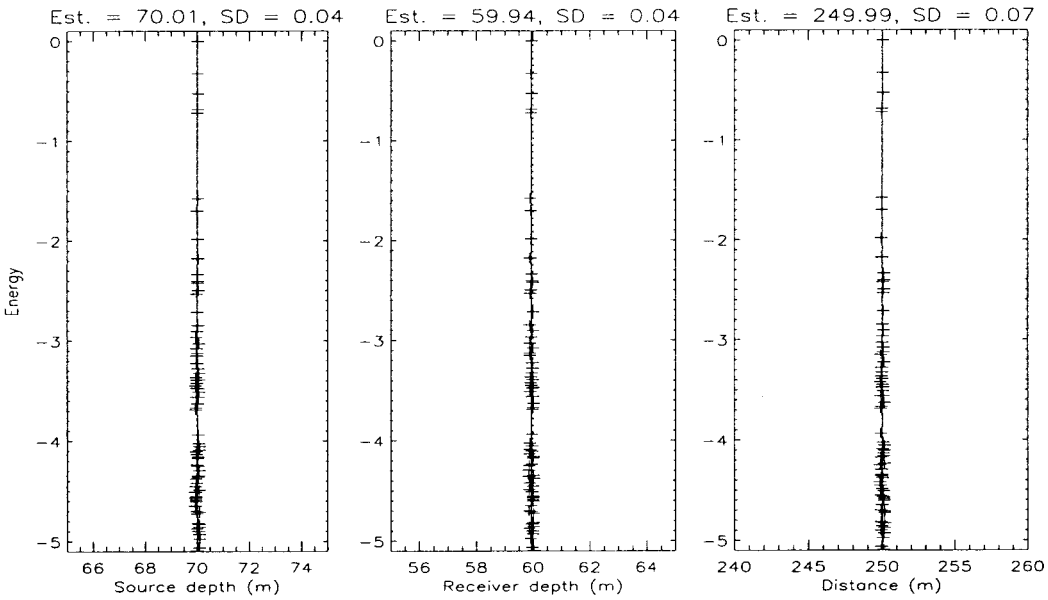


FIG. 8. Estimated source depth, receiver depth, and distance of one element of the array for the noise-free simulation.

estimation accuracy of the geoacoustic parameters decreases with range. This effect is clearly shown in Figs. 9 and 10. For the closer segments, the estimated values are accurate and the distribution of solutions is narrow. Estimation accuracy decreases and the width of the distribution increases with increasing range of the segment. Nevertheless, the algorithm has no stability problem and always quickly converges to a solution or a set of possible solutions. Each simulation requires less than 1 s (using a Pentium 200 MHz) and the estimation of a set of parameters is made in around 10 or 15 min.

B. Experimental data

The inversion has been carried out using data from three different shots that were received at the SW array (27, 29, and 32; see Fig. 1). The three shots and the array are aligned approximatively in a plane. The SNR for the shot signals ranged between 15 and 20 dB at the array, and the recordings for all three shots show the six paths corresponding to propagation in the water and three other paths from a subbottom layer, as in Fig. 4 for shot 29. A geoacoustic model with one sediment layer is appropriate for these data. The inversion

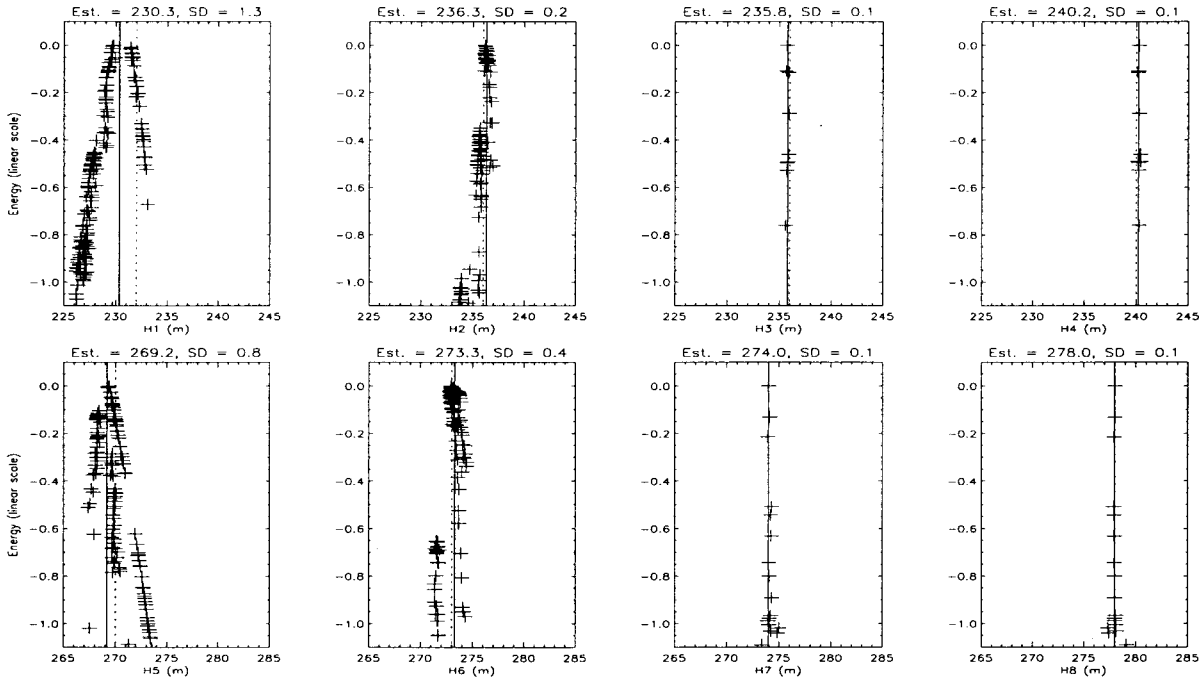


FIG. 9. Estimated thicknesses for the noise-free simulation. The top four panels show the results for the first sediment layer, and the lower four panels show the results for the second sediment layer.

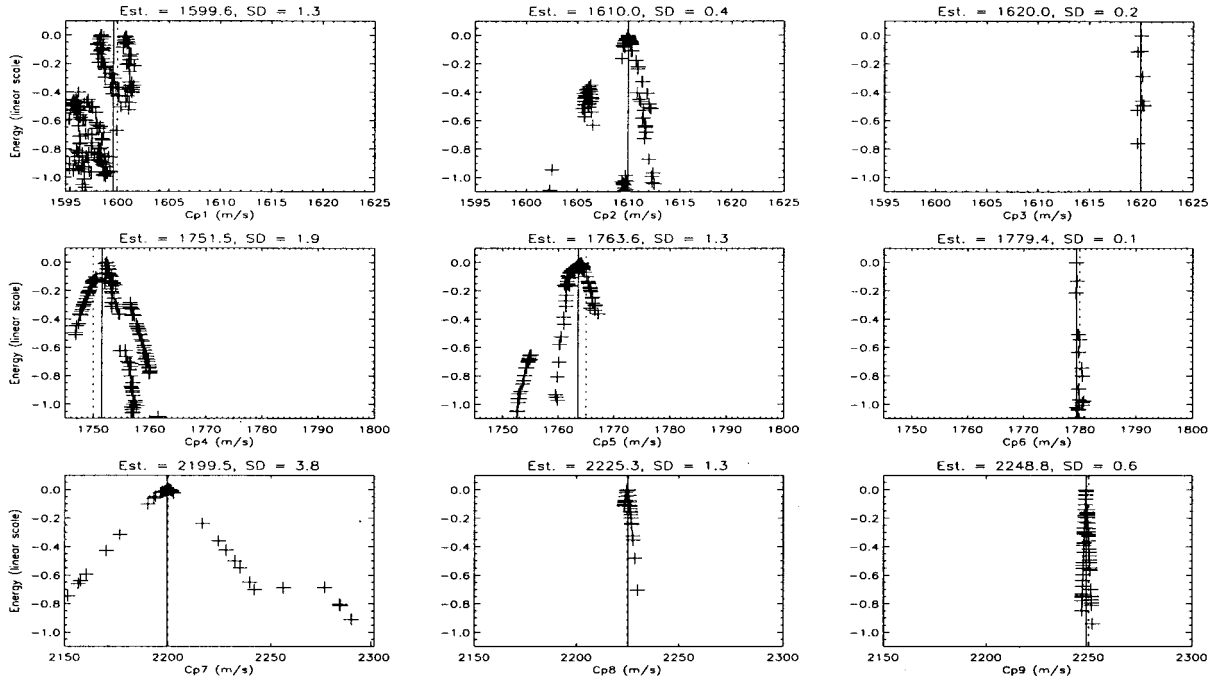


FIG. 10. Estimated sound speed for the noise-free simulation. The top three panels show the results for the first sediment layer, the middle three panels show the results for the second sediment layer, and the lower three panels show the results for half-space.

method described in Sec. III was applied to the data for the three shots to estimate first the location of the array hydrophones and the source, and then the range-dependent sound speed in the layer and in the subbottom, and the depth of the layer. The bathymetric measurements from the experiment were used for the water depths in solving the forward model, and the sound speed of the water was set at the measured value of 1482.5 m/s. The densities have been fixed at 1.7 in the bottom layer and 1.9 in the subbottom.[16]

The inversion results are summarized in Fig. 11, and the distribution functions for E are presented in Figs. 12 and 13. The best inversion result is obtained for the estimation of the experimental geometry, using the method described in Sec. III B. The source and receiver depths were estimated with an error less than 1 m, and the range estimation error was around 5 m over a total distance of 1 km. Figure 12 and Fig. 13 (top row) show that the sensitivities of the range-dependent layer depths and the sound speeds in the first layer are high, since the distributions of the different possible solutions are relatively narrow. On the other hand, the sensitivity of the sound speed in the subbottom is very low and the distribution is very wide, ranging between 2000 and 2500 m/s. Figure 14 shows a comparison between the recorded data for shot 29 (top) and the synthetic waveforms after inversion (bottom). The data indicate strong reflections from the sea bottom, and also comparatively strong reflections from the subbottom layer. By comparing the plots in the figure, the arrivals corresponding to the bottom-reflected paths are very well estimated in terms of travel time, amplitude, and phase, except for the path corresponding to the SBS reflection (at 140 ms for channel 1), which has a lower energy than that for the recorded data. However, the relative difference in signal strength between the bottom-reflected and subbottom-reflected paths is not well modeled, although the relative strength of the three subbottom-reflected paths for the modeled signals is consistent with the data.

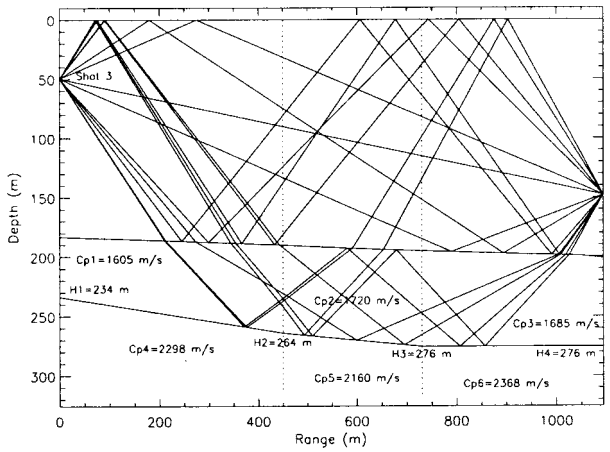


FIG. 11. Range-dependent inversion results for the experimental data. A subset of the acoustic paths for the farthest shot is shown in the figure to indicate the ray coverage. Segment 1 is closest to the array.

The fact that the subbottom-reflected paths and the bottom-reflected paths have nearly the same energy provides information about the impedance contrast at the two interfaces. The inversion algorithm can adopt two strategies to model the data. With the densities held constant, the sound speed in the sediment layer can be set to a high value, providing a high reflection coefficient for the bottom-reflected paths and a low energy for the subbottom-reflected paths; or, the sound speed in the layer can be estimated at a low value, providing a good fit between the experimental and modeled energies of subbottom-reflected paths but a high mismatch for the bottom-reflected paths. The inversion algorithm converges to a compromise between these two possibilities, selecting bottom and subbottom reflection coefficients that

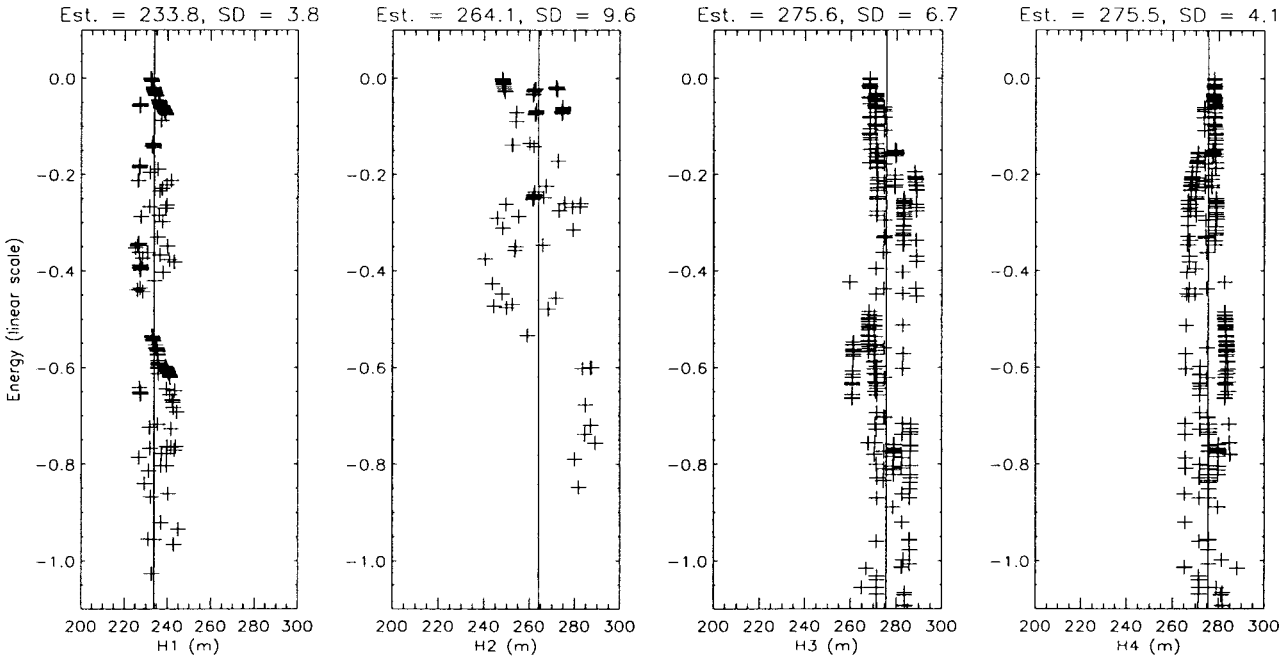


FIG. 12. Estimated layer thicknesses for experimental data.

model the bottom-reflected paths well and provide a reasonable fit for the later subbottom-reflected paths (after 100 ms for channel 1). Figure 14 shows a good fit in terms of propagation time and phase between the experimental and synthetic subbottom-surface (about 135 ms for channel 1) and surface-subbottom (about 130 ms for channel 1) reflected paths, but a mismatch occurs between the experimental and synthetic subbottom reflected path. The inversion estimates presented in Fig. 11 show that the slope of the subbottom layer increases towards the array, a result that is consistent with the measured seafloor slope. The sediment sound speed in the two segments between shot 2 and the array are con-

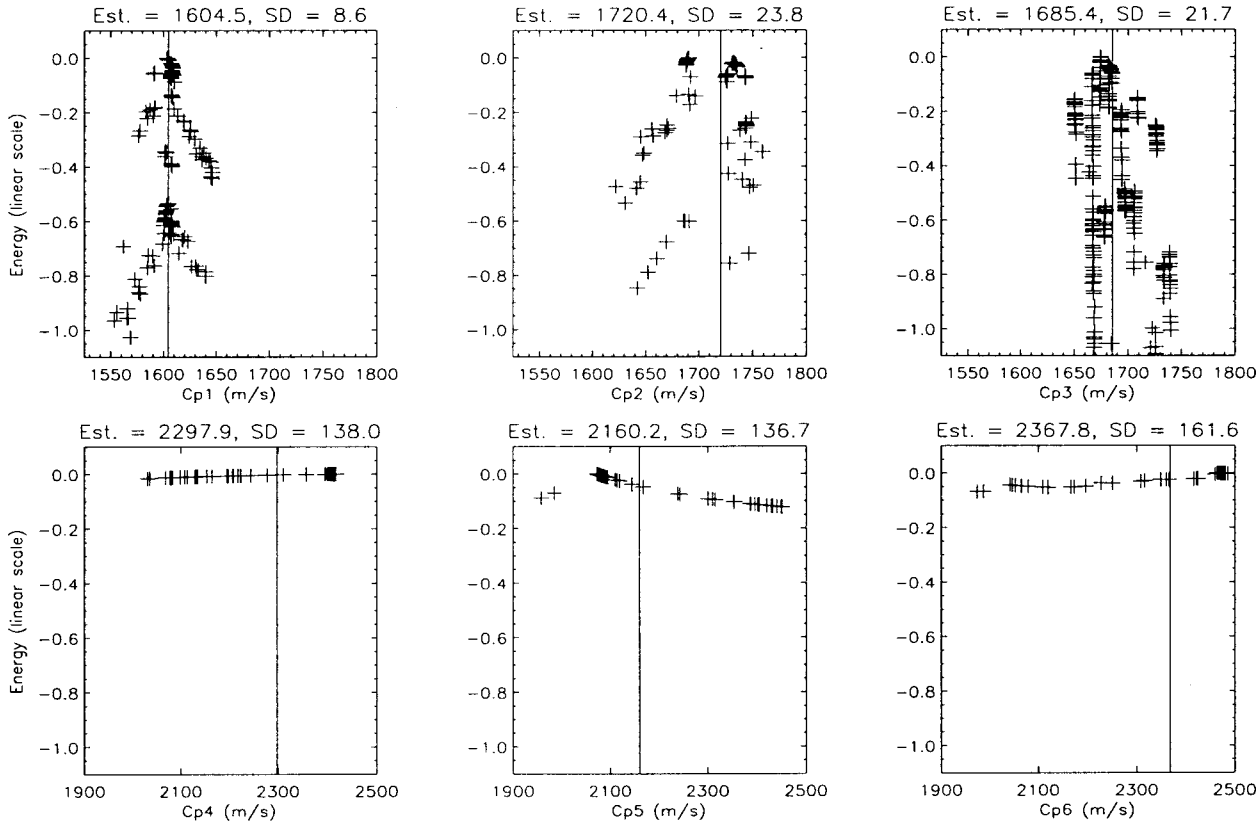


FIG. 13. Estimated sound speeds for experimental data. The top three panels show the estimates for the sediment layer, and the lower three panels show the estimates for the half-space.

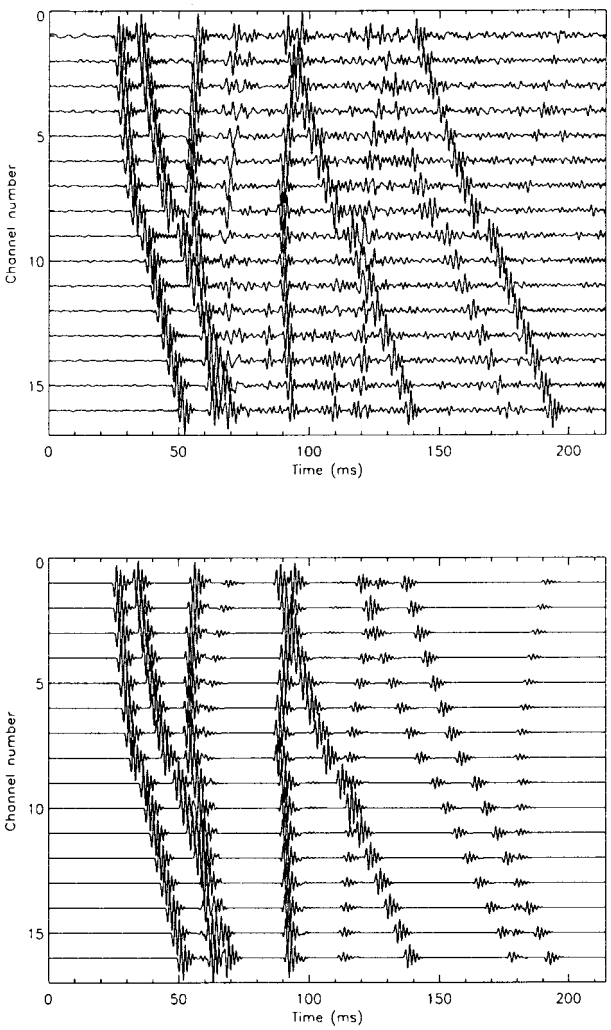


FIG. 14. Comparison between the recorded data for shot 29 (upper panel) and the calculated waveforms using the estimated model parameters (lower panel).

sistent at about 1700 m/s. The estimated sound speed in the corresponding layer of segment 3 is much lower than the value of 1700 m/s. This result may be due to the reduced estimated layer thickness. Also, the error is expected to be largest for segment 3 (between shot 3 and shot 2) since the inversion for this segment depends on the estimates from the two first segments, and since the density of rays is lowest in the bottom layer for this segment. Figure 11 clearly shows that the highest angle subbottom-reflected rays strike the interface in segment 3 between the 350- and 450-m range. No significant eigenrays are reflected on the subbottom interface for the first 350-m range.

The estimated values in this inversion are based on a homogeneous layer model. This constraint excludes the possibility that the relatively strong subbottom-reflected paths could be due to refraction by gradients within the sediment layers.

V. CONCLUDING REMARKS

In this paper, we have presented an inversion method capable of estimating range-dependent geoacoustic properties. The method is based on the relative travel time and magnitude measurements of the most significant signal paths. A method for automatic identification of signal ray paths was developed to improve the efficiency of the inversion process. The inversion provides a measure of the confidence limit of the estimate and indications of sensitivities. The drawback of the method is that the errors in the estimated values are cumulative, since the inversion for each successive segment assumes known geoacoustic properties from the inversions using the preceding segment. Modifying the algorithm to update previously determined estimates is possible, but at the cost of increasing the number of parameters that are estimated in the successive inversions. The inversion method has been applied to synthetic waveforms and experimental data from the Haro Strait sea trial. The main results for experimental data have shown that the slope of the subbottom layer can be determined with a relatively good accuracy, and that the sound speed of the sediment layer is uniform along most of the radial track. The observed uncertainties in experimental data at around 15 dB are consistent with those obtained for the simulation with SNR of around 10 dB.

ACKNOWLEDGMENTS

This work is supported by a grant from the Ocean Acoustics team of the Office of Naval Research, grant N00014-96-1-0721. The assistance and co-operation of the captain and crew of the R. B. YOUNG during the HARO Strait sea trial is greatly appreciated.

APPENDIX

HARORAY is a broadband 2D propagation model that uses ray theory to predict the acoustic field in range-dependent multilayered environments. This ray code has been written to provide synthetic waveforms that simulate the signals used in the Haro Strait experiment. The model was developed for a single point source and a system of many receivers aligned vertically. The sound speed of the water column or of the bottom layers has to be homogeneous versus depth but may change versus range.

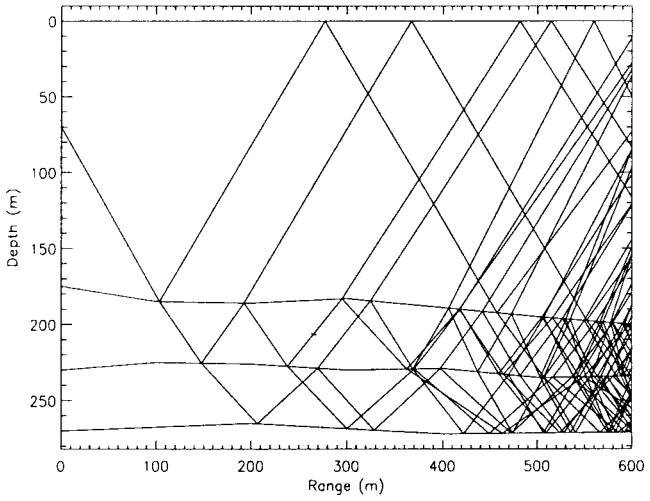


FIG. A1. Proliferation of rays in a multilayered environment.

The significant feature of the HARORAY code is the determination of the different eigenrays that connect a source point to a receiver point in terms of the propagation time, amplitude, and phase of the ray. The presence of a multilayered system implies a proliferation of rays at the different interfaces, since each ray is divided into two parts at each interface, providing a transmitted and a reflected ray. The HARORAY code determines all the eigenrays over a range of angles at the source point defined by a minimum and a maximum angle value. Starting with a given angle θ at the source point, the HARORAY code determines all the trajectories of the initial ray that arrive at the receiver range as presented in Fig. A1. Each trajectory is classified by a binary code describing the order of the reflections (0) and transmissions (1) on the interfaces. For example, the binary code 010 corresponds to a reflected-transmitted-reflected path. For a small variation of the starting angle $\theta+\Delta\theta$, the HARORAY code computes the new trajectories. Let Z_θ^i be the set of the different arrival depths corresponding to the starting angle θ and $Z_{\theta+\Delta\theta}^j$ the arrival depths corresponding to the starting angle $\theta+\Delta\theta$. The eigenrays are determined by comparing the receiver depth and the different values of Z_θ^i and $Z_{\theta+\Delta\theta}^j$. An eigenray is found each time a receiver point is located between two arrival points Z_θ^i and $Z_{\theta+\Delta\theta}^j$, for rays with the same trajectory class (same binary code). A bisection method is then used to find the exact starting angle between θ and $\theta+\Delta\theta$, providing an arrival depth of the ray equal to the receiver depth. The eigenray is defined by its starting angle θ_l, its propagation time t_l, its amplitude A_l, and its phase ϕ_l.

The HARORAY code can calculate the transfer function (complex field versus frequency) or impulse response between a source and a receiver. Elastic properties may be taken into account in the calculation of the transmission or reflection coefficient.

[1] N. R. Chapman and C. E. Lindsay, "Matched field inversion for geoacoustic model parameters in shallow water," IEEE J. Ocean Eng. **21**, 347–354 (1996).

[2] J-P. Hermand and P. Gerstoft, "Inversion of broadband multitone acoustic data from the Yellow-Shark experiments," IEEE J. Ocean Eng. **21**, 324–346 (1996).

[3] A. Tolstoy, "Using matched field processing to estimate shallow water properties from shot data," IEEE J. Ocean Eng. **21**, 471–479 (1996).

[4] D. F. Gingras and P. Gerstoft, "Inversion for geometric and geoacoustic parameters in shallow water: Experimental results," J. Acoust. Soc. Am. **97**, 3589–3598 (1995).

[5] 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).

[6] D. P. Knobles and R. A. Koch, "A time series analysis of sound propagation in a strongly multipath shallow water environment with an adiabatic normal mode approach," IEEE J. Ocean Eng. **21**, 1–13 (1996).

[7] A. Tolstoy, "Matched field tomographic inversion to determine environmental properties," Current Topics in Acoustics Research **1**, 53–61 (1994).

[8] A. Tolstoy, O. Diachok, and L. N. Frazer, "Acoustic tomography via matched field processing," J. Acoust. Soc. Am. **89**, 1119–1127 (1991).

[9] N. R. Chapman and L. Jaschke, "Matched field geoacoustic tomography experiments using light bulb sound sources in the HARO Strait sea trial," Proceedings of the Oceans'97 International IEEE Congress, Halifax, 763–768 (1997).

[10] L. Jaschke and N. R. Chapman, "Matched field inversion of broadband data using the freeze bath method," J. Acoust. Soc. Am. **106**, 1838–1851 (1999).

[11] J. G. Heard, M. McDonald, N. R. Chapman, and L. Jaschke, "Underwater light bulb implosions: A useful acoustic source," Proceedings of the Oceans'97 International IEEE Congress, Halifax, pp. 755–762 (1997).

[12] M. B. Porter, S. M. Jesus, Y. Stephan, E. Coelho, and X. Demoulin, "Single phone source tracking in a variable environment," in *Proceedings of the 4th European Conference on Underwater Acoustics*, edited by A. Alippi and G. B. Cannelli (CNR-IDaC, Rome, 1998), pp. 575–580.

[13] N. R. Chapman, and K. S. Ozard, "Matched field inversion for geoacoustic properties in young oceanic crust," in *Full Field Inversion Methods in Ocean and Seismo-Acoustics*, edited by O. Diachok, A. Caiti, P. Gerstoft, and H. Schmidt (Kluwer Academic, Dordrecht, 1995), pp. 165–170.

[14] J. A. Nelder and R. Mead, "A simplex method for function minimization," Comput. J. (UK) **7**, 308–313 (1965).

[15] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C* (Cambridge University Press, Cambridge, 1989).

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

Thin coating characterization by Rayleigh waves: An analytical model based on normal-mode theory

Filip W. Windels,[a] S. Vanaverbeke,[b] and O. Leroy
Interdisciplinary Research Center, KULAK, Etienne Sabbelaan 53, 8500 Kortrijk, Belgium

(Received 16 October 2000; revised 26 March 2001; accepted 12 June 2001)

The acousto-optic technique to evaluate coated substrates, presented by Devolder *et al.* [Appl. Phys. Lett. **68**, 1732 (1996)], uses the phase dependence of the reflected beam on the coating parameters. In this theoretical paper, a fully analytical model, based on normal-mode theory, is applied to this problem. It gives a transparent expression for the phase shift between the reflected and the incident field due to the presence of thin coatings (coating thickness much smaller than wavelength sound). The known numerical results from Fourier simulations, i.e. the phase dependency on the coating parameters and the advantages of using higher frequencies and wider beams, are reproduced. However, thanks to the analytical nature of the new theory, additional results are obtained. The calculation time is reduced by two orders of magnitude, which is important for the inverse problem. A technique for absolute thickness measurements is proposed. Measuring more to the right of the second reflected maximum increases the phase sensitivity. The versatility of the normal-mode approach is illustrated by applying it to other nondestructive testing (NDT) applications, such as hardness determination and adhesion testing, which are briefly discussed. [DOI: 10.1121/1.1391242]

PACS numbers: 43.35.Ns, 43.20.Ks, 43.35.Pt, 43.35.Zc, 43.20.Mv [SGK]

I. INTRODUCTION

Ultrasonic methods have found a wide range of application in the field of nondestructive testing (NDT).[1] For example, thin opaque coatings, which cannot be examined by conventional optical methods, constitute a fruitful topic of acoustical research.[2–6] Devolder *et al.*[2,3,7] generate a Rayleigh wave by insonifying with a bounded beam a semi-infinite substrate under its Rayleigh angle. When there is no coating present, a reflected wave is observed which consists of two parts:[8] the specularly reflected part and a second part, some what shifted in space and exponentially decaying in amplitude; see Fig. 1. Moreover, these two waves are 180 deg out of phase and are separated by a zone with no disturbance, the so-called null zone.

An acousto-optic technique[2,3] (see Fig. 2) was developed to detect the phase of the reflected field. A laser beam strikes the reflected acoustic field at normal incidence. Behind the sound field, the laser beam has split up—due to the acousto-optic effect—into different beams, each having a different direction, amplitude, and frequency. The intensity of their interference pattern behind the sound column is a copy of the sound field.[9] This intensity is recorded by a photodiode, and by a comparison with the reference signal, it provides the phase of the reflected field.

The presence of a coating changes this behavior drastically: the two reflected parts come closer and finally overlap (the null zone disappears) and the phase difference between them is no longer 180 deg. Attempts were made[2,3] to correlate the observed phase change with the parameters of the coating. These phase changes appear to be far more sensitive to the coating parameters than the more conventional amplitude information. In order to put these experimentally observed facts onto a more firm theoretical base, a Fourier beam model[8] was used to describe the incident bounded beam, in addition to the transfer matrix method[10]—modeling the coating—to simulate numerically the reflected beam profile and its corresponding phase difference. The numerically predicted phase shifts agreed quite well with the observed measurements.

However, the above theoretical model has some drawbacks:

(i) It is a fully numerical theory. Therefore, there is no transparent analytical formula which allows one to isolate or understand the influence of all the different coating parameters—like thickness, Young's modulus, Poisson ratio—on the phase shift. A numerical model is also far more calculation time-intensive than an analytical model. This slows down the inversion problem.

(ii) The lack of a transparent analytical formula limits the interpretation and possibilities of the acousto-optic technique: it measures only relative changes in thickness during a scan. Absolute thickness measurements of the film from a single point measurement are not possible. Also, the contributions to the phase shift due to a sudden change of the surface height are not taken into account.

(iii) It does not provide one underlying principle which explains the experimentally observed phenomenon, that not only coatings, but also defects like hardness, debondings, etc. give a similar phase change.

[a]Research Assistant of the Fund for Scientific Research—Flanders (Belgium); electronic mail: filip.windels@kulak.ac.be
[b]Research Assistant of the Flemish Institute for the encouragement of the scientific and technological research in industry (I. W. T.).

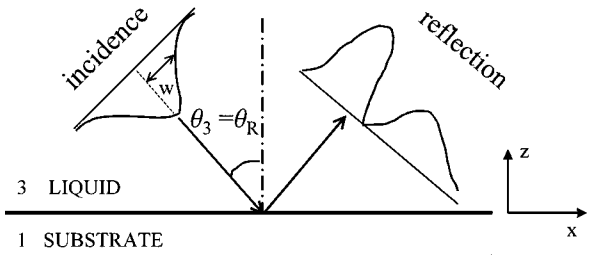


FIG. 1. A Gaussian beam, incidenting here under the Rayleigh angle ($\theta_3 = \theta_R$), is reflected by the substrate into a profile containing two lobes, separated by a null zone.

(iv) It lacks a profound physical insight into the mechanism why and how the reflected beam properties change with the coating.

To overcome these problems listed above, an *analytical* expression for the reflected field is derived in the second section, valid up to a first order, for thin coatings on a half space. Instead of the Fourier method,[2,3,8] the normal-mode theory[11–13] is used, not to be confused with the radiation mode theory,[14,15] which tackles similar problems in NDT. Unlike the radiation mode theory, which decomposes the Rayleigh wave into a continuous sum over radiation modes (a typical kind of combination of incident, reflected, and transmitted waves which constitute a radiation mode), the normal-mode theory considers the Rayleigh wave as a fundamental entity.[11,12] This simplifies the mathematics considerably, and for this reason the normal-mode theory is used in the present work.

The normal-mode method describes the interaction of the incident beam with the normal modes (also called eigenmodes or resonances) of the considered structure. An analytical expression is derived for the reflected field from a liquid-coating-substrate configuration in three steps in Sec. II.

First, in subsection II A, the considered structure is a liquid–solid (LS), which supports a Rayleigh wave as normal mode, as was shown by Jia.[11,12] An analytical formula is developed by means of normal-mode theory describing the reflected field for incidence at or near the Rayleigh angle.

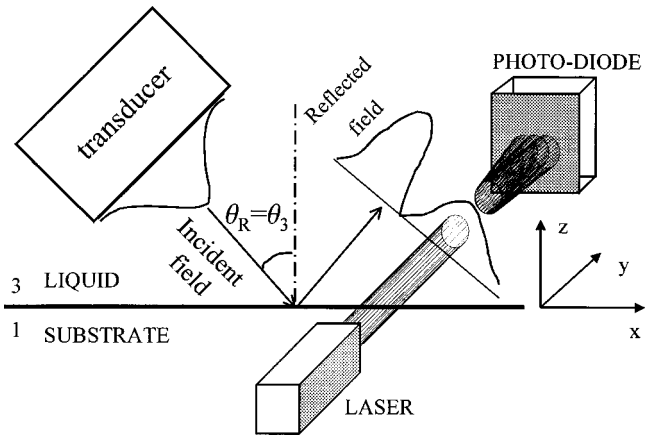


FIG. 2. The acoustic-optic technique to determine the phase in the reflected wave.

The reflected profile is shown to be highly sensitive to the Rayleigh resonance properties.

Second, in subsection II B, the liquid–coating–solid (LCS) structure is considered. Its normal mode still remains a Rayleigh wave but now with altered properties due to the small perturbing coat layer. The new characteristics of the perturbed Rayleigh mode are determined.

Finally, the subsection II C deals with the interaction of the incident bounded beam with the LCS structure, which has the perturbed Rayleigh wave as normal mode. The interaction with this altered Rayleigh mode results in a modified reflected field in comparison to the LS case. The obtained analytical expression shows the separate influence of all the different coating and beam parameters onto the reflected field.

In Sec. III simulations, based on the developed analytical formulas, show the dependence of the phase on the coating and beam properties. Some additional applications are also given.

Conclusions are presented in Sec. IV.

II. THEORY

A. The reflection of a bounded beam, incidenting at or near the Rayleigh angle, on a liquid–solid structure

The problem is considered to be two-dimensional as in Auld.[13] An incident Gaussian field u_i (half-width w), see Fig. 1, is considered with circular frequency ω, amplitude U_i, and x-projected half-width $w_0 = w/\cos\theta_3$, incidenting from medium 3: a liquid (L) onto medium 1: the substrate (S) under the angle θ_3, equal or close to the Rayleigh angle θ_R

$$u_i(0,x) = e^{-i\omega t} U_i \exp[-(x/w_0)^2 + ik_i x], \tag{1}$$

where $k_i = k_3 \sin\theta_3$ is the component of the incidenting wave number parallel to the substrate surface (the x component) and k_3 is the wave number in the liquid.

The incidenting field excites, according to normal-mode theory, a normal mode of the considered LS structure. A normal mode (or equivalently the terms eigenmode or resonance are used throughout this work) for a general layered structure is determined by its wave vector component k_x parallel to the interfaces; see Fig. 1. This k_x of the normal mode is defined as the zero of the denominator of the plane-wave reflection coefficient of the considered layered structure. Such a solution describes a field with reflection amplitude infinity, which means that the reflected wave can exist (amplitude nonzero) even with a zero incident wave. So, this solution represents a wave which can propagate without the presence of a driving incident force: it is an eigenmode, normal mode, or resonance of the structure.

According to this definition, the eigenmode of the considered LS structure is the Rayleigh wave of the liquid–substrate configuration.[11,12] The amplitude a_n of the n-th normal mode—in this case the Rayleigh wave—excited by the incident beam, is described by[11,12]

$$\frac{\partial a_n(x)}{\partial x} + (-ik_n + \alpha_n) a_n = -\frac{\omega^2 Z_3 \cos\theta_3 u_{nz}^*(0) u_i(x,0)}{2P_n}, \tag{2}$$

where k_n and α_n are the k_x component and the leak rate of the eigenmode, $Z_3=\rho_3 v_{3d}/\cos\theta_3$ is the acoustical impedance of the liquid, $u^*_{nz}(0)$ is the complex conjugate of the amplitude of the z-component of the eigenmode at the surface, and P_n is the average modal power flow in the x direction, per unit width in the y direction.

Integrating Eq. (2) from $-\infty$ to x, and substituting Eq. (1) in it, yields

$$a_n(x)=e^{+ik_n x-\alpha_n x}(-\beta_{ni})U_i \times\int_{-\infty}^{x} dx'\, e^{-(x'/w_0)^2+\alpha_n x'} e^{+i(k_i-k_n)x'}, \tag{3}$$

with

$$\beta_{ni}=\frac{\omega^2 Z_3\cos\theta_3 u^*_{nz}(0)}{2P_n}. \tag{4}$$

Equation (3) is a very important physical expression as it shows that the growing of the Rayleigh wave amplitude a_n along the excitation area is due to the overlap of the incidenting Gaussian profile with the exponentially decaying Rayleigh field, but there is a *phase mismatch* factor k_i-k_n preventing this energy transfer.

The boundary conditions, expressing the continuity of the stresses and normal displacements over the boundary, relate the incident, reflected, and Rayleigh wave irradiation field to each other. By the appropriate substitutions[11,12] a final expression for the reflected field is obtained

$$u_r(x,0)=u_i(x,0)\{1-\sqrt{\pi}h\,\exp[\gamma_n^2]\mathrm{erfc}[\gamma_n]\}, \tag{5}$$

with

$$\gamma_n=-x/w_0+(h+ip)/2, \tag{6a}$$

$$h=\alpha_n w_0, \tag{6b}$$

$$p=(k_i-k_n)w_0, \tag{6c}$$

where erfc is the complementary error function.[16] So, it is seen that the reflected field is totally characterized by three parameters: w_0 the projected half width, h the parameter introduced by Bertoni and Tamir,[8] and p a dimensionless phase mismatch factor.

The parameter h determines the two-lobe structure of the reflected field.[8] If $h\simeq 0.5$, $p=0$, the reflected field consists of two reflected lobes separated by a null zone without acoustical disturbance. If h is kept constant, but p starts to differ from zero (which means that the incidence angle is no longer equal to the Rayleigh angle) the null zone begins to disappear. Also, the phase difference between the two zones is no longer 180 deg. However, the same phenomenon was also seen in a totally different experiment:[2,3] keeping the angle fixed but changing the coating thickness produced the same effects.

These results are quite remarkable: insonifying the liquid–substrate (LS) at an angle close to, but *different* from the Rayleigh angle of the substrate, gives the same kind of reflected field as exciting a liquid–*coating*–substrate (LCS) structure under *precisely* the Rayleigh angle of the substrate. This symmetry is depicted in Fig. 3. So, changing the incidence angle for the LS structure gives the same effects as keeping the angle constant but increasing the thickness for an LCS structure! The underlying mechanism of this intriguing symmetry will be revealed in the next section.

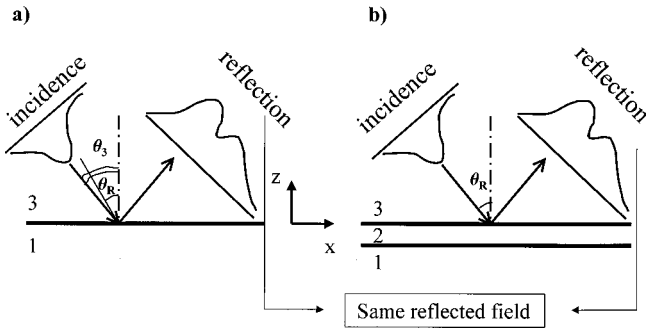


FIG. 3. In part (a) of the figure, the reflection on the LS structure for an incidence angle different from the Rayleigh angle is shown. The reflected field has no real null zone anymore (compare to Fig. 1). Part (b), on the contrary, shows the reflection for incidence under the Rayleigh angle, but this time for the LCS structure. The same kind of reflected field is observed.

B. The perturbed Rayleigh mode for the liquid–coating–solid structure

The reason for the striking similarity discussed in the previous subsection and visualized in Fig. 3, is to be found in the meaning of the variable p. What does it mean? The phase mismatch factor p is directly proportional to the difference between the mode wave vector and the incident wave vector. This difference can be influenced in two ways: first and most trivially, by changing the incident wave vector k_i [see Fig. 3(a)]. This occurs by altering the incidence angle. Alternatively, however, p can be controlled by changing the modal wave vector k_n. And this is just what happens when the angle is kept constant but the coating thickness is changed: the presence of a coating on the substrate alters the Rayleigh resonance conditions and as a consequence the k_n's are shifted [see Fig. 3(b)].

The LCS structure, as considered from now on, still supports a Rayleigh wave as normal mode. But, the properties of this Rayleigh resonance have changed in comparison to the LS case. The task is to derive the modified resonance conditions for the Rayleigh wave when there is a small coating layer present. This is accomplished by a modification of the reflection coefficient for the structure liquid–coating–substrate (LCS). To this end an approximation of the transfer matrix,[10] correct up to the first order in the coating thickness d, was used. This expression can be found in the context of adhesion models.[17] The Appendix provides the necessary expressions. The new reflection coefficient will result into a different resonance angle than we had for the LS structure. The new resonance k_x, the wave number component parallel to the surface, is determined up to a first order in d.

The LCS configuration with all the corresponding symbols is shown in Fig. 4: ρ_i, v_{id}, and v_{is} are the respective density, dilatational, and shear velocity of medium i, k_i and κ_i are the respective dilatational and shear wave number of medium i; θ_i and γ_i are the respective angles of the dilatational and shear waves in medium i.

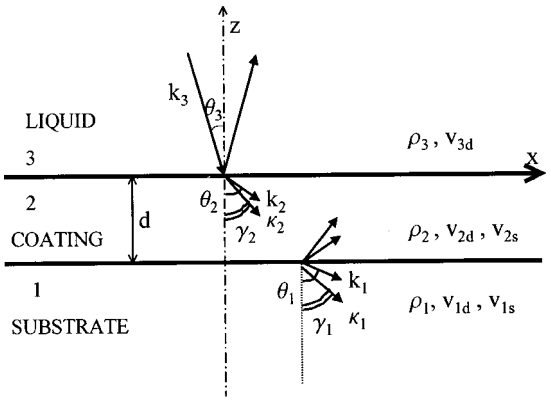


FIG. 4. The liquid–coating–substrate structure and all its defined symbols.

To find the eigenmodes of the LCS structure one needs the expression (A1) for the LCS-reflection coefficient. Setting the denominator equal to zero provides the dispersion equation and its solution determines the eigenmode of the LCS structure, i.e., the perturbed Rayleigh wave. The denominator N of the reflection coefficient in Eq. (A1) is

$$N(k_x)=Z_{in}+Z_3. \tag{7}$$

This can be rewritten formally as a power series, up to the first order, in the small parameter d, the coating thickness, and this denominator should be zero

$$N=N^{(o)}(k_x)+d\,N^{(1)}(k_x)=0, \tag{8}$$

where $N^{(o)}$ is the expression when the coating is absent, and the second term $N^{(1)}$ is the first-order correction due to the coating. This equation can be solved analytically by means of a perturbation technique. Suppose $\hat{k}_x$ is the total solution of Eq. (8), i.e., the x component of the Rayleigh wave number of the LCS configuration. As a consequence it can, according to perturbation theory, be rewritten as

$$\hat{k}_x=\hat{k}_x^{(0)}+d\hat{k}_x^{(1)}. \tag{9}$$

Substituting Eq. (9) into Eq. (8), using Taylor expansions around $\hat{k}_x^{(0)}$ and keeping terms up to the first order in d, results in two equations

$$N^{(0)}[\hat{k}_x^{(0)}]=0, \tag{10a}$$

$$\hat{k}_x^{(1)}=-\frac{N^{(1)}[\hat{k}_x^{(0)}]}{\dfrac{\partial N^{(0)}[k_x^{(0)}]}{\partial k_x}}. \tag{10b}$$

The zero-order part of the solution, $\hat{k}_x^{(0)}$, is given by the Rayleigh solution for the LS structure [Eq. (10a)]. The first-order part is given by d times the expression in Eq. (10b).

Now, the formal solution of Eq. (10b) is made more explicit by rewriting N, this time given by Eq. (7) and the further equations in the Appendix, into the form of Eq. (8). After some cumbersome but straightforward calculations, the following result is obtained:

$$N^{(1)}=-i\rho_2\omega\{1-[\rho_3c_3/(\rho_2c_2\cos\theta_3)]^2+F_2G\}. \tag{11}$$

The function F_2 is given by Eq. (A6), and the function G, independent of the coating, is

$$G=(v_{1d}/v_{1s})\frac{\cos(2\gamma_1)\sin(2\gamma_1-\theta_1)\sin\gamma_1}{\cos^2\theta_1}\times\left[1-(v_{1s}/v_{1d})\frac{2\cos\theta_1\cos\gamma_1}{\cos(2\gamma_1)}\right]. \tag{12}$$

The goniometric functions in Eqs. (11) and (12) can be expressed in terms of the wave number variables k_x, as shown in the Appendix. The physical meaning of θ_i, γ_i as refraction angles is shown in Fig. 4. Also, the derivative of $N^{(0)}$ in Eq. (10b) can be rewritten as

$$\frac{\partial N^{(0)}}{\partial k_x}=\frac{\omega\rho_1}{\kappa_1^4\sqrt{k_1^2-(k_x^{(0)})^2}}\left[\frac{\partial N_{\mathrm{BT}}}{\partial k_x}\right](\hat{k}_x^{(0)}), \tag{13}$$

with N_{BT} the denominator as expressed by Bertoni and Tamir:[8]

$$N_{\mathrm{BT}}(k_x)=(2k_x^2-\kappa_1^2)^2-4k_x^2\sqrt{k_x^2-\kappa_1^2}\sqrt{k_x^2-k_1^2}+i(\kappa_1^4\rho_3/\rho_1)\frac{\sqrt{k_x^2-k_1^2}}{\sqrt{k_3^2-k_x^2}}. \tag{14}$$

This change in the k_x component for the LCS configuration, $\Delta\hat{k}_x$, is found from Eq. (9)

$$\Delta\hat{k}_x=d\hat{k}_x^{(1)}, \tag{15}$$

and with the aid of Eqs. (11) and (14), Eq. (15) becomes

$$\Delta k_x=\frac{i\rho_2\omega d\{1-[\rho_3c_3/(\rho_2c_2\cos\theta_3)]^2+[1-4(k_x^{(0)}/\kappa_2)^2(1-(v_{2s}/v_{2d})^2]G\}}{[\omega\rho_1/(\kappa_1^4 i\sqrt{-k_1^2+(k_x^{(0)})^2})][\partial N_{\mathrm{BT}}/\partial k_x](\hat{k}_x^{(0)})}. \tag{16}$$

Equation (16) is rewritten into an alternative more transparent expression. As the velocities v_{2s}, v_{2d} are not independent, they are removed by use of the independent Lamé constants $\lambda_2+2\mu_2=\rho_2v_{2d}^2$ and $\mu_2=\rho_2v_{2s}^2$. To extract the exact frequency dependence of Δk_x, Eq. (16) is written as a function of the slowness $s_x=k_x/\omega$, $s_{id}=k_{id}/\omega$ $(i\in\{1,2,3\})$, $s_{is}=\kappa_i/\omega$ $(i\in\{1,2\})$. The following expression is obtained:

$$\Delta k_x = \frac{-\omega^2 d\sqrt{-s_{1d}^2+(s_x^{(0)})^2}\left\{(1+G)\rho_2-[\rho_3 c_3/(\cos\theta_3)]^2\frac{1}{\lambda_2+2\mu_2}-4(s_x^{(0)})^2 G\frac{\mu_2(\lambda_2+\mu_2)}{\lambda_2+2\mu_2}\right\}}{\rho_1 v_{1s}^4\left[\frac{\partial \tilde{N}_{\mathrm{BT}}}{\partial s_x}\right](\hat{s}_x^{(0)})}, \tag{17}$$

with

$$\tilde{N}_{\mathrm{BT}}(s_x)=\left\{(2s_x^2-s_{1s}^2)^2-4s_x^2\sqrt{s_x^2-s_{1s}^2}\sqrt{s_x^2-s_{1s}^2}+i(s_{1s}^4\rho_3/\rho_1)\frac{\sqrt{s_x^2-s_{1d}^2}}{\sqrt{s_{3d}^2-s_x^2}}\right\}. \tag{18}$$

Equation (16) or (17) shows that the presence of the coating, expressed by the proportionality to d, shifts the initial Rayleigh resonance value k_x to the new one, $k_x+\Delta k_x$. As a result the phase mismatching factor p of Eq. (6c) changes even when the incidence angle is kept constant (and equal to the Rayleigh angle of the LS configuration). As a consequence the same effects are observed as if there was no coating present and only the angle was varied.

Once this remarkable symmetry is understood, it can be fully exploited for coating characterization in the next section.

C. The reflected beam from the liquid-coating-solid structure

Equation (5), together with expressions (6b) and (6c), shows that the reflected field for the LS configuration depends strongly on the properties of the Rayleigh wave. As the presence of a coating alters these Rayleigh wave parameters, Eq. (17), the resulting reflected field for the LCS configuration will also change.

The reflected beam for the LCS structure is given by Eq. (5), but with the phase mismatch factor p and the h factor modified. Out of the definitions for p and h, respectively, Eqs. (6c) and (6b), and the correction to the resonance, Eq. (16) or (17), one readily finds

$$p=p^{(0)}-\mathrm{Re}[\Delta k_x]w_0, \tag{19a}$$

$$h=h^{(0)}+\mathrm{Im}[\Delta k_x]w_0, \tag{19b}$$

where $p^{(0)}$ and $h^{(0)}$ are the values for the LS structure. Assuming incidence under the Rayleigh angle of the substrate, one can put $p^{(0)}=0$. The reflected field, in terms of the new h and p of Eqs. (19a) and (19b), can be expressed as

$$u_r(x,0)=u_i(x,0)\{1-\sqrt{\pi}h\exp[(-x/w_0+(h+ip)/2)^2]\times\mathrm{erfc}[-x/w_0+(h+ip)/2]\}. \tag{20}$$

III. SIMULATIONS

Before performing some simulations, it is very useful to obtain a first-order expression for the phase of the reflected beam as predicted by normal-mode theory.

The reflected field was given by Eq. (20) and can be written in the form

$$u_r=u_i\{1-\sqrt{\pi}(h_0+\Delta h)\exp[\gamma_n^2]\mathrm{erfc}(\gamma_n)\}, \tag{21}$$

with

$$\gamma_n=\gamma_{n0}+\Delta\gamma, \tag{22}$$

$$\gamma_{n0}=-x/w_0+h_0/2, \tag{23}$$

$$\Delta\gamma=(\Delta h+i\Delta p)/2, \tag{24}$$

in Eq. (24) is $\Delta h=h-h^{(0)}$ and $\Delta p=p-p^{(0)}$. The phase $\tilde{\phi}$ of the reflected signal $u_r(x)$, see the above Eq. (21), is always defined relative to the reference $u_i(x)$

$$\tilde{\phi}=\mathrm{Arg}\left[\frac{u_r(x)}{u_i(x)}\right]. \tag{25}$$

A first-order approximation of the phase can be derived. By making a Taylor expansion, correct up to the first order in $\Delta\gamma$, of the exp and the erfc,[16] one finds

$$u_r=u_i\{1-\sqrt{\pi}h_0\exp[\gamma_{n0}^2]\mathrm{erfc}(\gamma_{n0})+\sqrt{\pi}\exp[\gamma_{n0}^2]\times[2(h_0/\sqrt{\pi})\exp[-\gamma_{n0}^2]\Delta\gamma-2h_0\gamma_{n0}\Delta\gamma\,\mathrm{erfc}(\gamma_{n0})-\Delta h\,\mathrm{erfc}(\gamma_{n0})]\}. \tag{26}$$

The first-order approximation to the phase, ϕ, can then be derived out of Eqs. (26) and (25)

$$\phi=\mathrm{Arg}[1-\sqrt{\pi}h_0\exp[\gamma_{n0}^2]\mathrm{erfc}(\gamma_{n0})]+h_0\frac{1-\sqrt{\pi}\gamma_{n0}\exp[\gamma_{n0}^2]\mathrm{erfc}(\gamma_{n0})}{1-\sqrt{\pi}h_0\exp[\gamma_{n0}^2]\mathrm{erfc}(\gamma_{n0})}\Delta p, \tag{27}$$

where Δp is defined by Eqs. (19a) and (17).

Armed with the expression for the phase predicted by normal-mode theory, Eq. (25), or its first-order approximation, Eq. (27), a profound insight into the results of the next subsections is available.

As the analytical solution is based on some approximations, the full Fourier theory has also been implemented. The Fourier model makes no simplifications on the reflection coefficient and the transfer matrix. This was also done by Devolder *et al.*[2,3] With this model a numerical alternative theory is at our disposal to check the results of the analytical but approximated normal-mode expression.

A. Influence of the material parameters

In this section all the plots are done with the following beam parameters: $f=4$ MHz, $w=4$ mm, and the measurement position for the phase is located at the second maximum of the reflected profile; see Fig. 1. Unless otherwise mentioned, a water–copper–steel configuration is considered. The corresponding values for the material parameters are listed in Table I. The incidence angle is the Rayleigh angle for the water–steel configuration (30.9°).

TABLE I. The material parameters of water (medium 3), copper (medium 2), and steel (medium 1) used in the simulations.

v_{3d}	1480 m/s
ρ_3	1000 kg/m^3
v_{2d}	4660 m/s
v_{2s}	2260 m/s
ρ_2	8900 kg/m^3
v_{1d}	5790 m/s
v_{1s}	3100 m/s
ρ_1	7900 kg/m^3

1. Thickness

First, the influence of the Cu film thickness on the phase is studied. Figure 5(a) displays this behavior. The phase grows with growing coating thickness. For thicker coatings a kind of saturation effect appears. So, the technique has a high sensitivity to thin coatings. The range 0–40 μm is shown into more detail in Fig. 5(b). The phase happens to grow initially linearly proportional to the thickness. This was predicted by Eq. (27) and Eq. (17). Starting from 20 μm the analytical model starts to overestimate the phase shift. An idea of the discrepancies in Fig. 5(b) is given in Table II. Up to coatings of 20 μm the normal-mode theory agrees very well with the Fourier model. The errors at higher thicknesses are due to the fact that the approximations of the analytical model break down when $k_x d \sim fd$ is too high.

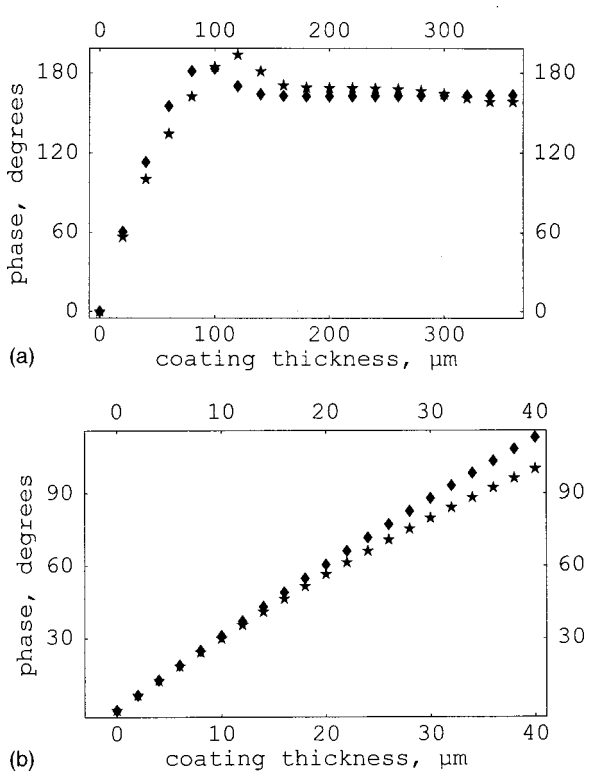


FIG. 5. (a) The phase$=\phi(d_2)-\phi(0)$ of the reflected profile, measured in the second maximum at 4 MHz, as a function of the thickness. The diamonds give the prediction by the analytical model; the stars are the values for the numerical Fourier model. Steps of 20 μm from 0–360 μm. (b) The phase$=\phi(d_2)-\phi(0)$ of the reflected profile, measured in the second maximum, as a function of the Cu-coating thickness. The diamonds give the prediction by the analytical model; the stars are the values for the numerical Fourier model. Steps of 2 μm from 0–40 μm.

TABLE II. The discrepancy in the phase shift calculated by the analytical model and by Fourier theory for different coating thicknesses.

Thickness (μm)	Analytical (deg)	Fourier (deg)	Error (deg)
0	0.0	0.0	0.0
5	15.7	15.3	0.4
10	31.0	29.8	1.2
15	45.9	43.6	2.3
20	60.4	56.4	4.0
25	74.4	68.5	5.9
30	87.9	79.7	8.2
35	100.9	90.3	10.6
40	113.2	100.2	13.0

An important point can be made on the calculation time for the two models. In Fig. 5(a) the calculation time for the Fourier results amounts to 180 s, whereas the calculation time for the normal mode results of Eq. 27 amounts to 1.8 s. The simulations were both done on the same computer with MATHEMATICA 4.0. The formulas provided by normal-mode theory are, thanks to their analytical nature, very computationally efficient. This ease of computation is important for the inverse problem: determining the coating thickness out of the phase shift.

2. The density and the elastical constants

The dependence of the phase shift on the Lamé constants of the layer is depicted in Fig. 6. It is clearly seen that the phase is far more sensitive to changes in μ than to variations in λ(lambda). If the density is altered, the behavior becomes that of Fig. 7. A constant shift proportional to the density change is observed. This is understood from Eqs. (27) and (17). Changing the density does not affect the terms in the elastical constants and as a consequence the plane of Fig. 6 is just shifted upwards into that of Fig. 7.

B. Influence of the beam parameters

By adjusting the parameters of the incident beam, it is shown in this section that the results of the previous section can be highly optimized. All calculations are done for the water–copper–steel configuration.

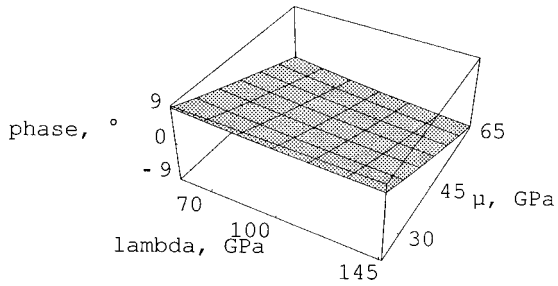


FIG. 6. The phase of the reflected profile, $\phi(\mu,\lambda)-\phi(\mu_{Cu},\lambda_{Cu})$, for a coating of 5 μm, measured in the second maximum at 4 MHz, as a function of the Lame constants λ(lambda) and μ. $\rho=8900$ kg/m^3, steps of 0.15 μ_{Cu} and 0.15 λ_{Cu}. Bounds: (1 ± 0.45) μ_{Cu},λ_{Cu}. The maximum error is 5 percent with the Fourier model.

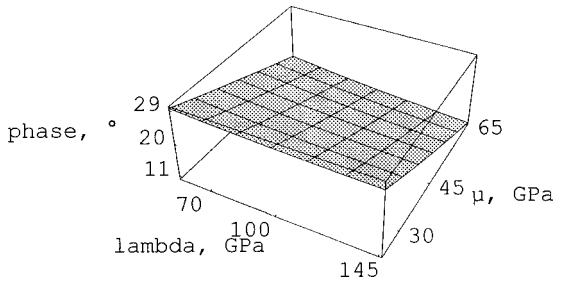


FIG. 7. The phase of the reflected profile, $\phi\ (\mu,\lambda,\rho=13\,900\ \mathrm{kg/m^3})-\phi\ (\mu_{\mathrm{Cu}},\lambda_{\mathrm{Cu}},\rho_{\mathrm{Cu}}=8900\ \mathrm{kg/m^3})$, for a 5-$\mu$m film, measured in the second maximum at 4 MHz, as a function of the Lamé constants λ(lambda) and μ. $\rho=13\,900\ \mathrm{kg/m^3}$, steps of $0.15\,\mu_{\mathrm{Cu}}$ and $0.15\,\lambda_{\mathrm{Cu}}$. Bounds: $(1\pm0.45)\ \mu_{\mathrm{Cu}},\lambda_{\mathrm{Cu}}$.

1. The frequency

The influence of the frequency on the phase of the reflected profile is quite drastic. This can be seen in Fig. 8. The general tendency is: higher frequencies cause higher phase shifts. For frequencies up to 6 MHz, the two normal-mode expressions for the phase, Eqs. (25) and (27), coincide nicely with the phase as defined by Fourier theory. This means, cf. Eqs. (27) and (17), that the phase depends quadratically on the frequency. For higher frequencies, there is still a good agreement between the general analytical expression for the phase, Eq. (25) and the numerical Fourier one, although the first-order approximation of Eq. (27) breaks down. Starting from 20 MHz a saturation is observed. So, there is an upper bound to the frequency in order to achieve a maximal phase shift. The difference between the analytical and the Fourier model is given in more detail in Fig. 9. The quadratic curvature is nicely observed and it is also clear that the analytical model slightly overestimates the Fourier theory. Some idea of the error between the two theories: at 4 MHz the phase is 11.2° with an error of 1.3°, at 8 MHz the phase is 47.8° with a corresponding error of 2.6°.

2. The measurement position in the reflected profile

At what position in the reflected profile should one measure the phase shift? That question is answered in Fig. 10. When there is no coating present, one observes in the second part of the reflected profile a phase shift of −180°. A coating of 10 μm alters the phase significantly, as is seen in the dashed curve. Moreover, it is seen that the further one measures along the *x* axis, the higher is the phase shift. This is an important point as Devolder *et al.*[2,3] did their measurements in the second maximum (max) of the reflected field. As can be concluded here this is not at all necessary, and one can measure better even further away. It seems that the Rayleigh wave, measured at a further located position, has propagated over a longer distance through the coating and hence will suffer from a higher phase shift. To optimize sensitivity, it is good to keep in mind that by measuring two half-widths further in the reflected beam, one can gain a phase shift double as high!

However, the phase sensitivity cannot be raised continuously, just by arbitrary measuring far away in the second lobe. As the second lobe drops exponentially, one should measure at a point in the secondary lobe where the signal-to-noise ratio is still high enough. This ratio depends not only on the present experimental noise level, but also on the leak rate of the Rayleigh resonance as slowly decaying Rayleigh modes will generate an extended second lobe. So, liquid–solid configurations supporting Rayleigh waves, which reradiate over long distances, benefit the most from measuring further away in the second lobe.

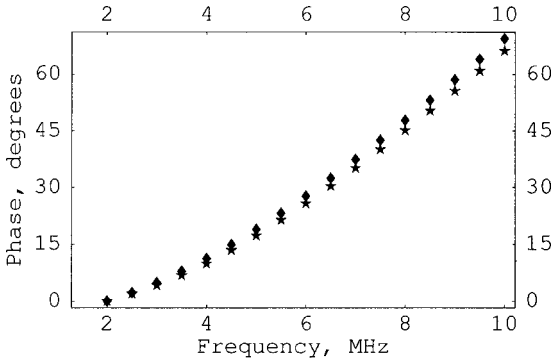


FIG. 9. The phase$=\phi(f)-\phi(2\ \mathrm{MHz})$ of the reflected profile, measured in the second maximum of a 4-MHz beam, as a function of the frequency for a coating of 5 μm. The diamonds give the prediction by the analytical model; the stars are the values for the numerical Fourier model. Steps of 0.5 MHz from 2–10 MHz.

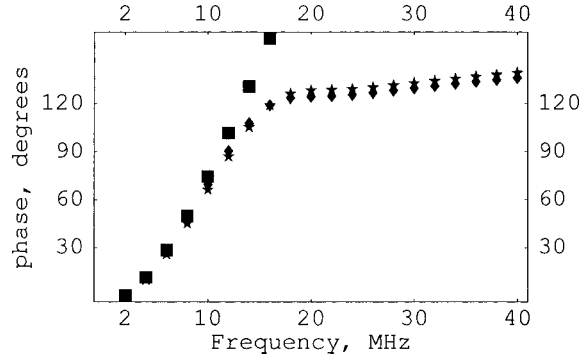


FIG. 8. The phase$=\phi(f)-\phi(2\ \mathrm{MHz})$ of the reflected profile, measured in the second maximum at 4 MHz, as a function of the frequency for a coating of 5 μm. The diamonds give the prediction by the analytical model [Eq. (25)]; the stars are the values for the numerical Fourier model, and the squares are the values for the first-order approximation of the analytical phase [Eq. (27)]. Steps of 2 MHz from 2–40 MHz.

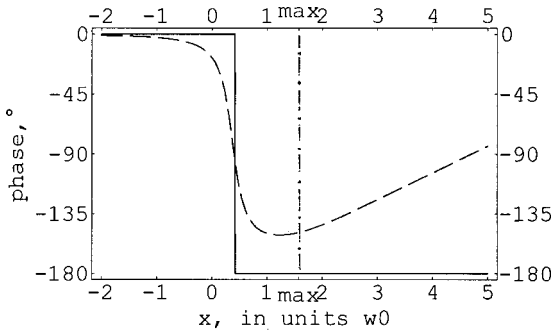


FIG. 10. The phase $\mathrm{Arg}[u_R(x)/u_i(x)]$ of the reflected profile, as a function of the position. All data are obtained with the analytical formulas. The full line gives the phase behavior when there is no coating present; the dashed curve gives the phase when there is a coating of 10 μm present. The max tick denotes the position in the reflected profile where the amplitude is maximal.

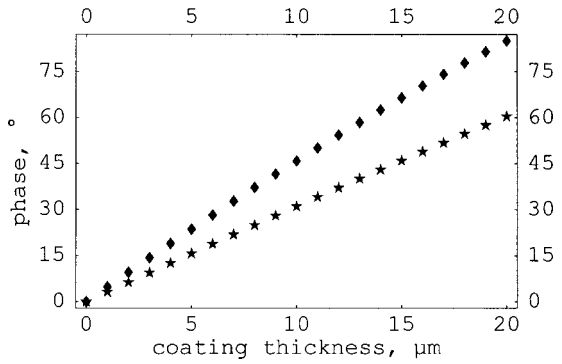


FIG. 11. The phase $=\phi(d_2)-\phi(0)$ of the reflected profile, measured in the second maximum (for that beamwidth) at 4 MHz, as a function of the coating thickness. All data are obtained with the analytical formulas. The diamonds give the values for a beamwidth $w=8$ mm; the stars are the values for $w=4$ mm, $f=4$ MHz, water–Cu–steel configuration. Steps of 1 μm from 0–20 μm.

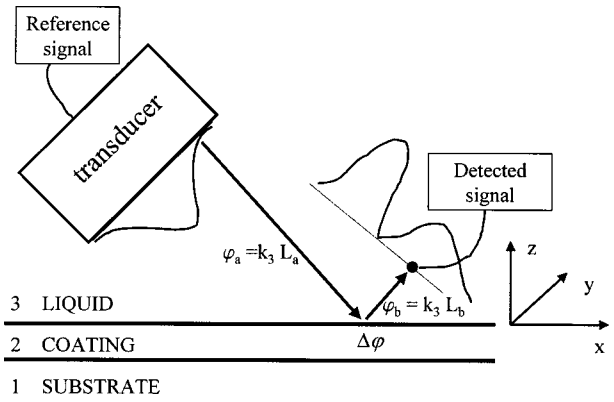


FIG. 12. Analysis of the phase shift between the incident and detected signal. ϕ_a and ϕ_b are the phase shifts along the respective propagation paths source-interface (with length L_a) and interface-detection point (with length L_b). An additional phase shift $\Delta\phi$ occurs due to the reflection at the water–coating interface.

3. The beamwidth

In Fig. 11 one sees the dependence of the phase on the thickness for two beamwidths: 8- and 4-mm as half-width. The phase was measured at the position of the second maximum for that beamwidth. The phase shift happens to be higher for wider beams. The mechanism behind this observation is related to the phase which grows with the measurement position along the x axis (see Fig. 10). As the second maximum of the wider beam is located further away along the x axis than the secondary maximum of the more narrow beam, it will exhibit a higher phase shift.

C. Other applications

1. Absolute thickness measurements

The acousto-optic technique[2,3] only measures changes in thickness. In other words: it is impossible to measure at a single point and to determine the thickness of the coating over there. Only the changes of the thickness during a scan from point to point are detected.

However, the *analytical* formula from Eqs. (27) and (17) gives a hint towards the solution of this problem. It is important to understand how the phase difference between the reference signal and the detected signal occurs. This can be seen from Fig. 12. The phase ϕ measured at the detection point is a sum of two contributions: one from the phase accumulation along the propagation paths source-interface (with length L_a) and interface-detection point (with length L_b) which amounts to $k_3(L_a+L_b)$, and another shift due to the reflection at the coat layer as defined in Eqs. (27). The total phase is

$$\phi(\omega)=\frac{\omega}{v_{3d}}(L_a+L_b)+A(\omega)\Delta p(\omega)+B(\omega), \tag{28}$$

with

$$A(\omega)=\frac{h_0[1-\sqrt{\pi}\gamma_{n0}\exp[\gamma_{n0}^2]\mathrm{erfc}[\gamma_{n0}]]}{1-\sqrt{\pi}h_0\exp[\gamma_{n0}^2]\mathrm{erfc}[\gamma_{n0}]}, \tag{29}$$

and

$$B(\omega)=\mathrm{Arg}[1-\sqrt{\pi}h_0\exp[\gamma_{n0}^2]\mathrm{erfc}(\gamma_{n0})]. \tag{30}$$

An absolute thickness determination can be performed by measuring the phase ϕ_1 at a first frequency ω_1 and then doing the same at a second frequency $\omega_2=b\omega_1$, which gives another phase ϕ_2. With the aid of the analytical formula (28), one can construct the following linear combination of the measured phases:

$$\phi_2-b\phi_1=A(\omega_2)\Delta p(\omega_2)-bA(\omega_1)\Delta p(\omega_1)+[B(\omega_2)-bB(\omega_1)], \tag{31}$$

which is independent of the propagation path and contains only the coating information. With the definition $\Delta p'=\Delta p/d$, one finds readily the absolute thickness d

$$d=\frac{\phi_2-b\phi_1-[B(\omega_2)-bB(\omega_1)]}{A(\omega_2)\Delta p'(\omega_2)-bA(\omega_1)\Delta p'(\omega_1)}. \tag{32}$$

As a control to the accuracy of this formula, the following experiment was simulated: the phase shift was calculated with the Fourier theory for a 5-μm coating and a half beamwidth of 4 mm. This was done once for a frequency of 3.5 MHz and once for a frequency of 4.5 MHz ($b=1.29$ in this case). With the aid of the analytical formula (32), a coating thickness of 5.16 μm was found from a single point measurement simulation.

The theoretical formulas, as developed by Devolder *et al.*,[2,3] did not take into account a small complication. In the experiments with the acousto-optic technique[2,3] only phase *changes* and the corresponding thickness *changes* during a scan are measured. Therefore, samples were used which had a zone without a coating, serving as a reference, followed by a coated region. The phase shift when coming from the uncoated zone onto the coated region provided the thickness of the coating. However, for such samples, will the coated zone be somewhat closer to the sound source and the detection zone? (see Fig. 13). This means that not only the material parameters change but also the propagation path source–receiver. This will result in an additional phase change due to the shorter propagation path, as can be seen in Eq. (28), which was not taken into account previously.

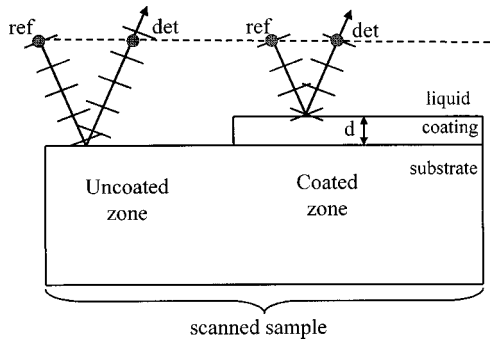


FIG. 13. Phase shift between the detection point (det) and the reference point (ref), due to the different propagation path, when the phase is measured either above the uncoated region or above the coated zone.

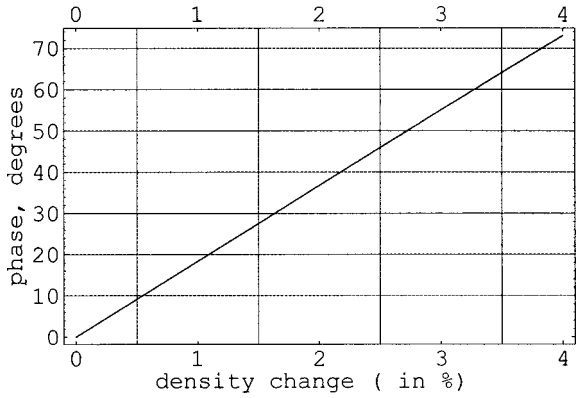


FIG. 14. The phase of the reflected profile, measured in the secondary maximum at 4 MHz, as a function of the density changes in the substrate.

To calculate this phase shift, it is easiest to consider the case of a beam reflecting from the uncoated region and of a beam reflecting from the coated zone, which lies a distance d closer to the detection point than the uncoated area, as seen in Fig. 13. An elementary calculation shows

$$\Delta\phi = 2dk_3 \cos\theta_3 = 2d(\omega/v_3)\cos\theta_3 . \tag{33}$$

This expression (33) is only valid for insonification and detection either above the uncoated region or above the coated zone. This is because mode conversion at the up-step has been neglected. A coating of thickness $d=5\ \mu$m will produce a phase shift of 15° [Eq. (27)], but due to the changed path length, there is an additional 8° shift [Eq. (33)], which is a correction of 50 percent. So, the total shift in that case will not be the theoretical 15°, given in Refs. 2, 3, but the 23° which includes the effect of the raised upper surface.

Equation (33) has two implications: first, it quantifies a possible source of noise for the phase determination: unwanted vibrations can move up and down the surface (which means a change in the propagation path L_a+L_b), which disturbs the phase stability. Second, Eq. (33) shows that an increase in height of the upper coating interface gives a positive contribution to the phase. But, the presence of the coating also gives a positive change in the phase. So, when one puts a coating on a substrate the two effects on the phase, one due to the coating presence and the other due to the increased height, will add up constructively and give a higher resolution for such coatings.

2. Other NDT applications

The formalism of normal-mode theory, as expressed by Eqs. (19a), (19b), and (20), predicts that not only coatings, but every defect which alters the resonance conditions Δk_x of the considered normal mode, can be detected by this phase-sensitive technique.

For example, the hardness of a solid also can be studied by this method. As some hardness variations in some steel types are mainly due to changes in density of the material,[7] it is interesting to study the sensitivity of the reflected phase to the density of the solid substrate. The considered configuration is water–steel, having the Rayleigh wave as normal mode. The shift in Rayleigh resonance position Δk_x due to different density values ρ_1 of the solid, was calculated from Eq. (14). The corresponding phase was determined by formulas (19b), (19a), (20), and (25).

The solid density, starting at 7900 kg/m^3, was gradually raised in steps of 0.5 percent up to a total amount of 4 percent. The corresponding simulated phase shifts are shown in Fig. 14. It shows that this Rayleigh phase technique is also very sensitive to hardness variations.

A brief remark about another application: adhesion testing. Coat layers, which are not rigidly bonded to the substrate, can also be modeled by the normal-mode technique. The weakened adhesion of the coating to the substrate, is introduced[17] by imposing a discontinuity on the stresses and displacements at the interface film–substrate. This shift of the stresses and displacements when crossing the interface film–substrate, is mathematically modeled by an adhesion transfer matrix[17] which relates the stresses and displacements at the bottom of the film to the stresses and displacements at the upper face of the substrate. Hence, the reflection coefficient for an imperfectly bonded film to a substrate, is found by multiplying the matrix describing the coating, Eq. (A5), with this adhesion transfer matrix.[17] From the denominator of the obtained reflection coefficient, the resonance position as a function of the adhesion quality can be determined. When there is a change in the adhesion status, the resonance position shifts and due to Eq. (27) a different phase is detected.

IV. CONCLUSIONS

The interaction of a bounded Gaussian beam with a thin coating attached to a substrate has been studied before,[2] both experimentally and theoretically by a Fourier method. In the present article the phenomenon has been studied theoretically by a normal-mode theory, which establishes for the first time a transparent analytical formula for the reflected field and its phase. The known Fourier results, i.e., that the phase is extremely sensitive to the coating parameters (thickness, density, and elastical constants), and that the phase sensitivity can be enhanced by using wider beams and higher frequencies, are recovered. Thanks to the analytical nature of the normal-mode formulas, some new results are found.

(i) Measuring more to the right of the second reflected maximum improves the phase sensitivity.
(ii) Absolute thickness determination from a single point measurement is possible.
(iii) The calculation time for the phase shift is reduced by two orders of magnitude, which is important for the inverse problem.
(iv) The physical mechanism behind the phase sensitivity is understood as a change in the resonance conditions of the Rayleigh mode due to the coating perturbation.
(v) Thanks to this interpretation provided by normal-mode theory, every perturbation which alters the resonance properties of the considered normal mode can be detected by the resulting phase shift. Two examples, i.e., hardness determination and adhesion testing, are briefly discussed.

ACKNOWLEDGMENTS

The authors wish to thank the Fund for Scientific Research—Flanders (Belgium) and the Flemish Institute for the encouragement of the scientific and technological research in industry (I.W.T.) for their financial support, and the reviewers for their valuable suggestions to improve this paper.

APPENDIX: THE REFLECTION COEFFICIENT FOR THE LCS STRUCTURE

The formulas given in this appendix come from the book of Brekhovskikh.[10] However, some errors were still contained in their formulas for the LCS case, so the correct expressions are listed here. Figure 4 shows the considered configuration.

The reflection coefficient for the case with coating

$$R=\frac{Z_{in}-Z_3}{Z_{in}+Z_3}, \tag{A1}$$

where $Z_3=\rho_3 c_3/\cos\theta_3$ is the impedance of the liquid, and Z_{in}, the input impedance of the CS configuration, is defined as: $Z_{in}=iA'/\omega B'$ and

$$A'=k_{1z}M_{32}-i\omega^2\rho_1\left[\cos(2\gamma_1)M_{33}-\frac{v_{1s}^2}{v_{1d}^2}\sin(2\theta_1)M_{34}\right]-[k_xM_{32}-i\omega^2\rho_1(\sin(2\gamma_1)M_{33}+\cos(2\gamma_1)M_{34})]\frac{P}{W}, \tag{A2a}$$

$$B'=k_{1z}M_{22}-i\omega^2\rho_1\left[\cos(2\gamma_1)M_{23}-\frac{v_{1s}^2}{v_{1d}^2}\sin(2\theta_1)M_{24}\right]-[k_xM_{22}-i\omega^2\rho_1(\sin(2\gamma_1)M_{23}+\cos(2\gamma_1)M_{24})]\frac{P}{W}, \tag{A2b}$$

where the goniometric functions can be expressed in terms of the wave number components

$$\sin\theta_1=k_x/k_1, \tag{A3a}$$

$$\sin\gamma_1=k_x/\kappa_1. \tag{A3b}$$

The other goniometric functions can be derived with elementary goniometric relations. The ratio P/W

$$P/W=-\frac{k_xA_{41}-k_{1z}A_{42}+i\omega^2\rho_1\cos(2\gamma_1)A_{43}-i\omega^2\rho_1(v_{1s}^2/v_{1d}^2)\sin(2\theta_1)A_{44}}{\kappa_1\cos\gamma_1A_{41}+k_xA_{42}-i\omega^2\rho_1\sin(2\gamma_1)A_{43}-i\omega^2\rho_1\cos(2\gamma_1)A_{44}}. \tag{A4}$$

The symbols M_{ij} and A_{ij} are defined by: $M_{ik}=A_{ik}-A_{i1}A_{4k}/A_{41}$ with $i,k\in\{2,3,4\}$. The A matrix can be found in Brekhovskikh.[10] To calculate the reflection exactly, one needs to use that expression. This was done by Devolder[2,3] and is also done here in the Fourier simulations as a means to control the results of the approximated analytical theory.

However, under the conditions that the coating thickness is negligible in comparison to the sound wavelength $k_xd\simeq 0$, the A matrix simplifies to[17,10]

$$A=\begin{pmatrix}1 & 0 & 0 & d/(\rho v_{2s}^2)\\ 0 & 1 & +d/(\rho v_{2d}^2) & 0\\ 0 & -\rho\omega^2 d & 1 & 0\\ -\rho\omega^2 dF_2 & 0 & 0 & 1\end{pmatrix}, \tag{A5}$$

with

$$F_2=1-4(k_x/k_{2s})^2(1-(v_{2s}/v_{2d})^2). \tag{A6}$$

[1] P. McIntire, *Nondestructive Testing Handbook* (The American Society for Nondestructive Testing, Columbus, 1991), Vol. 7: Ultrasonic Testing.
[2] M. Wevers, S. Devolder, P. Demeester, and O. Leroy, "Acousto-optic technique: A new nondestructive technique to evaluate thin layered structures," Appl. Phys. Lett. **66**(12), 1466–1468 (1995).
[3] S. Devolder, M. Wevers, P. Demeester, and O. Leroy, "Thin layer measurements based on the acousto-optic technique," Appl. Phys. Lett. **68**(12), 1732–1734 (1996).
[4] Z. Guo, J. D. Achenbach, A. Madan, K. Martin, and M. E. Graham, "Integration of modeling and acoustic microscopy measurements for thin films," J. Acoust. Soc. Am. **107**, 2462–2471 (2000).
[5] A. H. Nayfeh, D. E. Chimenti, L. Adler, and R. L. Crane, "Ultrasonic leaky waves in the presence of a thin layer," J. Appl. Phys. **52**(8), 4985–4994 (1981).
[6] T. Kundu, A. K. Mal, and R. D. Weglein, "Calculation of the acoustic material signature of a layered solid," J. Acoust. Soc. Am. **77**, 353–361 (1985).
[7] S. Devolder, "A new acousto-optic nondestructive technique for the quality control of layered materials," Ph.D. thesis, K. U. Leuven Campus Kortrijk (1996).
[8] H. L. Bertoni and T. Tamir, "Unified theory of Rayleigh-angle phenomena for acoustic beams at liquid–solid interfaces," Appl. Phys. **2**, 157–172 (1973).

[9] E. Blomme and O. Leroy, ''Plane wave analysis of the near field of light diffracted by ultrasound,'' J. Acoust. Soc. Am. **91**, 1474–1483 (1992).
[10] L. M. Brekhovskikh, *Waves in Layered Media* (Academic, San Diego, 1980), pp. 53–60.
[11] X. Jia, ''Normal-mode theory of nonspecular phenomena for a finite-aperture ultrasonic beam reflected from layered media,'' Appl. Phys. Lett. **70**(2), 1–4 (1997).
[12] X. Jia, ''Modal analysis of Lamb wave generation in elastic plates by liquid wedge transducers,'' J. Acoust. Soc. Am. **101**, 1–9 (1997).
[13] B. A. Auld, *Acoustic Fields and Waves in Solids* (Wiley, New York, 1973), Vol. 2, Chap. 10.
[14] R. Briers, O. Leroy, and G. N. Shkerdin, ''Influence of an inclusion on the phase of a Stoneley wave,'' J. Acoust. Soc. Am. **101**, 1358–1365 (1997).
[15] J. Vandeputte, O. Leroy, R. Briers, and G. Shkerdin, ''Extension of the mode method for viscoelastic media and focused ultrasonic beams,'' J. Acoust. Soc. Am. **108**, 1614–1621 (2000).
[16] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables* (Wiley, New York, 1972), pp. 297–300.
[17] 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).

Vibratory characteristics of multistep nonuniform orthotropic shear plates with line spring supports and line masses

Q. S. Li
Department of Building and Construction, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong

(Received 25 July 2000; accepted for publication 22 May 2001)

A new exact approach for free-vibration analysis of multistep nonuniform orthotropic shear plates with line spring supports and line masses is presented. The governing differential equation for free vibrations of an orthotropic shear plate with variably distributed mass and stiffness is established. It is proved that it is possible to separate a shear plate as two independent shear beams for free-vibration analysis. The jkth natural frequency of a shear plate is equal to the square root of the square sum of the jth natural frequency of a shear beam and the kth natural frequency of another shear beam. The jkth mode shape of the shear plate is the product of the jth mode shape of a shear beam and the kth mode shape of another shear beam. In this paper, the function for describing the distribution of mass of each step plate can be selected as an arbitrary one, and the distribution of shear stiffness is expressed as a functional relation with the mass distribution, and vice versa. The exact solutions of one-step shear plates with varying cross section are obtained first for eight cases. Then, the derived exact solutions are used to establish the frequency equation of a multistep nonuniform orthotropic shear plate with spring supports and line masses using the transfer matrix method and the recurrence method developed in this paper. The numerical example shows that the calculated results are in good agreement with the experimental data, and the proposed procedure is an exact and efficient method. [DOI: 10.1121/1.1387995]

PACS numbers: 43.40.At [PJR]

I. INTRODUCTION

Experimental results obtained from dynamic testing of buildings (e.g., Wang, 1978; Li, 1985; Li *et al.*, 1994; Li *et al.*, 1996; Li *et al.*, 1999; Jeary, 1997) have shown that for many cases the shear deformation is dominant in the total deformation of multistory frame buildings in their horizontal vibrations. Such buildings are usually called shear-type buildings. Korqingskee (1952) investigated the free vibration of frame buildings that are considered as a multistep cantilever shear beam, in which each step of the beam has constant parameters (mass and stiffness). Wang (1978) suggested that frame buildings and other shear-type buildings could be treated as a one-step cantilever shear beam with variably distributed mass and stiffness along the height of a beam for the analysis of free vibration. He derived the closed-form solutions for such a problem. But, he assumed that the mass of the shear beam is proportional to its stiffness. Li (1998) recently proposed an exact approach to determine the dynamic characteristics of cantilever shear beams with variably distributed mass and stiffness. In his study, the value of mass of a shear beam is not necessarily proportional to its stiffness. However, if a building has a narrow rectangular plane configuration (narrow buildings), $B/L<\frac{1}{4}$, where B and L are the width and length of the rectangular plane, respectively, the stiffness of each floor of the building cannot be treated as infinitely rigid. Such a building may not be simplified as a cantilever shear beam for free-vibration analysis. When analyzing free vibration of narrow buildings, it is reasonable to regard such structures as a shear plate in which shear deformation is dominant (Li, 2000). Because the stiffness of such a plate in the x direction in Fig. 1 is different from that in the y direction, the shear plate considered in this paper is an orthotropic shear plate with variably distributed stiffness and mass.

In fact, there are very few equations of vibrating plates with variable cross section where exact analytical solutions can be obtained. An analytical approach for the vibration of a simply support flexural plate with one change in thickness was developed by Chopra (1974). Guo *et al.* (1997) found the closed-form solutions for the free vibration of a stepped, simply supported flexural plate with uniform thickness and abrupt thickness changes. The concept of orthotropic shear plates was developed and used by Beiner and Librescu (1984). They have presented an analysis of weight minimization for rectangular flat panels with fixed flutter speed. To simplify the problem, a structural model that considers transverse shear deformation only and neglects the bending stiffness of the plate was adopted in their study. This has the effect of reducing the linear partial differential equation for this problem from the fourth to the second order. However, Beiner and Librescu (1984) did not consider free-vibration analysis of orthotropic shear plates. Wang (1978) derived the exact analytical solutions for free vibration of cantilever shear plates with uniformly distributed mass and stiffness. However, it is obvious that the distributions of mass and shear stiffness of most narrow buildings are actually not uniform, especially along the building heights. In general, the variation of mass and stiffness along the longitudinal axis of a narrow building (the x direction in Fig. 1) can be neglected (Li *et al.*, 1994). Thus, it is assumed that the narrow buildings considered herein have uniformly distributed mass and shear stiffness along the longitudinal axis, but variably dis-

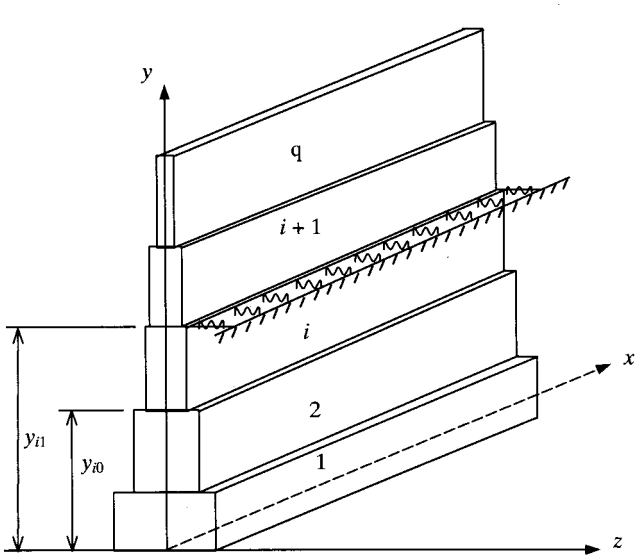


FIG. 1. A multistep shear plate with a line spring support at upper edge of the *i*th step plate.

tributed mass and shear stiffness along the heights of the narrow buildings. In this paper, the function for describing the distribution of mass of each step plate can be selected as an arbitrary one, and the distribution of shear stiffness is expressed as a functional relation with the mass distribution, and vice versa. Thus, an exact solution obtained herein actually represents a class of exact solutions for the title problem. Exact solutions of one-step shear plates for eight cases are obtained first. The analysis of free vibration of a nonuniform multistep shear plate with spring supports and line masses is a complex problem and the exact solution of this problem has not been obtained previously. Use of the derived exact solution of a one-step shear plate with variable cross section, together with a transfer matrix method and a recurrence method proposed in this paper, is presented in order to resolve this problem. Numerical example shows that the calculated results are in good agreement with the experimental data and it is convenient to apply the proposed method to free-vibration analysis of elastically restrained shear plates with varying cross section.

II. THEORY

In order to establish the differential equation for vibration of a multistep orthotropic shear plate (Fig. 1) in which shear deformation is dominant, an infinitesimal plate element is cut from the *i*th step plate, as shown in Fig. 2. The size of the element is $dx \times dy$. The dynamic loading acting on the element is $q(x,y,t)dx\,dy$. The inertial force is $-\overline{m}_{ixy}(\partial^2 W_i/dt^2)dx\,dy$ and the damping force is $-C_{ixy}(\partial W_i/dt)dx\,dy$, where $\overline{m}_{ixy}$, W_i, and C_{ixy} are the mass intensity (mass per unit area), dynamic displacement in the z direction, and viscous damping coefficient at the point (x,y), respectively. The element shown in Fig. 2 is rotated over an angle of 90°. According to the d'Alembert principle, all forces acting on the element including the inertial force should satisfy the equilibrium conditions. From $\Sigma F_z = 0$, we obtain

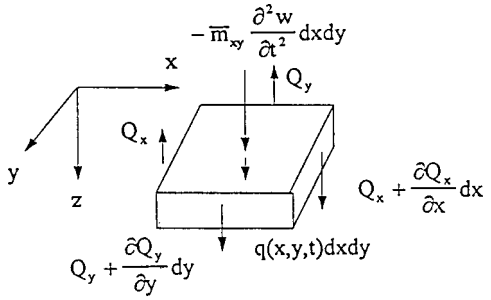


FIG. 2. An element of the shear plate.

$$\frac{\partial Q_{ix}}{\partial x} + \frac{\partial Q_{iy}}{\partial y} - C_{ixy}\frac{\partial W_i}{\partial t} - \overline{m}_{ixy}\frac{\partial^2 W_i}{\partial t^2} = -q(x,y,t), \quad (1)$$

where Q_{ix} and Q_{iy} are the transverse shear forces in the x direction and in the y direction, respectively.

For free vibration, $q(x,y,t)=0$. Considering that the effect of damping on the natural frequency, in general, is not significant and setting $C_{ixy}=0$, one thus obtains the governing differential equation for undamped free vibration of the *i*th step orthotropic shear plate as follows:

$$\frac{\partial Q_{ix}}{\partial x} + \frac{\partial Q_{iy}}{\partial y} - \overline{m}_{ixy}\frac{\partial^2 W_i}{\partial t^2} = 0. \quad (2)$$

The shear forces can be expressed as

$$Q_{ix} = K_{ix}\frac{\partial W_i}{\partial x}, \quad Q_{iy} = K_{iy}\frac{\partial W_i}{\partial y}, \quad (3)$$

in which K_{ix} and K_{iy} are the transverse shear stiffness in the x direction and in the y direction, respectively.

Substituting Eq. (3) into Eq. (2) yields

$$\frac{\partial}{\partial x}\left(K_{ix}\frac{\partial W_i}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_{iy}\frac{\partial W_i}{\partial y}\right) - \overline{m}_{ixy}\frac{\partial^2 W_i}{\partial t^2} = 0. \quad (4)$$

The method of separation of variables is adopted herein

$$W_i(x,y,t) = X_i(x)Y_i(y)\sin(\omega t + \gamma_0). \quad (5)$$

It is assumed that K_{iy} is a function of y; K_{ix} and $\overline{m}_{ixy}$ are also functions of y

$$K_{iy} = K_{i1}f_i(y), \quad K_{ix} = K_2\varphi_i(y), \quad \overline{m}_{ixy} = \overline{m}\varphi_i(y), \quad (6)$$

i.e., it is assumed that K_{ix} is directly proportional to $\overline{m}_{ixy}$. Since the values of K_{ix} and $\overline{m}_{ixy}$ are mainly dependent on the dimensions and materials of building floors, this assumption is thus reasonable for most narrow buildings.

Substituting Eqs. (5) and (6) into Eq. (4), one obtains

$$\frac{\frac{d}{dy}\left[K_{i1}f_i(y)\frac{dY_i(y)}{dy}\right]}{Y_i(y)\varphi_i(y)} = -\frac{K_2\frac{d^2X_i(x)}{dx^2}}{X_i(x)} - \overline{m}\omega^2. \quad (7)$$

The left-hand side of this equation is a function of y and the right-hand side is a function of x. Thus, the entire equation can be satisfied for arbitrary values of x and y only if both sides are equal to a constant. It is assumed that the constant is $-\overline{m}\theta^2$; then, the following two independent ordinary differential equations are obtained from Eq. (7):

$$K_2\frac{d^2X_i(x)}{dx^2}+\overline{m}\Omega^2X_i(x)=0, \tag{8}$$

$$\frac{d}{dy}\left[K_{i1}f_i(y)\frac{dY_i(y)}{dy}\right]+\overline{m}\varphi_i(y)\theta^2Y_i(y)=0, \tag{9}$$

in which

$$\Omega^2=\omega^2-\theta^2, \quad \omega=\sqrt{\Omega^2+\theta^2}. \tag{10}$$

It is obvious that Eqs. (8) and (9) are two governing equations of vibration mode shapes of two independent shear beams. One is a shear beam in the x direction, where K_2, $\overline{m}$, Ω are the stiffness, mass intensity, and circular natural frequency of this shear beam, respectively; its boundary conditions are the same as those of the shear plate in the x direction. The other one is a shear beam in the y direction, where $K_{i1}f_i(y)$, $\overline{m}\varphi_i(y)$, θ are the stiffness, mass intensity, and circular natural frequency of this shear beam, respectively; its boundary conditions are the same as those of the shear plate in the y direction. The natural frequency of the plate is equal to the square root of the square sum of the two natural frequencies of the two shear beams. The mode shape of the plate is the product of the corresponding two mode shapes of the two beams. This suggests that a shear plate may be simplified as two independent shear beams with the same boundary conditions as those of the shear plate for free-vibration analysis. The general solution of Eq. (8) is easily found as

$$X_i(x)=D_1\sin\frac{\Omega}{\alpha_2}x+D_2\cos\frac{\Omega}{\alpha_2}x \quad (i=1,2,...,q), \tag{11}$$

where q is the number of steps divided, and

$$\alpha_2=\sqrt{\frac{K_{i2}}{\overline{m}_i}}. \tag{12}$$

It is assumed that α_2 is a constant. This assumption is reasonable for many narrow buildings, because both K_{i2} and $\overline{m}_i$ are mainly dependent on the dimensions and materials of the ith floor of the building. Since the general solutions of $X_i(x)$ $(i=1,2,...,q)$ depend on α_2, which is a constant for all steps, $X_i(x)$ takes the same function for all steps. The subscript i in $X_i(x)$ can be eliminated in the rest of this paper.

The frequency equation and mode shape in the x direction can be determined using Eq. (11) and the boundary conditions in the x direction of the shear plate as follows:

(1) A shear plate with free–free (F–F) edges in the x direction. The boundary conditions for this case are given by

$$\frac{dX(x)}{dx}=0 \quad \text{at } x=0 \quad \text{and } x=L. \tag{13}$$

Using the boundary condition at $x=0$ and Eq. (11), one obtains $D_1=0$; then, the frequency equation can be obtained from the boundary condition at $x=L$ as

$$\sin\frac{\Omega}{\alpha_2}L=0. \tag{14}$$

The kth circular natural frequency and mode shape are found as

$$\Omega_k=\frac{\alpha_2(k-1)\pi}{L} \quad k=1,2,..., \tag{15}$$

$$X_k(x)=\cos\frac{(k-1)\pi x}{L}. \tag{16}$$

(2) A shear plate with clamped–clamped (C–C) edges or simply supported edges in the x direction.
The boundary conditions for this case are

$$X(x)=0 \quad \text{at } x=0 \quad \text{and } x=L. \tag{17}$$

Using Eqs. (11) and (17) gives the kth circular natural frequency and mode shape as follows:

$$\Omega_k=\frac{\alpha_2k\pi}{L} \quad k=1,2,..., \tag{18}$$

$$X_k(x)=\sin\frac{k\pi x}{L}. \tag{19}$$

(3) A shear plate with clamped–spring (C–S) edges and a line mass is attached at the spring edge in the x direction. If the edge at $x=0$ is clamped, then the boundary condition, $X(0)=0$, is substituted into Eq. (11), leading to $D_2=0$.
Since the edge at $x=L$ is a spring-supported one with line mass, the spring stiffness and the line mass intensity (mass per unit length) are K_{uL} and $\overline{m}_L$, respectively; the boundary condition at this edge is

$$X'(L)=-\frac{K_{uL}-m_L\Omega^2}{K_2}X(L), \tag{20}$$

where

$$X'(L)=\left.\frac{dX}{dx}\right|_{x=L}. \tag{21}$$

Using Eq. (20) one obtains the frequency equation as

$$\tan\frac{\Omega}{\alpha_2}L=-\frac{K_2\Omega}{\alpha_2(K_{uL}-m_L\Omega^2)}. \tag{22}$$

The kth mode shape in the x direction can be written as

$$X_k(x)=\sin\frac{\Omega_kx}{\alpha_2}, \tag{23}$$

in which Ω_k is the kth circular natural frequency of the beam in the x direction.

(4) A shear plate with spring–spring (S–S) edges and line masses in the x direction. If the opposite edges of a shear plate in the x direction are spring-supported ones with line masses, then the boundary conditions can be written as

$$X'(0)=-\frac{K_{u0}-m_0\Omega^2}{K_2}X(0), \tag{24}$$

$$X'(L)=-\frac{K_{uL}-m_L\Omega^2}{K_2}X(L), \tag{25}$$

where K_{u0} and m_0 are the spring stiffness and line mass intensity, respectively, at the edge $x=0$.

Using Eqs. (24), (25), and (11) one obtains the frequency equation of the shear beam in the x direction as follows:

$$\frac{\Omega}{\alpha_2}\tan\frac{\Omega L}{\alpha_2}-\frac{K_{uL}-m_L\Omega^2}{K_2} =\frac{K_{u0}-m_0\Omega^2}{K_2}\left[1+\frac{\alpha_2(K_{uL}-m_L\Omega^2)}{\Omega K_2}\tan\frac{\Omega L}{\alpha_2}\right]. \tag{26}$$

Substituting the kth circular natural frequency Ω_k into Eq. (11), one obtains the kth mode shape of the shear beam in the x direction.

As mentioned previously in this paper, the function for describing the distribution of mass of each step plate along the plate height can be selected as an arbitrary one, and the distribution of shear stiffness is expressed as a functional relation with the mass distribution, and vice versa. In order to solve Eq. (9) the following functional transformations are introduced:

$$Y_i(y)=Y_i(\zeta),\quad \overline{m}_{ixy}=\text{arbitrary function of } y,\quad K_{iy}=\overline{m}_{xy}^{-1}P_i(\zeta) \tag{27}$$

or

$$Y_i(y)=Y_i(\zeta),\quad K_{iy}=K_{il}f(y)=\text{arbitrary function of } y,\quad \overline{m}_{ixy}=K_{iy}^{-1}P_i(\zeta), \tag{28}$$

in which

$$\zeta=\int K_{iy}^{-1}\,dy. \tag{29}$$

Substituting Eq. (27) or Eq. (28) into Eq. (9) yields

$$\frac{d^2Y_i(\zeta)}{d\zeta^2}+\theta^2P_i(\zeta)Y_i(\zeta)=0. \tag{30}$$

Since ζ is a function of y, $y(\zeta)$ is a functional expression; a solution of Eq. (30) actually represents a class of solution.

The solutions of Eq. (30) depend on the expression of $P(\zeta)$ (in order to simplify the expressions of formulas, the subscript i is eliminated in the following parts of this paper). Several important cases of the expression of $P(\zeta)$ will be considered and discussed as follows:

$$\text{Case 1:}\quad P(\zeta)=(a+b\zeta)^c. \tag{31}$$

Substituting Eq. (31) into Eq. (30) and then using the following functional transformation:

$$\eta=(a+b\zeta)^{1/2\nu},\quad Y=\eta^\nu V,\quad \nu=\frac{1}{c+2}, \tag{32}$$

leads to a Bessel equation of order ν.

The special solutions of $Y(\zeta)$ are found as

$$S_1(\zeta)=(a+b\zeta)^{1/2}J_\nu[\alpha(a+b\zeta)^{1/2\nu}], \tag{33}$$

$$S_2(\zeta)=\begin{cases}(a+b\zeta)^{1/2}J_{-\nu}[\alpha(a+b\zeta)^{1/2\nu}], & \nu=\text{a noninteger}\\ (a+b\zeta)^{1/2}Y_\nu[\alpha(a+b\zeta)^{1/2\nu}], & \nu=\text{an integer}\end{cases} \tag{34}$$

in which

$$\alpha^2=4\nu^2\theta^2/b^2, \tag{35}$$

and J_ν and Y_ν are Bessel functions of the first and second kind of order ν, respectively.

For a special case, $c=-2$, Eq. (30) becomes an Euler equation; the special solutions are

$$\left.\begin{aligned}S_1(\zeta)&=(a+b\zeta)^{1/2}\sin[\overline{\alpha}\ln(a+b\zeta)]\\ S_2(\zeta)&=(a+b\zeta)^{1/2}\cos[\overline{\alpha}\ln(a+b\zeta)]\end{aligned}\right\}\quad \text{for } 4\theta^2-b^2>0, \tag{36}$$

or

$$\left.\begin{aligned}S_1(\zeta)&=(a+b\zeta)^{(1/2)+\overline{\alpha}}\\ S_2(\zeta)&=(a+b\zeta)^{(1/2)-\overline{\alpha}}\end{aligned}\right\}\quad \text{for } 4\theta^2-b^2<0, \tag{37}$$

or

$$\left.\begin{aligned}S_1(\zeta)&=(a+b\zeta)^{1/2}\\ S_2(\zeta)&=(a+b\zeta)^{1/2}\ln(a+b\zeta)\end{aligned}\right\}\quad \text{for } 4\theta^2-b^2=0, \tag{38}$$

in which

$$\overline{\alpha}^2=\frac{4\theta^2-b^2}{4b^2}. \tag{39}$$

$$\text{Case 2:}\quad P(\zeta)=d\zeta^c. \tag{40}$$

This is a special form of case 1. The special solutions, $S_1(\zeta)$ and $S_2(\zeta)$, can be obtained from Eq. (33) to Eq. (39) by setting $a=0$ and $b=d^{1/c}$.

$$\text{Case 3:}\quad P(\zeta)=a(1+b\zeta)^c. \tag{41}$$

This case is an alteration of case 2. The special solutions are given by

$$S_1(\zeta)=(1+b\zeta)^{1/2}J_\nu[\overline{\alpha}(1+b\zeta)^{1/2\nu}], \tag{42}$$

$$S_2(\zeta)=\begin{cases}(1+b\zeta)^{1/2}J_{-\nu}[\overline{\alpha}(1+b\zeta)^{1/2\nu}], & \nu=\text{a noninteger}\\ (1+b\zeta)^{1/2}Y_\nu[\overline{\alpha}(1+b\zeta)^{1/2\nu}], & \nu=\text{an integer.}\end{cases} \tag{43}$$

If $c=-2$, then the special solutions are similar to those given in Eqs. (36), (37), and (38).

$$\text{Case 4:}\quad P(\zeta)=a(\zeta^2+b)^{-2},\quad a>0,\quad b>0. \tag{44}$$

The special solutions for this case are found as

$$\left.\begin{aligned}S_1(\zeta)&=(\zeta^2+b)^{1/2}\sin\xi\\ S_2(\zeta)&=(\zeta^2+b)^{1/2}\cos\xi,\end{aligned}\right. \tag{45}$$

where

$$\xi=\left(\frac{a+b}{b}\right)^{1/2}\arctan\frac{\zeta}{b^{1/2}}. \tag{46}$$

$$\text{Case 5:}\quad P(\zeta)=a(\zeta^2-b)^{-2},\quad a>0,\quad b>0.$$

The special solutions for this case are

$$\left.\begin{aligned}S_1(\zeta)&=(b-\zeta^2)^{1/2}\sin\xi\\ S_2(\zeta)&=(b-\zeta^2)^{1/2}\cos\xi,\end{aligned}\right. \tag{47}$$

where

$$\xi=\frac{1}{2}\left(\frac{a-b}{b}\right)^{1/2}\ln\frac{b^{1/2}+\zeta}{b^{1/2}-\zeta}. \tag{48}$$

$$\text{Case 6: } P(\zeta)=-c[\zeta^2-(a+b)\zeta+ab]^{-2}. \tag{49}$$

The special solutions for case 6 are given by

$$\begin{aligned} S_1(\zeta)&=|\zeta-a|^{(1+\gamma)/2}|\zeta-b|^{(1-\gamma)/2}\\ S_2(\zeta)&=|\zeta-a|^{(1-\gamma)/2}|\zeta-b|^{(1+\gamma)/2}, \end{aligned} \tag{50}$$

where

$$\gamma^2=\frac{4c}{(a-b)^2}+1\neq 0, \quad a\neq b. \tag{51}$$

$$\text{Case 7: } P(\zeta)=ae^{b\zeta}-c. \tag{52}$$

The special solutions for this important case are found as

$$S_1(\zeta)=J_\nu(\lambda e^{b\zeta/2}), \tag{53}$$

$$S_2(\zeta)=\begin{cases} J_{-\nu}(\lambda e^{b\zeta/2}), & \nu=\text{a noninteger}\\ Y_\nu(\lambda e^{b\zeta/2}), & \nu=\text{an integer}\end{cases} \tag{54}$$

in which

$$\lambda^2=\frac{4\theta^2 a}{b^2}, \quad a>0 \tag{55}$$

$$\nu^2=\frac{4c}{b^2}, \quad c>0. \tag{56}$$

If $c=0$, then $\nu=0$.

$$\text{Case 8: } P(\zeta)=k^2=\text{constant}. \tag{57}$$

This case represents a uniform plate. The special solutions for this case are

$$S_1(\zeta)=\sin k\zeta, \quad S_2(\zeta)=\cos k\zeta, \tag{58}$$

or

$$S_1(y)=\sin\frac{\theta}{\alpha_1}y, \quad S_2(y)=\cos\frac{\theta}{\alpha_1}y, \tag{59}$$

in which $\alpha_1=\sqrt{K_1/m}$ and K_1 is the shear stiffness of the shear plate in the y direction.

The complete solutions of the ith step plate in the y direction can be written as

$$Y_i(y)=D_{i1}S_{i1}(y)+D_{i2}S_{i2}(y). \tag{60}$$

In order to establish the frequency equation and the equation of mode shape of a multistep shear plate in the y direction, a transfer matrix method and a recurrence method are introduced as follows.

A. Transfer matrix method

The displacement $Y_i(y)$ and the shear force $Q_{iy}(y)$ can be expressed as

$$\begin{bmatrix} Y_i(y)\\ Q_{iy}(y)\end{bmatrix}=\begin{bmatrix} S_{i1}(y) & S_{i2}(y)\\ K_{iy}S'_{i1}(y) & K_{iy}S'_{i2}(y)\end{bmatrix}\begin{bmatrix} D_{i1}\\ D_{i2}\end{bmatrix}. \tag{61}$$

The relationship between the parameters, $Y_i(y)$ and $Q_{iy}(y)$ at the two ends of the ith step can be expressed as

$$\begin{bmatrix} Y_{i1}\\ Q_{i1}\end{bmatrix}=[T_i]\begin{bmatrix} Y_{i0}\\ Q_{i0}\end{bmatrix}, \tag{62}$$

in which

$$[T_i]=[S(y_{i1})][S(y_{i0})]^{-1},$$

$$[S(y)]=\begin{bmatrix} S_{i1}(y) & S_{i2}(y)\\ K_{iy}S'_{i1}(y) & K_{iy}S'_{i2}(y)\end{bmatrix},$$

$$Y_{i0}=Y_i(y_{i0}), \quad Y_{i1}=Y_i(y_{i1}),$$

$$Q_{i0}=Q_{iy}(y_{i0}), \quad Q_{i1}=Q_{iy}(y_{i1}).$$

$[T_i]$ is called the transfer matrix, because it transfers the parameters at the end y_{i0} to those at the end y_{i1} of the ith step plate.

As is well known, the displacement and slope at all common interfaces of two neighboring steps are required to be continuous

$$Y_{i0}=Y_{i-1,1}, \quad Q_{i0}=Q_{i-1,1} \quad \text{or} \quad Y'_{i0}=\frac{K_{i-1,1}}{K_{i0}}Y'_{i-1,1}, \tag{63}$$

in which

$$K_{i-1,1}=K_{i-1,y}(y_{i-1,1}), \quad K_{i0}=K_{iy}(y_{i0}).$$

Substituting Eq. (63) into Eq. (62) leads to

$$\begin{bmatrix} Y_{i1}\\ Q_{i1}\end{bmatrix}=[T_i][T_{i-1}]\begin{bmatrix} Y_{i-1,0}\\ Q_{i-1,0}\end{bmatrix}. \tag{64}$$

Using Eqs. (64) and (63) repeatedly yields

$$\begin{bmatrix} Y_{q1}\\ Q_{q1}\end{bmatrix}=[T]\begin{bmatrix} Y_{10}\\ Q_{10}\end{bmatrix}, \tag{65}$$

in which

$$[T]=[T_q][T_{q-1}]\cdots[T_1]. \tag{66}$$

$[T]$ is a matrix which can be expressed as

$$[T]=\begin{bmatrix} T_{11} & T_{12}\\ T_{21} & T_{22}\end{bmatrix}. \tag{67}$$

The elements of this matrix can be determined from Eq. (66).

If there are a line mass and a line spring attached to the interface of the ith step and the $i+1$th step (Fig. 1), then the displacement and shear force are required to satisfy the following conditions:

$$Y_{i+1,0}=Y_{i1}$$

$$Q_{i+1,0}=Q_{i1}+(K_i-m_i\theta^2)Y_{i1}$$

or

$$Y'_{i+1,0}=\frac{K_{i1}}{K_{i+1,0}}Y'_{i1}+\frac{K_i-m_i\theta^2}{K_{i+1,0}}Y_{i1}, \tag{68}$$

where m_i is the intensity of the line mass and K_i is the stiffness of the line spring. It can be seen from Eq. (68) that if a line spring and a line mass are attached at y_{i1}, $[T_i]$ should be replaced by

$$[T_{mi}]=\begin{bmatrix} 1 & 0 \\ K_i-m_i\omega^2 & 1 \end{bmatrix}[T_i]. \tag{69}$$

The frequency equation can be established by use of Eq. (65) and the boundary conditions as follows:

(1) A multistep shear plate with F–F edges in the y direction.
The boundary conditions for this case are given by

$$Q_{10}=0 \quad \text{and} \quad Q_{q1}=0. \tag{70}$$

Substituting Eq. (70) into Eq. (65), one obtains

$$\begin{bmatrix} Y_{q1} \\ 0 \end{bmatrix}=\begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix}\begin{bmatrix} Y_{10} \\ 0 \end{bmatrix}. \tag{71}$$

From this equation we have

$$T_{21}Y_{10}=0. \tag{72}$$

Because $Y_{10}\neq 0$, one obtains

$$T_{21}=0. \tag{73}$$

This is the frequency equation of a multistep shear plate (in the y direction) with F–F edges.

(2) A multistep shear plate with C–C edges in the y direction.
The boundary conditions for this case are

$$Y_{10}=0 \quad \text{and} \quad Y_{q1}=0. \tag{74}$$

Using Eqs. (65) and (74) gives the frequency equation as

$$T_{12}=0. \tag{75}$$

(3) A multistep shear plate with C–F edges in the y direction.
If the edge at $y=0$ is clamped, then $Y_{10}=0$. For the free edge at $y=H$, we have $Q_{q1}=0$.
Using Eq. (65) and these boundary conditions, one obtains the frequency equation as

$$T_{22}=0. \tag{76}$$

(4) A multistep shear plate with C–S edges and with a line mass at spring edge $y=H$ in the y direction.
If the edge at $y=0$ is clamped and the edge at $y=H$ is a line spring-supported one with a line mass, then the boundary conditions can be written as

$$Y_{10}=0, \quad Q_{q1}=-(K_{uH}-m_H\theta^2)Y_{q1}. \tag{77}$$

The boundary conditions are substituted into Eq. (65), leading to the frequency equation

$$T_{22}+(K_{uH}-m_H\theta^2)T_{12}=0, \tag{78}$$

in which m_H and K_{uH} are the line mass intensity and line spring stiffness, respectively, attached at $y=H$.

(5) A multistep shear plate with S–S edges and line masses at $y=0$ and $y=H$.
The boundary conditions for this case can be written as

$$\begin{aligned} Q_{10}&=(K_{u0}-m_0\theta^2)Y_{10}, \\ Q_{q1}&=-(K_{uH}-m_H\theta^2)Y_{q1}, \end{aligned} \tag{79}$$

in which m_0 and K_{u0} are the line mass intensity and line spring stiffness, respectively, attached at $y=0$.

Substituting Eq. (79) into Eq. (65) yields the frequency equation

$$\begin{aligned} &(K_{uH}-m_H\theta^2)[T_{11}+T_{12}(K_{u0}-m_0\theta^2)] \\ &\quad +T_{21}+T_{22}(K_{u0}-m_0\theta^2)=0. \end{aligned} \tag{80}$$

It is necessary to point out that Ω_k and θ_j obtained from the frequency equation in the x direction and y direction are not the circular natural frequencies of the shear plate in general cases. The circular natural frequencies should be determined from

$$\omega_{jk}=\sqrt{\theta_j^2+\Omega_K^2}. \tag{81}$$

The corresponding mode shape function, $Z_{jk}(x,y)$, can be found from

$$Z_{jk}(x,y)=Y_j(y)X_k(x), \tag{82}$$

where $X_k(x)$ and $Y_j(y)$ are the kth mode shape and jth mode shape in the x direction and y direction, respectively. $X_k(x)$ can be found from Eqs. (16)–(26). For a multistep shear plate with C–F edges in the y direction, as an example, $Y_j(x)$ can determined using the following procedure:

(1) Setting $i=1$, $Y_{10}=0$, and $Q_{10}=1$ (or any value), and substituting θ_j into Eq. (62), one obtains Y_{11j} and Q_{11j}.
(2) Using Eq. (63) leads to $Y_{20j}=Y_{11j}$, $Q_{20j}=Q_{11j}$ and substituting Y_{20j}, Q_{20j} into Eq. (62) one obtains Y_{21j} and Q_{21j}.
(3) Repeating the above procedures yields Y_{i1j} and Q_{i1j} ($i=1,2,...,q$); then, substituting Y_{i1j} and Q_{i1j} into Eq. (61), one obtains D_{i1j} and D_{i2j}. Thus, $Y_j(y)$ is determined. In general, Y_{i1j} and Q_{i1j} are required only.

B. Recurrence method

In order to simplify the analysis for the title problem, the linearly independent fundamental solutions, $\overline{S}_{i1}(y)$ and $\overline{S}_{i2}(y)$, satisfy the following normalization condition at the origin of coordinate system:

$$\begin{bmatrix} \overline{S}_{i1}(0) & \overline{S}'_{i1}(0) \\ \overline{S}_{i2}(0) & \overline{S}'_{i2}(0) \end{bmatrix}=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{83}$$

can be easily constructed by using the linearly independent special solutions, $S_{i1}(y)$ and $S_{i2}(y)$ derived previously, and the following matrix equation:

$$\begin{bmatrix} \overline{S}_{i1}(y) \\ \overline{S}_{i2}(y) \end{bmatrix}=\begin{bmatrix} S_{i1}(0) & S'_{i1}(0) \\ S_{i2}(0) & S'_{i2}(0) \end{bmatrix}^{-1}\begin{bmatrix} S_{i1}(y) \\ S_{i2}(y) \end{bmatrix}. \tag{84}$$

In general, the mode shape function of the ith step plate in the y direction can be expressed in terms of the fundamental solutions as follows:

$$Y_i(y)=Y_{i0}\overline{S}_{i1}(y)+Y'_{i0}\overline{S}_{i2}(y), \tag{85}$$

where Y_{i0} and Y'_{i0} are the displacement and slope of the displacement curve of the ith step plate at $y=0$, respectively; the origin of the coordinate is set at the lower end of the ith step plate.

Substituting Eq. (68) into Eq. (85) leads to

$$Y_i(y)=Y_{i-1,1}\overline{S}_{i1}(y)+\frac{1}{K_{i0}}\times[K_{i-1}Y'_{i-1,1}+(K_{i-1}-m_{i-1}\theta^2)Y_{i-1,1}]\overline{S}_{i2}(y). \tag{86}$$

This is a recurrence formula. Using this formula and $Y_1(y)$, which has only one unknown parameter for any type of the lower end of the first step, we can determine the mode shape function of the i-th step ($i=2,3,...,q$), which has the same unknown parameter as that in the expression of $Y_1(y)$.

The frequency equation of a multistep shear plate in the y direction can be established by using $Y_1(y)$, $Y_q(y)$ and the boundary conditions as follows:

(1) A multistep shear plate with F–F edges in the y-direction.
The boundary conditions for this case are given in Eq. (70), and can be written as

$$Y_1(y)=Y_{10}\overline{S}_{12}(y). \tag{87}$$

Using Eqs. (87) and (86), one obtains $Y_i(y)$ ($i=2,3,...,q$). The frequency equation of the shear plate in the y direction is

$$Y'_{q1}=0. \tag{88}$$

(2) A multistep shear plate with C–C edges in the y direction.
The boundary conditions for this case are the same as those given in Eq. (74). $Y_1(y)$ can be written as

$$Y_1(y)=Y'_{10}\overline{S}_{12}(y). \tag{89}$$

The frequency equation is

$$Y_{q1}=0. \tag{90}$$

(3) A multistep shear plate with C–F edges in the y direction.
$Y_1(y)$ for this case is given by

$$Y_1(y)=Y'_{10}\overline{S}_{12}(y). \tag{91}$$

The frequency equation is

$$Y'_{q1}=0. \tag{92}$$

It is necessary to point out that Eq. (92) is different from Eq. (88). It can be seen from Eq. (91) that Y'_{10} is an unknown parameter in Eq. (92), but Y_{10} is an unknown parameter in Eq. (88).

(4) A multistep shear plate with C–S edges and with line mass at spring edge, $y=H$, in the y direction.
$Y_1(y)$ for this case is

$$Y_1(y)=Y'_{10}\overline{S}_{12}(y). \tag{93}$$

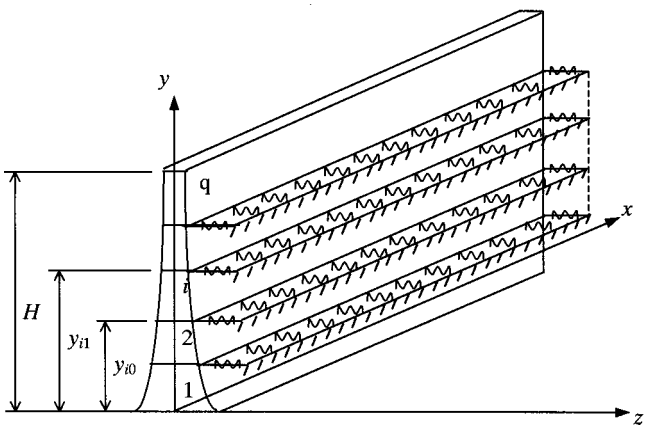


FIG. 3. A one-step nonuniform shear plate with C–F edges and line spring supports and line masses. (Note: the line masses are not shown in Fig. 3).

The frequency equation is

$$K_{q1}Y'_{q1}+(K_{uH}-m_H\theta^2)Y_{q1}=0. \tag{94}$$

(5) A multistep shear plate with S–S edges and line masses at $y=0$ and $y=H$.
$Y_1(y)$ for this case is

$$Y_1(y)=\left[\overline{S}_{11}(y)+\frac{K_{u0}-m_0\theta^2}{K_{10}}\overline{S}_{12}(y)\right]Y_{10}. \tag{95}$$

The frequency equation is the same as Eq. (94).

(6) A one-step shear plate with C–F edges and line spring supports, line masses at the $(q-1)$ intermediate lines in the y direction (Fig. 3).
It is assumed that the fundamental solutions for this case are $\overline{S}_1(y)$ and $\overline{S}_2(y)$. If the lower edge of the first segment is clamped, $Y_1(y)$ is given by

$$Y_1(y)=Y'_{10}\overline{S}_2(y).$$

The mode shape of the other segments can be written as

$$Y_i(y)=Y_1(y)+\sum_{j=1}^{i-1}\frac{1}{K_y(y_{j1})}(K_j-m_j\theta^2)Y_j(y_{j1})\times\overline{S}_2(y-y_{j1})H(y-y_{j1}),\quad i=2,3,...,q, \tag{96}$$

in which $H(\cdot)$ is a Heaviside function and $\overline{S}_2(y)$ is the same as $\overline{S}_{12}(y)$.

The frequency equation is

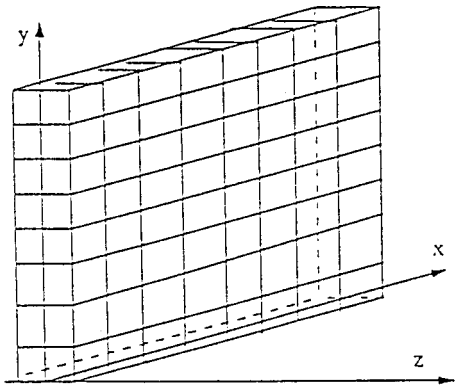
$$Y'_q(h)=0, \tag{97}$$

or

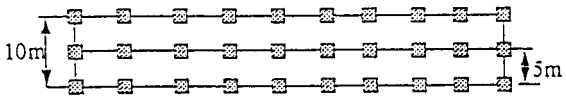
$$S'_2(H)+\sum_{j=1}^{n-1}\frac{1}{K_y(y_{j1})}(K_j-m_j\theta^2)Y_j(y_{j1})\overline{S}'_2(H-y_{j1})=0. \tag{98}$$

III. NUMERICAL EXAMPLE

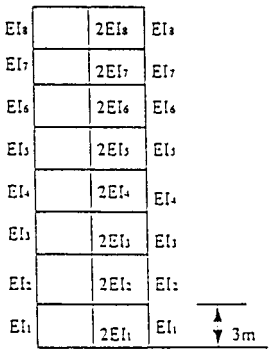
A ten-story frame-light panel building is shown in Fig. 4. The mass of each story is lumped to the corresponding floor. The masses in different stories are as follows:



(a) Perspective drawing



(b) Plane



(C) A transverse frame

FIG. 4. A narrow building.

$m_1 = 4.7820 \times 10^5$ kg;

$m_2 = m_3 = m_4 = m_5 = 4.5448 \times 10^5$ kg;

$m_6 = m_7 = m_8 = m_9 = 4.3194 \times 10^5$ kg;

$m_{10} = 4.1052 \times 10^5$ kg.

The values of stiffness of different floors in the x direction are

$GF_1 = 7.3487 \times 10^9$ N;

$GF_2 = GF_3 = GF_4 = GF_5 = 6.9842 \times 10^9$ N;

$GF_6 = GF_7 = GF_8 = GF_9 = 6.3786 \times 10^9$ N;

$GF_{10} = 6.3087 \times 10^9$ N.

The values of stiffness of columns in different stories are

$EI_1 = 4.5340 \times 10^7$ Nm^2;

$EI_2 = EI_3 = EI_4 = EI_5 = 4.3091 \times 10^7$ Nm^2;

$EI_6 = EI_7 = EI_8 = EI_9 = 4.0954 \times 10^7$ Nm^2;

$EI_{10} = 3.8923 \times 10^7$ Nm^2.

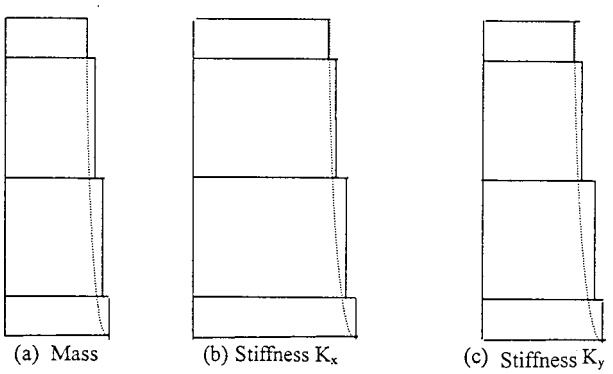


FIG. 5. The distribution of mass and stiffness. Note: The solid lines represent the real distributions and the dashed lines represent the assumed exponential distributions.

The story height is a constant, h=3.0 m. The distance between the transverse frame is 6 m. The foundation of the building can be treated as a line spring support; its stiffness is found as $K_{u0} = 7.8613 \times 10^{11}$ Nm^{-1}. The procedure for determining the natural frequencies and mode shapes of this narrow building is as follows:

(1) Determination of the mass intensity (mass per unit area). The mass intensity is equal to the mass of a story divided by the length of the building and the story height, i.e., the mass intensity of the first story is

$$\overline{m}_1 = \frac{4.7820 \times 10^5}{54 \times 3} = 2.9519 \times 10^3 \ \text{kg m}^{-2}.$$

The mass intensity in the range from the second story to the fifth story is

$$\overline{m}_2 = 2.8054 \times 10^3 \ \text{kg m}^{-2}.$$

The mass intensity in the range from the sixth story to the ninth story is

$$\overline{m}_3 = 2.6663 \times 10^3 \ \text{kg m}^{-2}.$$

For the tenth story, we have

$$\overline{m}_4 = 2.5341 \times 10^3 \ \text{kg m}^{-2}.$$

The four-step distribution of the mass intensity is shown in Fig. 5(a). In order to apply the methods proposed in this paper to the analysis of free vibration of this narrow building, the four-step model and the one-step model are considered, respectively, as follows:
The mass distribution of the one-step model can be described by

$$\overline{m}_{xy} = \overline{m} e^{-0.1526(y/H)}, \quad \overline{m} = 2.9519 \times 10^3. \tag{99}$$

The distribution described by Eq. (99) is denoted by the dashed line shown in Fig. 5(a).

(2) Evaluation of the shear stiffness K_x.
K_x is the value of GF divided by the story height.
For the first story

$$K_{1x} = \frac{7.3487 \times 10^9}{3} = 2.4496 \times 10^9 \ \text{N m}^{-1}.$$

For the range from the second story to the fifth story

$$K_{2x}=\frac{6.9842\times10^9}{3}=2.3281\times10^9 \ \ \mathrm{N\,m^{-1}}.$$

For the range from the sixth story to the ninth story

$$K_{3x}=\frac{6.3786\times10^9}{3}=2.1262\times10^9 \ \ \mathrm{N\,m^{-1}}.$$

For the tenth story

$$K_{4x}=\frac{6.3087\times10^9}{3}=2.1029\times10^9 \ \ \mathrm{N\,m^{-1}}.$$

The four-step distribution of K_x is shown in Fig. 5(b). The stepped distribution of K_x is also represented by a continuous curve. The continuous distribution of K_x is described reasonably well by the following exponential function:

$$K_x=K_2e^{-0.1526(y/H)} \quad K_2=2.4496\times10^9. \tag{100}$$

(3) Evaluation of shear stiffness K_y.
Since the total number of transverse frames is ten, the transverse stiffness of the first story is

$$D_{1y}=\left[\frac{12EI_1}{h^3}\times2+\frac{12(2EI_1)}{h^3}\right]\times10$$

$$=8.060\times10^8 \ \ \mathrm{N\,m^{-1}}.$$

For the range from the second story to the fifth story

$$D_{2y}=7.660\times10^8 \ \ \mathrm{N\,m^{-1}}.$$

For the range from the sixth story to the ninth story

$$D_{3y}=7.280\times10^8 \ \ \mathrm{N\,m^{-1}}.$$

For the tenth story

$$D_{4y}=6.919\times10^8 \ \ \mathrm{N\,m^{-1}},$$

in which D_{iy} $(i=1,2,3,4)$ is the total shear force of the ith story caused by a unit story displacement.
K_{iy} $(i=1,2,3,4)$ is the value of $D_{iy}h$ divided by the length of the building

$$K_{1y}=\frac{8.06\times3}{54}\times10^8=4.478\times10^7 \ \ \mathrm{N\,m^{-1}},$$

$$K_{2y}=4.256\times10^7 \ \ \mathrm{N\,m^{-1}},$$

$$K_{3y}=4.045\times10^7 \ \ \mathrm{N\,m^{-1}},$$

$$K_{4y}=3.844\times10^7 \ \ \mathrm{N\,m^{-1}}.$$

The four-step distribution of K_y is shown in Fig. 5(c). The stepped distribution of K_y is also represented by a continuous curve, described by

$$K_y=4.478\times10^7e^{-0.1526(y/H)} \ \ \mathrm{N\,m^{-1}}.$$

Considering the effect of light panels on the structural stiffness, K_y should be increased by 20% (Li, 1985; Li *et al.*, 1994), so, it is changed to

$$K_{yw}=K_1e^{-0.1526(y/H)}, \quad K_1=5.374\times10^7. \tag{101}$$

(4) Determination of $P(\zeta)$.
From Eq. (27), we have

$$P(\zeta)=\overline{m}_{xy}K_{yw}. \tag{102}$$

Substituting Eqs. (99) and (101) into Eq. (102) leads to

$$P(\zeta)=\overline{m}K_1e^{-0.3052(y/H)}. \tag{103}$$

Using Eq. (29) yields

$$\zeta=\frac{H}{K_1}e^{0.1526(y/H)}. \tag{104}$$

Comparing Eq. (103) with Eq. (104) gives

$$P(\zeta)=(b\zeta)^{-2}, \quad b=(mK_1)^{1/2}\frac{K_1}{H}. \tag{105}$$

(5) Determination of the natural frequencies and mode shapes.
Equation (105) has the same form as Eq. (40), the case 2 discussed in the previous section; the special solutions are the same as those given in Eq. (36) if a is set to zero. The boundary conditions for this example are

$$Y'(0)=\frac{K_{u0}}{K(0)}Y(0), \quad Y'(H)=0. \tag{106}$$

The frequency equation can be established by using Eq. (106) and the complete solution expressed in Eq. (60) including the special solutions for this case as follows:

$$\left(\frac{\beta}{2}-\frac{K_{u0}H}{K(0)}\right)\left(\frac{\beta}{2}\sin\xi+\xi\cos\xi\right)=\xi\left(\frac{\beta}{2}\cos\xi-\xi\sin\xi\right), \tag{107}$$

where

$$\xi^2=\frac{H^2}{4}\left(\frac{4\overline{m}\theta^2}{K_1}-2.5874\times10^{-5}\right), \quad \beta=0.1526. \tag{108}$$

Solving Eq. (107) yields ξ_j; substituting ξ_j into Eq. (108) gives

$$\theta_1=7.092, \quad \theta_2=21.25, \quad \theta_3=34.43, \quad \theta_4=49.61.$$

Substituting ξ_j into the complete solution of mode shape of the shear plate in the y direction and using the boundary conditions yields

$$Y_j(y)=e^{(0.1526y/2H)}\left[\cos\xi_j\frac{y}{H}+\left(\frac{K_{u0}H}{\xi_jK(0)}-\frac{\beta}{2\xi_j}\right)\sin\xi_j\frac{y}{H}\right]. \tag{109}$$

Because the two opposite edges of this building in the longitudinal direction are free, the circular natural frequencies and mode shapes in the x direction of the plate are given by Eqs. (15) and (16). Using Eq. (15) gives the kth circular natural frequency as

$$\Omega_k=\frac{910.95(k-1)\pi}{54}.$$

Setting $k=1,2,3,4$, one obtains

$$\Omega_1=0, \quad \Omega_2=53.00, \quad \Omega_3=105.99, \quad \Omega_4=158.99,$$

TABLE I. The circular natural frequencies of the narrow building.

ω_{ij}	ω_{11}	ω_{21}	ω_{31}	ω_{41}	ω_{21}	ω_{22}	ω_{32}	ω_{42}
One-step model	7.09	21.25	34.43	49.61	53.47	57.10	63.20	72.60
Four-step model	7.10	21.26	34.45	49.64	53.49	57.13	63.23	72.64
Measured results	7.29	21.40	34.82	49.99				
ω_{ij}	ω_{13}	ω_{23}	ω_{33}	ω_{43}	ω_{14}	ω_{24}	ω_{34}	ω_{44}
One-step model	106.23	108.10	111.44	117.03	159.15	160.40	162.62	166.55
Four-step model	106.26	108.14	111.47	117.08	159.19	160.43	162.67	166.59

and the kth mode shape in the x direction as

$$X_k(x)=\cos\frac{(k-1)\pi x}{L}.$$

It is necessary to point out that the natural frequencies and mode shapes given above are not those of the shear plate. The circular natural frequencies and mode shapes of the shear plate representing the narrow building considered herein are given by

$$\omega_{jk}=\sqrt{\theta_j^2+\Omega_K^2},\quad Z_{jk}(x,y)=Y_j(y)X_k(x). \tag{110}$$

The natural frequencies ω_{jk} and mode shapes $Z_{jk}(x,y)$ determined by Eq. (110) are corresponding to the jth mode shape in the y direction and the kth mode shape in the x direction and are listed in Table I and shown in Fig. 6, respectively.

If the four-step model and the transfer matrix method are also adopted for free-vibration analysis of this building, then the frequency equation in the y direction can be established from Eq. (80) by setting $K_{uH}=m_0=m_H=0$ as follows:

$$T_{11}+K_{u0}T_{22}=0, \tag{111}$$

in which T_{11}, T_{22} are the elements of the matrix $[T]$

$$[T]=[T_1][T_2][T_3][T_4],$$

$$[T_i]=\begin{bmatrix}\sin\dfrac{\theta}{\alpha_i}y_{i1} & \cos\dfrac{\theta}{\alpha_i}y_{i1}\\ K_{iy}\dfrac{\theta}{\alpha_i}\cos\dfrac{\theta}{\alpha_i}y_{i1} & -K_{iy}\dfrac{\theta}{\alpha_i}\sin\dfrac{\theta}{\alpha_i}y_{i1}\end{bmatrix}\times\begin{bmatrix}\sin\dfrac{\theta}{\alpha_i}y_{i0} & \cos\dfrac{\theta}{\alpha_i}y_{i0}\\ K_{iy}\dfrac{\theta}{\alpha_i}\cos\dfrac{\theta}{\alpha_i}y_{i0} & -K_{iy}\dfrac{\theta}{\alpha_i}\sin\dfrac{\theta}{\alpha_i}y_{i0}\end{bmatrix}^{-1}, \tag{112}$$

$$\alpha_i=\sqrt{\frac{K_{iy}}{m}}. \tag{113}$$

By solving the frequency equation Eq. (111) and using Eq. (110), the natural frequencies are found and also are listed in Table I for comparison purposes. It can be seen from Table I that the calculated results by using the two models are almost identical.

The mode shapes of the shear plate in the y direction can be also determined by using the four-step model as follows:

For the ith step plate, we have

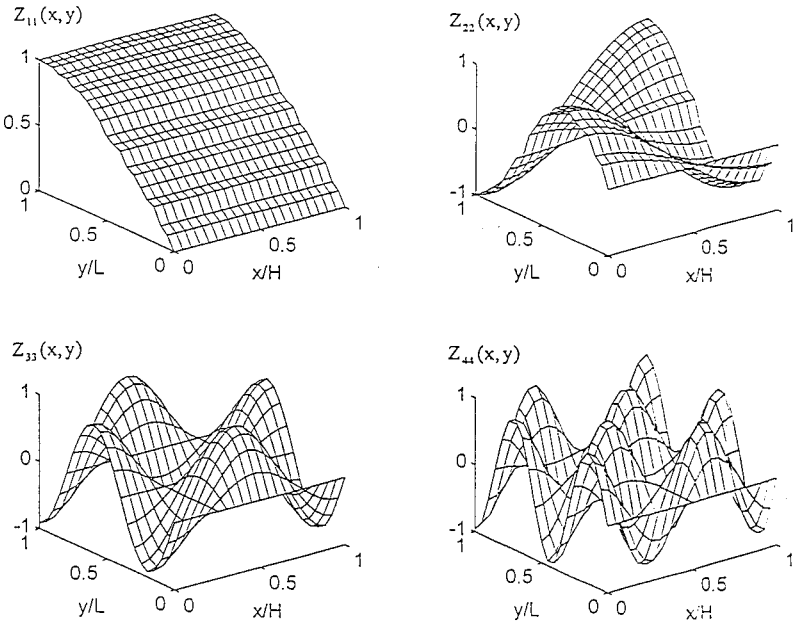


FIG. 6. The mode shapes of the narrow building.

$$\begin{bmatrix} Y_{i1} \\ Q_{i1} \end{bmatrix} = [T_i][T_{i-1}]\cdots[T_1]\begin{bmatrix} 1 \\ K_{u0} \end{bmatrix}, \quad (i=1,2,3,4). \quad (114)$$

Substituting θ_j in the expressions of $[T_i]$, Eq. (112), into Eq. (114) yields the jth ($j=1,2,...$) mode shape, which is almost the same as that obtained by using the one-step model.

We also calculated the natural frequencies and mode shapes using the four-step model and the recurrence method presented in this paper. It is found that all the calculated results are almost the same as those determined by the same model and the transfer matrix method.

The natural frequencies of this building measured by the author (Li *et al.*, 1994) are also listed in Table I for comparison purposes. It is clear that the calculated results are in good agreement with the measured data.

IV. CONCLUSIONS

In this paper, an exact approach to determine the natural frequencies and mode shapes of elastically restrained orthotropic shear plates with arbitrarily distributed mass or stiffness corresponding to several boundary conditions is proposed. It has been shown that a shear plate can be divided into two independent shear beams with the same boundary conditions as those of the shear plate in analyzing its free vibration. The jkth natural frequency of a shear plate is equal to the square root of the square sum of the jth natural frequency of a shear beam and the kth natural frequency of another shear beam. The jkth mode shape of the shear plate is the product of the jth mode shape of a shear beam and the kth mode shape of another shear beam. The exact solutions for free vibration of one-step shear plates with varying cross section are derived for eight cases. Since the function for describing the distribution of mass of a shear plate is an arbitrary one and the distribution of shear stiffness is expressed as a functional relation with the distribution of mass, and vice versa, an exact solution presented in this paper for each case actually represents a class of exact solutions for the title problem. The derived exact solutions of one-step shear plate are used to establish the frequency equation of a multistep orthotropic shear plate with spring supports and line masses by using the transfer matrix method and recurrence method. The numerical example shows that the calculated natural frequencies and mode shapes of a narrow building are in good agreement with the measured data, and the proposed procedure is an efficient and exact method.

Beiner, L., and Librescu, L. (**1984**). ''Minimum-weight design of an orthotropic shear panel with fixed flutter speed,'' AIAA J. **21**, 1015–1016.

Chopra, I. (**1974**). ''Vibration of stepped thickness plates,'' Int. J. Mech. Sci. **16**, 337–344.

Guo, S. J., Keane, A. J., and Mosherefi-Torbati, M. (**1997**). ''Vibration analysis of stepped thickness plates,'' J. Sound Vib. **204**, 645–657.

Jeary, A. P. (**1997**). *Designer's Guide to the Dynamic Response of Structures* (E & FN Spon, London).

Korqingskee, E. L. (**1952**). ''Vibrations of tall buildings,'' Moscow Press, pp. 25–36.

Li, G. Q. (**1985**). *Calculation Theory and Method of Seismic Structures* (Earthquake, Beijing), pp. 291–336.

Li, Q. S. (**1998**). ''Free vibration analysis of shear-type buildings,'' Adv. Struct. Eng. **2**(3), 163–172.

Li, Q. S. (**2000**). ''Exact solutions for free vibration of multi-step orthotropic shear plates,'' Struct. Eng. Mech. **9**(3), 269–288.

Li, Q. S., Cao, H., and Li, G. (**1994**). ''Analysis of free vibrations of tall buildings,'' J. Eng. Mech. **120**(9), 1861–1876.

Li, Q. S., Cao, H., and Li, G. (**1996**). ''Static and dynamic analysis of straight bars with variable cross section,'' Comput. Struct. **59**(6), 1185–1191.

Li, Q. S., Fang, J. Q., and Jeary, A. P. (**1999**). ''Vibration analysis of tall buildings with narrow rectangular plane configuration,'' Struct. Design Tall Buildings **7**(4), 307–322.

Wang, G. Y. (**1978**). *Vibration of Buildings and Structures* (Science, Beijing), pp. 168–178.

Mode isolation: A new algorithm for modal parameter identification

Michael V. Drexel and Jerry H. Ginsberg
The George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 29 June 2000; revised 14 May 2001; accepted 15 May 2001)

Multiple degree of freedom (MDOF) algorithms are the dominant methods for extracting modal parameters from measured data. These methods are founded on the notion that because the response of a linear dynamic system is the sum of many modal contributions, the extraction technique must deal with all of the modal parameters in a simultaneous fashion. The Mode Isolation Algorithm (MIA) described here is a frequency domain formulation that takes an alternative viewpoint. It extracts the modal parameters of each mode in an iterative search, and then refines the estimation of each mode by isolating its effect from the other modal contributions. The first iteration estimates modes in a hierarchy of their dominance. As each mode is estimated, its contribution is subtracted from the data set, until all that remains is noise. The second and subsequent iterations subtract the current estimates for all other modes to identify the properties of the mode under consideration. The various operations are described in detail, and then illustrated using data from a four-degree-of-freedom system that was previously used to assess the Eigensystem Realization Algorithm (ERA) and Enhanced ERA. Eigenvalues and mode shapes are compared for each algorithm. Another example analyzes simulated data for a cantilever beam with three suspended one-degree-of-freedom subsystems, in which the parameters are adjusted to bring two natural frequencies into close proximity. The results suggest that MIA is more accurate, and more robust in the treatment of noisy data, than either ERA version, and that it is able to identify modes whose bandwidth is comparable to the difference of adjacent natural frequencies. [DOI: 10.1121/1.1385902]

PACS numbers: 43.40.At, 43.20.Ks [CBB]

I. INTRODUCTION

Based on the assumption that measured data match the response of a linear mechanical system, experimental modal analysis seeks to identify the underlying modal structure. Such a response has a set of undetermined parameters that must be found by a ''best fit'' of the assumed solution to the data, which may be in either the frequency or time domain. A critical aspect of extracting modal parameters using multiple degree of freedom methods is identification of the number of modes contained in the system response. If the model order is assumed to be smaller than the true number of modes in the system, the number of modes found by the algorithm will be truncated and incorrect parameters will be ascribed to the identified modes. On the other hand, if the number of modes is assumed larger that the true number of modes, then there is the possibility of the creation of artificial modes or numerical instability.

Allemang *et al.*[1] provide an overview showing that many of the modal parameter extraction techniques share a common mathematical form. In each of the modal parameter extraction algorithms, it is difficult to determine if the choice of model order is an underestimation or an overestimation of the true number of modes. Many algorithms provide an error metric that compares the measured data to predicted values calculated from the assumed solution evaluated with the modal parameters identified by that algorithm. The magnitude of the error is then used to gauge the success of the extraction process. This, however, can be misleading, because it is possible for incorrect modal parameters to have a small error. In other words, a small error estimate merely implies that the modal parameters that were found correspond well to the measured data.

An example of the difficulties encountered in finding the number of modes contained in the system can be seen in the eigensystem realization algorithm (ERA) developed by Juang and Pappa.[2] This extraction strategy, which is derived from a control and systems identification viewpoint, has become a widely accepted approach. The method was initially developed as a time domain based algorithm; Juang and Suzuki[3] extended the ERA method to extract modal parameters from frequency domain test data. In ERA, because the model order is unknown, the number of modes is typically overestimated, and the irrelevant modes are truncated based on the information provided by a singular value decomposition. Doebling *et al.*[4] investigated the effect of overestimating the system order. They showed that the addition of data might not lead to an improved approximation because some modes might ''split'' into two modes as the model order is increased. In addition, the accuracy indicators showed high confidences in both the single and split modes. This means that the ERA error metrics would conclude that both solutions were correct. Thus, increasing model order might lead to multiple seemingly correct solutions, rather than convergence to a unique solution.

Other modal parameter extraction methods attempt to remedy this situation by utilizing peak counting in the fre-

quency domain to find the number of modes contained in the system. Systems in which the modal bandwidths are comparable to, or greater than, the separation between natural frequencies exhibit modal coupling, in which resonant peaks merge. This invalidates the use of peak counting as a way of identifying the number of natural frequencies contained in any frequency band. Assessment of the appropriate number of modes contained in the system is a primary source of error for any modal extraction algorithm.

Current MDOF methods are based on the assumption that the extraction technique must deal with all of the modal parameters in a simultaneous fashion. Although these methods recognize that the system's response is the sum of several terms, they do not exploit the possibility that a single mode may have characteristics that can be used to highlight that mode relative to the contribution of the other remaining modes; the possibility of doing so is mentioned by Ewins.[5] In contrast, the Mode Isolation Algorithm (MIA) proposes an iterative approach that highlights each mode in a hierarchal manner based on the dominance of the mode's contribution. As each mode is identified, its properties are used to isolate less dominant modes, until all that remains is noise, at which point a new iteration begins. The extraction procedure is halted when convergence criteria are met.

The presentation begins with the algorithm, and then the parameter identification scheme used to identify a specific mode. A four-degree-of-freedom system with moderately large damping, which previously was addressed with ERA and Enhanced ERA, provides the first of two numerical examples. Experimental error is simulated in the analytical data by the addition of Gaussian white noise. The first two iterative steps as the algorithm passes through the data are illustrated graphically. Comparison of results indicates that MIA is more accurate, especially for estimating modal damping ratios. The results also suggest that mode isolation is more robust with increasing noise levels. The second example is a lightly damped cantilever beam with suspended one-degree-of-freedom subsystems. The interesting feature of this example is that it illustrates the ability of MIA to identify modal properties when the difference of adjacent natural frequencies is comparable to the resonance bandwidths.

II. THE MODE ISOLATION ALGORITHM

Mode isolation has been formulated as a frequency domain technique. The data are taken to be a set of complex displacement amplitudes as a function of drive frequency, corresponding to application of a single harmonic generalized force, $Q_P = \mathrm{Re}\{F_P e^{i\omega t}\}$. The following will explain how the algorithm extracts the natural frequencies, modal damping ratios, and modal amplitudes, and the number of degrees of freedom associated with the data set, from data describing the frequency dependence of the Jth generalized coordinate.

The first step is to examine the complex frequency response (or acceleration) data Y_J for the selected generalized coordinate. If the system response consists of a superposition of undamped modal contributions, then

$$Y_J(\omega) = \sum_{k=1}^{N} \frac{A_{Jk}}{\omega_k^2 - \omega^2 + 2i\zeta_k\omega_k\omega}, \quad A_{Jk} = \Phi_{Jk}\Phi_{Pk}F_P. \tag{1}$$

The objective in the first iteration through the data is to obtain estimates for the natural frequencies ω_k, modal damping ratios ζ_k, and modal factors A_{Jk}, which are needed to initiate the second iteration. Subsequent iterations repeat the pattern of the second.

The first iteration begins by treating the overall $Y_J(\omega)$ data set as though it corresponded to a single-degree-of-freedom system. The parameter identification scheme described in the next section seeks evidence of the mode that gives the largest contribution to the overall $Y_J(\omega)$ data set. Such evidence can be the most visible peak exhibited by $|Y_J(\omega)|$ or the sharpest transition in the phase angle $\arg(Y_J(\omega))$. From that starting point, the parameter identification scheme estimates ω_1, ζ_1, A_{J1} for the most dominant mode. This step is much like standard single-degree-of-freedom methods, in that no consideration is given to the role of other modes. Note that the mode number k is assigned in the sequence in which modes are isolated, which is based on the dominance of their contributions within the processed data.

The second step in the first iteration is the place where mode isolation diverges from current techniques. In order to bring the next mode into dominance, the contribution of the first estimated mode to the frequency response function is subtracted from the measured data. Each further step repeats this pattern by subtracting an estimated mode from the data used to estimate that mode. To illustrate the general process for any step in the first iteration, let $Y_J^{(n)}(\omega)$ denote the residual of $Y_J(\omega)$ in the first iteration after the contributions of modes 1 to n have been subtracted. It follows that

$$\begin{aligned} Y_J^{(0)}(\omega) &= Y_J(\omega), \\ Y_J^{(n)}(\omega) &= Y_J^{(n-1)}(\omega) - \frac{A_{Jn}}{\omega_n^2 - \omega^2 + 2i\zeta_n\omega_n\omega}. \end{aligned} \tag{2}$$

At each step in the first iteration, estimates of ω_n, ζ_n and A_{Jn} are obtained from $Y_J^{(n-1)}(\omega)$ by the same procedures as those used to obtain preliminary estimates of the first mode's parameters.

The first iteration ends when there are no highlights in the residual frequency response functions. (This condition will be evident in the example that follows.) At this stage, estimates of ω_k, ζ_k and A_{Jk} have been obtained for each identified mode. The number of modes, N, described by the data is estimated as the number of steps required to reduce the residual $Y_J^N(\omega)$ to noise.

The second and following iterations are intended to isolate each mode and improve the estimates of the modal parameters. The concept is that any mode may be isolated by using the current estimates for the modal properties. These estimates are used to subtract the contribution of the other modes. Let $X_J^{(n)}(\omega)$ denote the $Y_J(\omega)$ data at a specified iteration step from which the estimated contributions of all except mode n have been subtracted:

$$X_J^{(n)}(\omega) = Y_J(\omega) - \sum_{k=1,k\neq n}^{N} \frac{A_{Jk}}{\omega_k^2 - \omega^2 + 2i\zeta_k\omega_k\omega}. \quad (3)$$

In this expression A_{Jk}, ω_k, and ζ_k are the current values, which are known from previous iterations and previous steps in the current iteration. The $X_J^{(n)}(\omega)$ data set is passed to the parameter identification scheme used in the first iteration. This yields improved estimates for ω_n, ζ_n and A_{Jn}. As was the case for the first iteration, the sequence in which modes are isolated is based on the current estimates for the modal amplitudes at resonance, which are $|A_{Jn}|/2\zeta_n$. An iteration progresses sequentially through the higher modes, until all modal parameters have been updated. Additional iterations repeat the steps of the second, until convergence criteria for the ω_n, ζ_n, and A_{Jn} parameters are met.

The second iteration provides an improved estimate of the value of the number of degrees of freedom. The iteration ends when the number of modes given by the current estimate for N have been isolated. Rather than immediately continuing to repeat the steps of the second iteration, the procedure goes on to form the last residual $X_J^{(N+1)}(\omega)$, from which the contributions of all estimated modal properties have been subtracted out. If this residual indicates that coherent data remains, then the value of N would need to be increased. In that case, the steps described by Eq. (2) may be implemented with $Y_J^{(N)}(\omega) = X_J^{(N+1)}(\omega)$ in order to obtain preliminary estimates for any additional modal properties.

III. PARAMETER IDENTIFICATION

Identification of ω_k and ζ_k values corresponding to the residual data is a crucial operation for each iterative stage. In the first iteration the data is the $Y_J^{(n)}$ successive data set obtained from subtracting lower modes, whereas the residuals for the second iteration are $X_J^{(n)}$, which are obtained by subtracting current estimates of the other modes. Let $\chi(\omega)$ denote whichever residual is under consideration. The complex displacement of the Jth generalized coordinate when ω is close to the natural frequency of the mode under consideration is

$$\chi(\omega) \approx \frac{A_{Jk}}{\omega_k^2 - \omega^2 + 2i\zeta_k\omega_k\omega}. \quad (4)$$

The parameter identification scheme uses values of $\chi(\omega)$ at several frequencies to find the values ω_k, ζ_k, and A_{Jk} that best fit the preceding description.

Because the approximate representation of $\chi(\omega)$ is to be matched to many data values, the solution of the overdetermined system requires a nonlinear least squares approach,[5,6] which leads to

$$\{D\} = [Z]\{\Delta P\}, \quad (5)$$

where $\{D\}$ is the error vector formed between the measured data and the assumed solution. The matrix $[Z]$ is the Jacobian of $Y_J^k(\omega)$ with respect to the variables ω_k, ζ_k, A_{jk}, and $\{\Delta P\}$ is a vector of the incremental variables ΔA_{jk}, $\Delta\omega_k$ and $\Delta\zeta_k$.

To determine $\{\Delta P\}$ Eq. (5) is premultiplied by $[Z]^T$, so that the coefficient matrix is square. This leads to

$$\{\Delta P\} = ([Z]^T[Z])^{-1}[Z]^T\{D\}. \quad (6)$$

(Gauss elimination, rather than a formal inversion, actually is the method by which $\{\Delta P\}$ is computed.) The incremental vector is used to update the current vector of modal parameters $\{P\}_m$, so the improved estimation of the modal parameters is

$$\{P\}_{m+1} = \{P\}_m + \{\Delta P\}. \quad (7)$$

The process iterates until convergence criteria are met.

The nonlinear least squares method for solving overdetermined systems requires that initial values for the modal parameters be obtained in order to begin the iteration process. The initial values for vector $\{P\}$ could be obtained by utilizing the peak-amplitude[7] or the Kennedy and Pancu[8] methods. The present work was performed by initiating ω_k as the value of ω where the imaginary part of the frequency response function reaches a global maximum,

$$\max(\mathrm{Im}(\chi(\omega))) \Rightarrow \omega_k. \quad (8)$$

The initial value of ζ_k is found by utilizing the definition of the half-power points ω_2 and ω_1 and the above approximation of ω_k, so that

$$\zeta_k = \frac{\omega_2 - \omega_1}{2\omega_k}. \quad (9)$$

Finally the initial estimation of A_{Jk} is

$$A_{jk} = 2\zeta_k\omega_k^2(\max(\mathrm{Im}(\chi(\omega)))). \quad (10)$$

The complete extraction of the modal parameters requires determination of the eigenvectors, as well as the eigenvalues. These values are derived from the parameters identified by MIA. The analysis returns to Eq. (1), which indicates that $A_{Jk} = \Phi_{Jk}\Phi_{Pk}F_P$. Because A_{Jk} and F_p are known, the eigenvector elements may be determined by setting $J = P$, which gives

$$\Phi_{Pk} = \left|\frac{A_{Pk}}{F_P}\right|^{1/2}. \quad (11)$$

This operation yields a positive value for Φ_{Pk}. The magnitude and sign of the other elements of the kth mode may be determined by

$$\Phi_{Jk} = \frac{A_{Jk}}{\Phi_{Pk}F_p}. \quad (12)$$

It follows from the development that consideration of a specific $Y_J(\omega)$ data set yields an estimate for all the natural frequencies and modal damping ratios, as well as row J of the normal mode matrix. The complex amplitudes for all generalized coordinates are only required to fill in the modal transformation matrix $[\Phi]$. In the first of the following examples the data set for each generalized coordinate was used to derive individual estimates for the ω_k and ζ_k values, which then were averaged. In contrast, the second example uses a composite data set obtained by summing the fre-

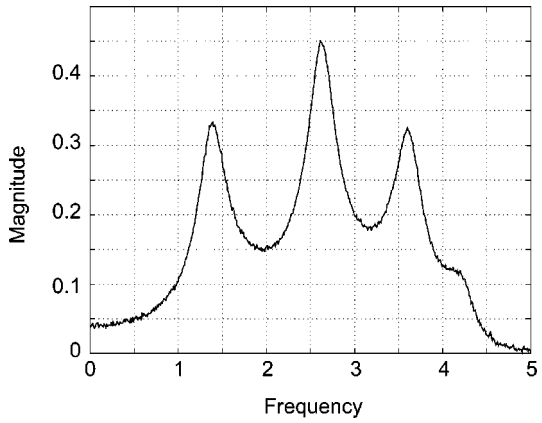


FIG. 1. Amplitude-frequency diagram of the original $Y_4(\omega)$ data.

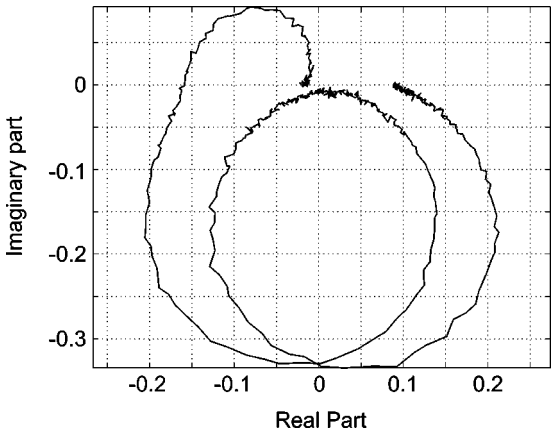


FIG. 3. First step in the first iteration yields $Y_4^{(1)}(\omega)$.

quency response functions. This leads to a single estimate for the ω_k and ζ_k values.

IV. NUMERICAL EXAMPLES OF MODE ISOLATION

Roemer and Mook[9,10] applied ERA and Enhanced ERA to a four-degree-of-freedom, discretized model of a cantilever wing, which originally was used by Meirovitch.[11] The mass, damping and stiffness matrices are given by

$$[M]=\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix},\quad [C]=\begin{bmatrix}0.3&0&0&0\\0&0.3&0&0\\0&0&0.3&0\\0&0&0&0.3\end{bmatrix},\tag{13}$$

$$[K]=\begin{bmatrix}10&-5&0&0\\-5&10&-5&0\\0&-5&10&-5\\0&0&-5&10\end{bmatrix}.$$

An exact solution of this system, which is proportionally damped, was found by solving the associated eigenproblem, and then analytically evaluating the modal response in the time domain. The time increment for the evaluation was $\Delta t=0.002$, which is 1% of the shortest natural period. To each response, a Gaussian white noise at approximately 10% of the signal amplitude was added. Frequency response functions were then obtained by FFT techniques. Contamination with noise adds to the parameter extraction difficulties associated with large (approximately 15%) damping ratios.

The performance of the algorithm will be illustrated by following the operations performed on the data for the fourth generalized coordinate at each stage in the first two iterations. Figure 1 shows the magnitude of $Y_4(\omega)$ as a function of drive frequency ω, after contamination with noise, while Fig. 2 is a Nyquist plot of the same data set. Three peaks are evident in the first figure, while the fourth, in the vicinity of $\omega=4$, is masked by the low modal participation factor and the added noise. The data set depicted by Figs. 1 and 2 is sent to the parameter identification scheme. Figure 3 shows $Y_4^{(1)}(\omega)$, which is the residual after the estimates for ω_1, ζ_1, and A_{41} have been used to subtract the first mode from $Y_4(\omega)$. In turn, $Y_4^{(1)}(\omega)$ is sent to the identification scheme in order to estimate the parameters of the next mode. Figure 4 depicts the residual after the estimated second mode has been subtracted from $Y_4^{(2)}(\omega)$, which is the residual after the two estimated modes have been subtracted. The first iteration

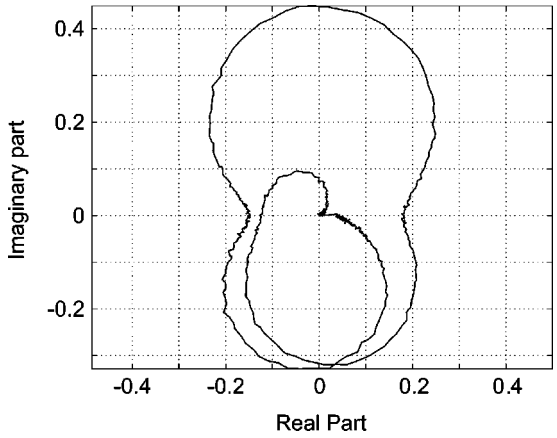


FIG. 2. Nyquist plot of the original $Y_4(\omega)$ data.

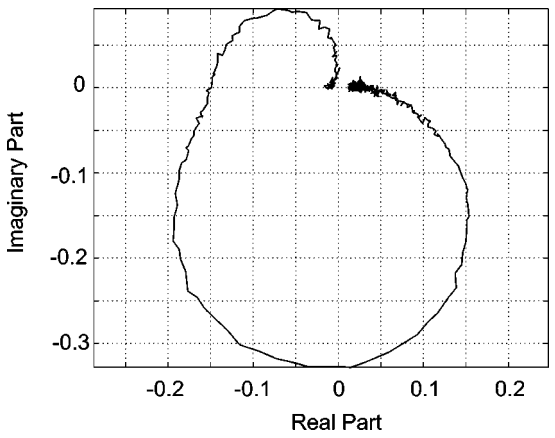


FIG. 4. Second step in the first iteration yields $Y_4^{(2)}(\omega)$.

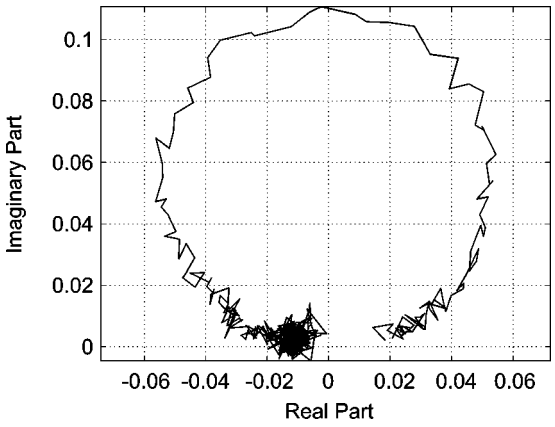


FIG. 5. Third step in the first iteration yields $Y_4^{(3)}(\omega)$.

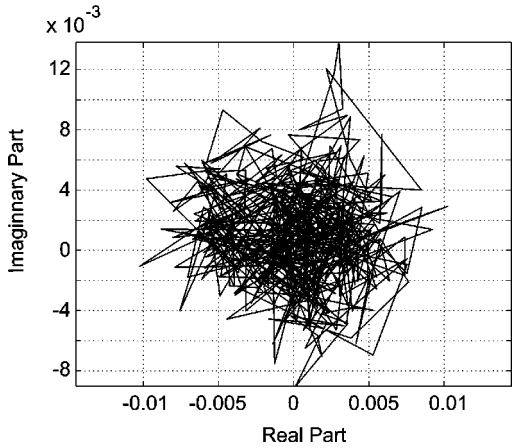


FIG. 6. Noise residual $Y_4^{(4)}(\omega)$ after the fourth step in the first iteration.

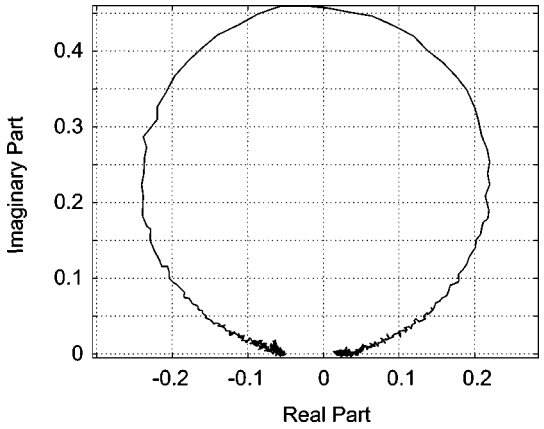


FIG. 7. Isolate the first mode: $X_4^{(1)}$ data resulting from the first step in the second iteration.

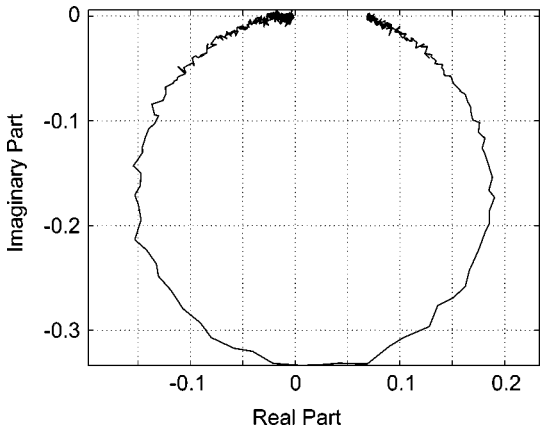


FIG. 8. Isolate the second mode: $X_4^{(2)}$ data resulting from the second step in the second iteration.

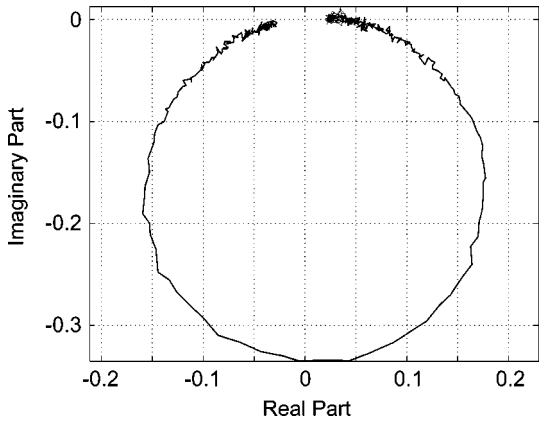


FIG. 9. Isolate the third mode: $X_4^{(3)}$ data resulting from the third step in the second iteration.

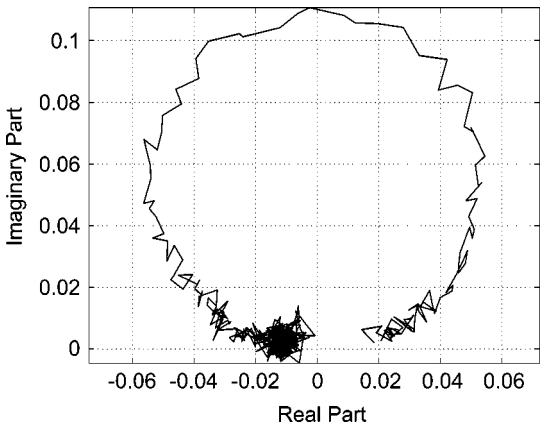


FIG. 10. Isolate the fourth mode: $X_4^{(4)}$ data resulting from the fourth step in the second iteration.

TABLE I. Comparison of ERA and Enhanced ERA to mode isolation eigenvalues for a four-degree-of-freedom system.

Exact solution	ERA	Enhanced ERA	Mode isolation (10%)	Mode isolation (20%)
-0.150 ± 1.374i	-0.176 ± 1.372i	-0.161 ± 1.387i	-0.151 ± 1.374i	-0.150 ± 1.375i
-0.150 ± 2.624i	-0.171 ± 2.663i	-0.176 ± 2.636i	-0.150 ± 2.624i	-0.151 ± 2.624i
-0.150 ± 3.615i	-15.93 ± 0.000i	-0.281 ± 3.795i	-0.150 ± 3.615i	-0.148 ± 3.616i
-0.150 ± 4.250i	-0.464 ± 4.039i	-0.299 ± 4.324i	-0.150 ± 4.251i	-0.155 ± 4.250i

continues to estimate the third mode's parameters, leading to the residual in Fig. 5. After subtraction of the estimated fourth mode, the state in Fig. 6 is attained. At this stage, only noise remains, indicating that no new modes can be estimated.

The first iteration has found an estimate for four modes. This estimation, however, has not addressed the simultaneous contribution of the modes. The second iteration addresses this matter. It begins by subtracting the second and higher modal displacements identified in the first iteration from the measured data according to Eq. (3). The residual that results is $X_4^{(1)}(\omega)$, shown in Fig. 7. Application of the nonlinear least squares procedure to this data yields a new estimate of ω_1, ζ_1, and A_{14}. These values and the current values for the other modes are utilized to form the residual $X_4^{(2)}(\omega)$ described by Eq. (3) and depicted in Fig. 8. The algorithm continues to isolate each mode in the manner described above. Figures 9 and 10 show the residuals $X_4^{(3)}(\omega)$ and $X_4^{(4)}(\omega)$, used to identify the third and fourth modes, respectively. The residual $X_4^{(5)}(\omega)$ is obtained by subtracting the four estimated modes from $Y_4(\omega)$. This residual, which is not shown, shows no discernible pattern. Additional passes through the data, in the manner of the second iteration, continue until the incremental variables, $\Delta\omega_k$, $\Delta\zeta_k$, and ΔA_{jk}, in the nonlinear least squares algorithm, lead to less than 0.1% change in the each value.

After convergent estimates for the ω_k, ζ_k, and A_{4k} parameters were obtained from $Y_4(\omega)$, the frequency response data $Y_1(\omega)$, $Y_2(\omega)$, and $Y_3(\omega)$ corresponding to the other generalized coordinates were subjected to the same treatment. Averaging the parameters identified from each of the five data sets yielded the final estimates of ω_k and ζ_k. In addition, the estimates of A_{Jk} extracted from each frequency response function were used directly to compute the elements of the eigenvectors. Table I shows the average eigenvalues obtained from ten individual simulations using different seeds for the random number generator for noise. The analytical values were obtained by performing state–space modal analysis on the system matrices, which is valid for arbitrary damping.[12] In terms of the properties of undamped modes, the magnitude of a complex eigenvalue's imaginary part is the damped natural frequency, while the real part is the negative product of the modal damping ratio and the natural frequency; the absolute value of the latter term is one half the bandwidth of a single-degree-of-freedom oscillator. Examination of the tabulation shows that ERA found four modes. However, one of the eigenvalues has only a real part, corresponding to an overdamped mode. Another eigenvalue has over 100% relative error in the estimation of the real part. The Enhanced ERA algorithm is able to find four meaningful modes from the data. The damped natural frequency reflects the improvement of the Enhanced ERA algorithm over the standard implementation of ERA. All of the damped natural frequencies are meaningful and have relative errors below 5%. However, the relative error in the real parts of two of the eigenvalues still approaches 100%.

TABLE II. Mode shape comparison of ERA, Enhanced ERA and Mode isolation.

Mode number	Exact solution	ERA	Enhanced ERA	Mode isolation (10%)	Mode isolation (20%)
	0.3717	0.4012	0.3705	0.3715	0.3681
Mode 1	0.6015	0.5743	0.6098	0.5979	0.6027
	0.6015	0.5912	0.6005	0.6004	0.6022
	0.3717	0.3992	0.3891	0.3782	0.3666
	-0.6015	-0.5945	-0.5685	-0.6021	-0.6024
Mode 2	-0.3717	-0.3622	-0.3810	-0.3696	-0.3827
	0.3717	0.4039	0.3711	0.3675	0.3695
	0.6015	0.6069	0.6395	0.6075	0.5936
	0.6015	0.2087	0.5645	0.6000	0.5992
Mode 3	-0.3717	0.2532	-0.4193	-0.3641	-0.3846
	-0.3717	0.1717	-0.3496	-0.3739	-0.3962
	0.6015	0.3515	0.6312	0.6082	0.6047
	-0.3717	-0.0751	-0.3223	-0.3773	-0.3811
Mode 4	0.6015	0.4508	0.6001	0.5930	0.5881
	-0.6015	-0.7844	-0.7101	-0.5922	-0.6008
	0.3717	0.6279	0.4541	0.3667	0.3408

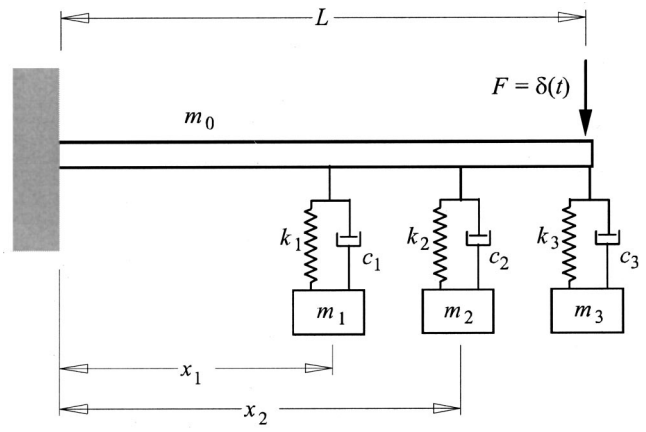


FIG. 11. A cantilever beam with suspended blocks.

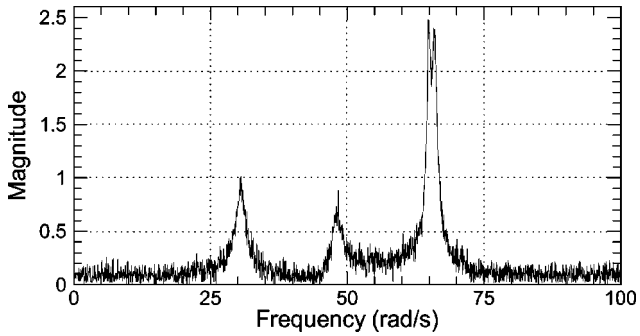


FIG. 12. Frequency response function for displacement at $x=L$ on the cantilever beam with attachments, with 15% white noise added.

Table I shows that mode isolation is more accurate than Enhanced ERA with respect to extraction of the damped natural frequencies, and it is substantially better in obtaining damping ratios. As a further illustration of the merits of mode isolation, the last column of Table I shows the mode isolation algorithm's performance when the white noise is raised to approximately 20% of the signal amplitude. The mode isolation method still is more accurate than Enhanced ERA, even though the parameters were extracted from a signal containing twice the noise of that for Enhanced ERA. Table II compares the extracted eigenvectors from the simulated data for the ERA, Enhanced ERA, and mode isolation algorithms. The ERA algorithm fails to extract the third mode correctly. The exact eigenvector is characterized by positive and negative values, whereas the extracted eigenvector reports only positive values. This is not a surprising outcome because of the failure of the ERA algorithm to extract four meaningful eigenvalues. As was the case for eigenvalues, Enhanced ERA shows improvement over ERA by providing four meaningful eigenvectors. Calculation of an average relative error between Enhanced ERA's extracted values and the exact values reveals an error of 6.5%. The mode isolation algorithm yields better results than Enhanced ERA for the case where the simulated data set is corrupted with 10% white noise, with an average relative error of 0.95%. The last column of Table II shows the mode isolation algorithm's performance when the white noise is raised to 20% of the signal amplitude. The relative error is 2.0%, which is much less that what Enhanced ERA obtained at half this noise level.

A cantilever beam with three spring-mass-damper systems suspended at $x_1=L/2$, $x_2=3L/4$, $x_3=L$, as shown in Fig. 11, represents another, and more difficult test of MIA, because the properties are adjusted to bring two natural frequencies into close proximity. The parameters of the attached systems are defined such that the stiffnesses are equal, the fixed-base natural frequencies increase linearly, and the fixed-base critical damping ratios are constant. Thus, the jth spring-mass-damper system is described by $k_j=\sigma m_0\Omega_0^2$, $\Omega_j/\Omega_0=[0.4663,0.7350,1.0025]$, $m_j=k_j/\Omega_j^2$ and $c_j=2\zeta(k_jm_j)^{1/2}$, where $\sigma=0.0001$, $\zeta=0.15$, and $\Omega_0=65.3594$ rad/s is the fundamental natural frequency of the cantilever beam without the three oscillators. Additional parameters required to describe the beam are: modulus of elasticity $E=2.07\times10^{11}$ Pa, density $\rho=7800\,\mathrm{kg/m^2}$, second moment of area $I=3.125\times10^{-7}\,\mathrm{m^4}$, cross sectional area $A=0.0015\,\mathrm{m^2}$, and beam length $L=2$ m.

A Ritz series whose basis functions are the normalized modes of the isolated cantilever beam was used to model the system. Although a ten term series was used for the simulation, the manner in which the system parameters are set is such that the first four natural frequencies are closely spaced, with significant modal motion for the beam and suspended masses. The higher natural frequencies are essentially those of the cantilever beam without attachments. The simulation was carried out by evaluating the temporal response at the tip of the beam, and of the suspended masses, to a unit impulsive force at the beam's tip. White noise at the appropriate level is added to the temporal response data for each response variable. FFT processing yielded frequency response functions for displacement at $x=L$ and for each block, for which the frequency increment is 0.5 rad/s. Figure 12 displays the frequency response function at the free end of the beam. A composite data set, obtained by summing the frequency response functions at the tip and for each of the suspended blocks up to a cutoff frequency of 100 rad/s, was used as input to MIA. The trial was limited to identification of the natural frequencies and modal damping ratios.

Table III compares the analytical complex eigenvalues to those obtained from MIA with either 15% or 45% white noise added to the simulated responses. At the 15% white noise level, the maximum error in the imaginary part of an eigenvalue, which represents the damped natural frequency, is less than 0.4%. The identified and analytical real parts, which are the product of the classical modal damping ratio and the undamped natural frequency, show noticeable differences. The errors are approximately 90% for the second mode, 23% for the fourth, and 7% for the first and third modes. The values extracted from the data contaminated by

TABLE III. Eigenvalues of the cantilever beam with attachments according to the mode isolation algorithm.

Exact solution	Mode isolation (15%)	Mode isolation (45%)
-0.750 ± 30.545i	-0.7618 ± 30.6102i	-0.5739 ± 30.6096i
-0.749 ± 48.024i	-1.4039 ± 48.2101i	-1.4086 ± 52.4717i
-0.312 ± 64.909i	-0.2916 ± 64.8429i	-0.3676 ± 63.1200i
-0.439 ± 65.990i	-0.5413 ± 65.7610i	-0.5085 ± 65.7884i

45% white noise show greater discrepancies. Nevertheless, the ability of the MIA to provide meaningful information for the two modes in the vicinity of 65 rad/s is noteworthy. Recall in this regard that the absolute value of the real part of an eigenvalue is half the single-degree-of-freedom bandwidth. This bandwidth is slightly smaller than the separation of the third and fourth natural frequencies.

V. SUMMARY AND CONCLUSIONS

MIA does not require *a priori* knowledge of the number of significant modes contained in the data. This relieves the user from guessing the system's degrees of freedom or, in the case of ERA and its descendants, from specifying the dimensions of the Hankel matrix. The benefit of the mode isolation's recursive search strategy was presented in the first of two numerical examples, in which the ERA algorithm was unable to detect four meaningful modes because of the presence of noise in the data. This implies that the singular value decomposition, on which ERA depends, produced several nonzero values instead of the four that correspond to actual modes. This caused the algorithm to produce artificial modes and large errors in the other modes. The Enhanced ERA algorithm showed an improvement over ERA in regard to the ability to detect all four modes, and to provide reasonable estimates of the damped natural frequencies and eigenvectors. However, Enhanced ERA still displayed large errors in the extraction of the real part of the eigenvalues. MIA was able to detect all four modes, to achieve excellent estimations of all natural frequencies, damping ratios, and eigenvectors, and to do so in an automated fashion. This suggests that the method is more accurate than ERA and Enhanced ERA. Its merit became even more evident when the noise level was doubled. The modal properties identified by MIA in that case were more accurate than those of either ERA version at the original noise level.

Although the modal damping ratios in the first example were somewhat large, the modal coordinates were not coupled by damping because the damping was proportional. This might seem to simplify the identification task, although the ERA efforts did encounter difficulty. The second example was a nonproportionally damped system. The parameters of this system were adjusted to have two natural frequencies whose separation is less than the modal bandwidth. The insertion of white noise at 15% and 45% of the overall amplitude did not inhibit the ability of MIA to identify all natural frequencies in the band of interest, although the estimated modal damping ratios for two modes did show significant discrepancies. It should be mentioned that in the latter example MIA showed an unexpected behavior when it operated on the simulated result with 45% white noise. Proper behavior would have been for the algorithm to recognize automatically that the residual produced by subtraction of all the identified modal contributions is incoherent. For some seed values in the 45% noise case, it was necessary to halt manually the second iteration because the incoherence was not recognized. However, this is merely a failure of the programing logic used to identify lack of coherence, rather than failure of MIA.

The primary objective of the present work is to illustrate mode isolation. MIA does not require a specific modal structure. An undamped modal formulation was used here because of its familiarity and accessibility. Current work is adapting the algorithm to operate in conjunction with a state space damped modal formulation,[12] which is valid for arbitrary damping ratios. Results for the second prototypical system have been obtained with this alternative analytical framework.[13] Discrepancies between identified and actual natural frequencies and damping ratios resulting from that analysis were an order of magnitude lower than those reported here. An important aspect of the algorithm is the simplicity and versatility of its approach. Changing the modal structure only requires that the nonlinear least squares parameter identification be altered, commensurate with the new modal structure. The iterative steps presented here remain the same.

ACKNOWLEDGMENTS

The authors acknowledge the support of the National Science Foundation under Grant No. CMS-0098539.

[1] R. J. Allemang, D. L. Brown, and W. Fladung, ''Modal Parameter Estimations: A Unified Matrix Polynomial Approach,'' Proceedings of the 12th International Modal Analysis Conference, Honolulu, HI, pp. 501–513 (1994).
[2] J. N. Juang and R. S. Pappa, ''An eigensystem realization algorithm for modal parameter identification and model reduction,'' J. Guid. Control Dyn. **8**, 620–627 (1985).
[3] J. N. Juang and H. Suzuki, ''An eigensystem realization algorithm in frequency domain for modal parameter identification,'' J. of Vibration, Stress, and Reliability in Design **110**, 24–29 (1988).
[4] S. W. Doebling, K. F. Alvin, and L. D. Peterson, ''Limitations of State-Space System Identification Algorithms for Structures with High Modal Density,'' Proceedings of the 12th International Modal Analysis Conference, Honolulu, HI, pp. 633–637 (1994).
[5] D. J. Ewins, *Modal Testing: Theory and Practice* (Research Studies Press, Wiley, New York, 1984).
[6] N. R. Draper and H. Smith, *Applied Regression Analysis* (Wiley, New York, 1998).
[7] C. W. de Silva, *Dynamic Testing and Seismic Qualification Practice* (Heath, Lexington, KY, 1983).
[8] C. C. Kennedy and C. D. P. Pancu, ''Use of vectors in vibration measurement and analysis,'' J. Aeronaut. Sci. **14**, 603–625 (1947).
[9] M. J. Roemer and D. J. Mook, ''An Enhanced Mode Shape Identification Algorithm,'' AIAA Paper 89–1245, Proceedings of the 30th AIAA SDM Conference, pp. 753–762 (1989).
[10] M. J. Roemer and D. J. Mook, ''Enhanced realization/identification of physical modes,'' J. of Aerospace Engineering **3**, 122–136 (1990).
[11] L. Meirovitch, *Elements of Vibrational Analysis* (McGraw-Hill, New York, 1975).
[12] J. H. Ginsberg, *Mechanical and Structural Vibrations* (Wiley, New York, 2001).
[13] M. V. Drexel and J. H. Ginsberg, ''Modal Parameter Isolation Using State Space Mode Isolation,'' Procedings of the 19th International Modal Analysis Conference, Orlando, FL, February 5–8, 2001.

An efficient formulation of Krylov's prediction model for train induced vibrations based on the dynamic reciprocity theorem

Geert Degrande[a] and Geert Lombaert
K.U. Leuven, Department of Civil Engineering, Kasteelpark Arenberg 40, B-3001 Leuven, Belgium

(Received 29 September 2000; accepted for publication 22 May 2001)

In Krylov's analytical prediction model, the free field vibration response during the passage of a train is written as the superposition of the effect of all sleeper forces, using Lamb's approximate solution for the Green's function of a halfspace. When this formulation is extended with the Green's functions of a layered soil, considerable computational effort is required if these Green's functions are needed in a wide range of source–receiver distances and frequencies. It is demonstrated in this paper how the free field response can alternatively be computed, using the dynamic reciprocity theorem, applied to moving loads. The formulation is based on the response of the soil due to the moving load distribution for a single axle load. The equations are written in the wave-number-frequency domain, accounting for the invariance of the geometry in the direction of the track. The approach allows for a very efficient calculation of the free field vibration response, distinguishing the quasistatic contribution from the effect of the sleeper passage frequency and its higher harmonics. The methodology is validated by means of *in situ* vibration measurements during the passage of a Thalys high-speed train on the track between Brussels and Paris. It is shown that the model has good predictive capabilities in the near field at low and high frequencies, but underestimates the response in the midfrequency band.
[DOI: 10.1121/1.1388002]

PACS numbers: 43.40.At, 43.50.Lj, 43.20.Bi [PJR]

I. INTRODUCTION

Train induced vibrations are an environmental concern, as waves propagate through the soil and interact with nearby buildings, where they may cause malfunctioning of sensitive equipment and discomfort to people, as well as re-radiated noise in properties. The rapid extension of the high speed rail network throughout Europe has initiated a lot of research on analytical and numerical prediction models for train induced vibrations.

These models differ in several aspects, such as the excitation mechanisms included, the soil model used (halfspace or layered halfspace) and the consideration of dynamic interaction between the track and the subsoil, as well as through-soil coupling of the sleepers and the ballast. The following excitation mechanisms are distinguished: (1) quasistatic excitation due to moving axle loads, (2) wheel and rail roughness, (3) transient excitation due to rail joints, switches, and wheelflats, and (4) parametric excitation due to the discrete support of a track by sleepers, resulting in a spatial variation of the stiffness experienced by the axles.

The quasistatic excitation is important in the near field and when the train speed is close to or larger than the critical phase velocity of the coupled track–soil system, a condition that can be met for high speed trains running on very soft soils. This results in high vibration levels and high displacements of the track, affecting track stability and safety.[1]

The response of a semi-infinite soil medium due to moving loads has received considerable attention in the literature. The steady-state response of a layered visco-elastic halfspace, subjected to a moving load with a constant amplitude, has been studied by de Barros and Luco.[2] Grundmann *et al.*[3] have used the wavelet transform[4] to compute the response of a layered halfspace due to moving loads.

Aubry *et al.*[5] have considered the dynamic interaction between a beam with a rigid cross section and a layered halfspace under the action of a moving load. A substructure approach is used in the wave-number-frequency domain, exploiting the invariance of the geometry in the longitudinal direction. The formulation has been applied and validated by *in situ* experiments for the case of road traffic induced vibrations,[6,7] as well as extended to include the interaction between the moving vehicle and the coupled track–soil system.[8] Sheng *et al.* have coupled an infinite layered beam model for the track to a layered model of the soil, using Haskell–Thomson transfer matrices, considering the case of fixed[9] and moving[10] harmonic point sources. Kaynia *et al.*[11] and Madshus and Kaynia[12] have coupled a beam model for the track and the embankment with the dynamic impedance of a layered soil, modeled with dynamic stiffness matrices.[13] The model accounts for quasistatic excitation and aims to study the response of the track and the subsoil at train speeds below and beyond the critical phase velocity of the track–soil system.

Knothe and Wu[14] have studied the receptance of a railway track and the subgrade, accounting for dynamic soil-structure effects and through-soil coupling of the sleepers. Similar developments have been presented by Van den Broeck and De Roeck,[15] who have, apart from quasistatic

[a]Electronic mail: Geert.Degrande@bwk.kuleuven.ac.be

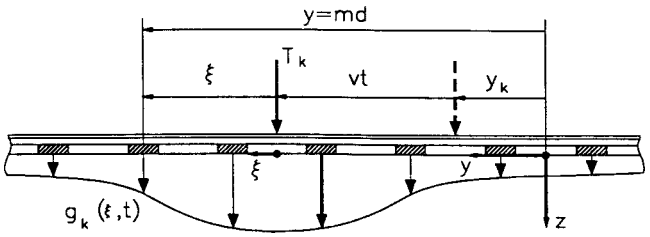


FIG. 1. Vertical deflection curve of the track.

and parametric excitation, also included the roughness excitation.

The present paper will concentrate on an analytical prediction model that has been presented by Krylov.[16–20] The model only includes quasistatic excitation and aims to predict ground vibrations at train speeds below and above the critical phase velocity of the track–soil system. Other excitation mechanisms as well as dynamic soil-structure effects and through-soil coupling of the sleepers are not accounted for.

The quasistatic force transmitted by a sleeper is derived from the deflection curve of the track, modeled as a beam on an elastic foundation. The free field response is calculated in the frequency domain as the superposition of all sleeper forces; wave propagation through the soil is represented by Lamb's approximate solution for the Green's function of a halfspace, accounting for the contribution of the surface wave only. This formulation can be easily extended to incorporate the Green's functions of a layered halfspace.[21,22] Depending on the train speed and length, the dynamic soil characteristics, and the frequency range envisaged, a considerable computational effort is needed, however, to compute the Green's functions for a wide range of source–receiver distances and frequencies.

The objective of this paper is to demonstrate how the free field response can be computed alternatively, relying on the dynamic reciprocity theorem of Betti–Rayleigh applied to moving loads, resulting in a very efficient computation method. Krylov's calculation method of the load, distributed by the track during the passage of a train, is briefly recapitulated; particular attention goes to the use of a uniform notation that can subsequently be used in both Krylov's formulation as well as the alternative formulation to compute the free field vibrations. The resulting formulation is finally validated by means of experimental results, obtained by *in situ* measurements during the passage of a Thalys high-speed train (HST) on the track Brussels–Paris.[21–23]

II. LOAD DISTRIBUTION DUE TO A TRAIN PASSAGE

The track is modeled as a beam with bending stiffness EI and mass m per unit length on an elastic foundation with subgrade stiffness k_s. It is assumed that the track is directed along the y direction, with the vertical z axis pointing downwards, and the horizontal x axis perpendicular to the (y,z) plane (Fig. 1). The train has K axles; the load and the initial position of the kth axle are denoted by T_k and y_k, respectively. The vertical deflection $w(y)$ of the track due to a single axle load T_k that moves with a speed v is governed by the following partial differential equation:

$$EI\frac{\partial^4 w}{\partial y^4}+m\frac{\partial^2 w}{\partial t^2}+k_s w=T_k\delta(y-y_k-vt). \tag{1}$$

A local coordinate $\xi=y-y_k-vt$ determines the position of a point y along the track with respect to the position y_k+vt of the axle load. The track deflection $w(\xi)$ is equal to:[20]

$$w(\xi)=\frac{T_k}{8EI\beta^3\delta}\exp(-\beta\delta|\xi|)\left(\cos\beta\eta\xi+\frac{\delta}{\eta}\sin\beta\eta|\xi|\right). \tag{2}$$

Herein, $\beta=(k_s/4EI)^{0.25}$, while $\delta=[1-(v/c_{\min})^2]^{0.5}$ and $\eta=[1+(v/c_{\min})^2]^{0.5}$, with $c_{\min}=(4k_sEI/m^2)^{0.25}$ the velocity of free track waves.

At time t, it is assumed that each sleeper m, located at $y=md$ (d is the sleeper distance) or $\xi=md-y_k-vt$ in the moving frame of reference, transfers a fraction of the axle load T_k proportional to its instantaneous deflection $w(\xi)$. The distribution of forces $g_k(\xi,t)$ can be written as the following summation:

$$g_k(\xi,t)=\frac{T_k}{N_{\text{eff}}^{\text{st}}}\frac{w(\xi)}{w_{\max}^{\text{st}}}\sum_{m=-\infty}^{\infty}\delta(\xi+y_k+vt-md). \tag{3}$$

The Dirac function assures that, at time t when the axle load T_k is located at y_k+vt, the distribution of forces in the moving frame of reference is transmitted at the position of the sleepers. $N_{\text{eff}}^{\text{st}}$ is the effective number of sleepers needed to support the axle load T_k, if all sleepers would take up a maximum load corresponding to the maximum quasistatic deflection $w_{\max}^{\text{st}}=T_k/(8EI\beta^3\delta)$ of the track. $N_{\text{eff}}^{\text{st}}$ is equal to $2y_0^{\text{st}}/\pi d$, with $y_0^{\text{st}}=\pi/\beta$ the effective quasistatic track deflection distance. The Fourier transform of the distribution $g_k(\xi,t)$ is equal to

$$\hat{g}_k(\xi,\omega)=\frac{1}{v}\frac{T_k}{N_{\text{eff}}^{\text{st}}}\frac{w(\xi)}{w_{\max}^{\text{st}}}\left[\frac{2\pi v}{d}\sum_{m=-\infty}^{\infty}\delta\left(\omega-m\frac{2\pi v}{d}\right)\right]\times\exp\left(i\omega\frac{\xi}{v}\right)\exp\left(i\omega\frac{y_k}{v}\right), \tag{4}$$

where the bracketed term is equal to the Fourier transform of a series of Dirac impulses, separated in time by d/v; it corresponds to a series of harmonics of the sleeper passage frequency $f_d=v/d$. The two last terms in Eq. (4) represent a phase shift.

The distribution of forces $f_z(\mathbf{x},t)$ transmitted by all sleepers due to the passage of a train with K axles is now written in the original fixed coordinate system as

$$f_z(\mathbf{x},t)=\int_{-\infty}^{\infty}\sum_{k=1}^{K}\delta(x)\,\delta(y-\xi-y_k-vt)\,\delta(z)g_k(\xi,t)\,d\xi. \tag{5}$$

The Dirac functions are introduced to express the coordinate ξ that moves with the source in terms of the fixed coordinate y, so that the resulting expression $f_z(\mathbf{x},t)$ depends on the fixed coordinates. The introduction of expression (3) for $g_k(\xi,t)$ into Eq. (5) results in

$$f_z(\mathbf{x},t)=\sum_{m=-\infty}^{\infty}\sum_{k=1}^{K}\delta(x)\,\delta(y-md)\times\delta(z)\,\frac{T_k}{N_{\text{eff}}^{\text{st}}}\,\frac{w(md-y_k-vt)}{w_{\max}^{\text{st}}}. \tag{6}$$

This expression is equivalent to the one proposed by Krylov; it is used in Sec. III where Krylov's computation method of the free field vibrations is briefly reviewed. In the subsequent derivation of an alternative formulation to compute the free field vibrations, based on the dynamic reciprocity theorem of Betti–Rayleigh applied to moving sources, expression (5) will be preferred to Eq. (6).

The Fourier transform of the distribution $f_z(\mathbf{x},t)$ is equal to

$$\hat{f}_z(\mathbf{x},\omega)=\hat{F}(\omega)\hat{C}(\omega)\sum_{m=-\infty}^{\infty}\delta(x)\,\delta(y-md)\,\delta(z)\times\exp\left(-i\omega\frac{md}{v}\right). \tag{7}$$

Denoting the point along the source line where the load is applied with a vector $\mathbf{x}^S=\{x^S,y^S,z^S\}^T$, where $x^S=0$, $y^S=md$ and $z^S=0$, this equation can be abbreviated as

$$\hat{f}_z(\mathbf{x},\omega)=\hat{F}(\omega)\hat{C}(\omega)\sum_{m=-\infty}^{\infty}\delta(\mathbf{x}-\mathbf{x}^S)\exp\left(-i\omega\frac{y^S}{v}\right). \tag{8}$$

The function $\hat{F}(\omega)$ represents the Fourier transform of the force transmitted by a single sleeper due to the passage of single unit axle load:

$$\hat{F}(\omega)=\frac{1}{N_{\text{eff}}^{\text{st}}}\frac{1}{\beta v}\left[\frac{\delta+\eta+\dfrac{\omega}{v\beta}}{\delta^2+\left(\eta+\dfrac{\omega}{v\beta}\right)^2}+\frac{\delta+\eta-\dfrac{\omega}{v\beta}}{\delta^2+\left(\eta-\dfrac{\omega}{v\beta}\right)^2}\right]. \tag{9}$$

Equation (9) demonstrates that $\hat{F}(\omega)$ depends on the characteristics of the track, the subgrade stiffness k_s, and the speed v of the train. As $1/(N_{\text{eff}}^{\text{st}}\beta)=d/2$, the force transmitted by a sleeper is proportional to the sleeper distance d. The quasi-static value $\hat{F}(\omega=0)$ is independent of β or k_s, while it decreases for increasing v. The frequency content $\hat{F}(\omega)$ increases for increasing k_s or β and increasing v. When v is much lower than $c_{\min}$, the beam inertial forces can be neglected and δ and η tend to 1, resulting in the original expression of Krylov and Ferguson[24] for low train speeds. It can be demonstrated that $\hat{F}(\omega)$ is also equal to

$$\hat{F}(\omega)=\frac{1}{vN_{\text{eff}}^{\text{st}}}\frac{\tilde{w}(k_\xi)}{w_{\max}^{\text{st}}}, \tag{10}$$

with $\tilde{w}(k_\xi)$ the forward Fourier transformation of $w(\xi)$ to the wave number domain $k_\xi=\omega/v$:

$$\tilde{w}(k_\xi)=\int_{-\infty}^{+\infty}w(\xi)\exp(+ik_\xi\xi)\,d\xi. \tag{11}$$

The function $\hat{C}(\omega)$ represents the composition of the train in the frequency domain:

$$\hat{C}(\omega)=\sum_{k=1}^{K}T_k\exp\left(i\omega\frac{y_k}{v}\right). \tag{12}$$

III. THE FREE FIELD RESPONSE: KRYLOV'S FORMULATION

The vertical free field displacement $\hat{u}_z(\mathbf{x}^R,\omega)$ in a receiver $\mathbf{x}^R$ is computed in the frequency domain as the integral along the source line of the product of the component $\hat{u}_{zz}^G(\mathbf{x}^R,\mathbf{x},\omega)$ of the Green's tensor $\hat{u}_{ij}^G(\mathbf{x}^R,\mathbf{x},\omega)$ and the force distribution $\hat{f}_z(\mathbf{x},\omega)$:

$$\hat{u}_z(\mathbf{x}^R,\omega)=\int_{-\infty}^{+\infty}\hat{u}_{zz}^G(\mathbf{x}^R,\mathbf{x},\omega)\hat{f}_z(\mathbf{x},\omega)\,d\mathbf{x}. \tag{13}$$

The Green's tensor $\hat{u}_{ij}^G(\mathbf{x}^R,\mathbf{x},\omega)$ represents the displacement vector at a point $\mathbf{x}$ in the direction $\mathbf{e}_j$ when a force is applied at $\mathbf{x}^R$ in the direction $\mathbf{e}_i$. Krylov uses Lamb's approximate solution for the Green's functions of a halfspace, which only accounts for the surface wave contribution.[25,26] The latter is replaced here by the Green's functions of a layered halfspace, which are calculated with a direct stiffness formulation in the wave-number-frequency domain.[13,27–29]

Equation (13) can be further elaborated, using expression (8) for $\hat{f}_z(\mathbf{x},\omega)$:

$$\hat{u}_z(\mathbf{x}^R,\omega)=\hat{F}(\omega)\hat{C}(\omega)\int_{-\infty}^{+\infty}\sum_{m=-\infty}^{+\infty}\hat{u}_{zz}^G(\mathbf{x}^R,\mathbf{x},\omega)\,\delta(\mathbf{x}-\mathbf{x}^S)\times\exp\left(-i\omega\frac{y^S}{v}\right)d\mathbf{x}. \tag{14}$$

Accounting for the discrete support of the sleepers, the integral along the source line disappears:

$$\hat{u}_z(\mathbf{x}^R,\omega)=\hat{F}(\omega)\hat{C}(\omega)\sum_{m=-\infty}^{+\infty}\hat{u}_{zz}^G(\mathbf{x}^S,\mathbf{x}^R,\omega)\exp\left(-i\omega\frac{y^S}{v}\right). \tag{15}$$

In Eq. (15), use is made of the reciprocity property $\hat{u}_{ij}^G(\mathbf{x}^R,\mathbf{x},\omega)=\hat{u}_{ji}^G(\mathbf{x},\mathbf{x}^R,\omega)$ to switch the position of the source and the receiver in the definition of the Green's tensor, so that $\hat{u}_{zz}^G(\mathbf{x}^R,\mathbf{x}^S,\omega)=\hat{u}_{zz}^G(\mathbf{x}^S,\mathbf{x}^R,\omega)$. As the force is applied in the vertical direction and the soil stratification is assumed to be horizontal, the resulting problem is axisymmetric. The source is located at the surface in $x^S=0$, $y^S=md$, and $z^S=0$, while it is assumed that the receiver is located at x^R, $y^R=0$ and z^R (Fig. 2). In a local cylindrical coordinate system with the origin in $\mathbf{x}^S$, the Green's function can be written as $\hat{u}_{zz}^{G\text{axi}}(r_m,z^R,\omega)$, where $r_m=\sqrt{(x^R)^2+(md)^2}$ is the radius, or the distance between source and receiver in the horizontal plane. The Green's function $\hat{u}_{zz}^{G\text{axi}}(r,z^R,\omega)$ is computed as the inverse Hankel transform of the solution $\tilde{u}_{zz}^{G\text{axi}}(k_r,z^R,\omega)$ in the radial wave-number-frequency domain, as explained in Appendix A.

As the Green's function is the same for sleepers at equal distance behind and ahead of a receiver, Eq. (15) becomes

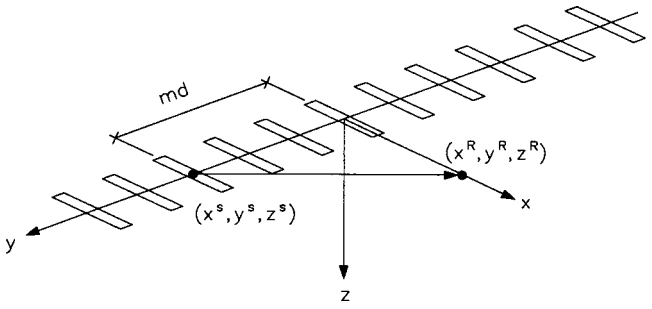


FIG. 2. Location of the source and the receiver.

$$\hat{u}_z(\mathbf{x}^R,\omega)=\hat{F}(\omega)\hat{C}(\omega)\left[\hat{u}_{zz}^{G\mathrm{axi}}(r_0,z^R,\omega)+\sum_{m=1}^{+\infty}2\cos\left(\omega\frac{md}{v}\right)\hat{u}_{zz}^{G\mathrm{axi}}(r_m,z^R,\omega)\right]. \quad (16)$$

When the bracketed term is denoted as $\hat{u}_{\delta z}(\mathbf{x}^R,\omega)$, Eq. (16) can be written in the following abbreviated form:

$$\hat{u}_z(\mathbf{x}^R,\omega)=\hat{F}(\omega)\hat{C}(\omega)\hat{u}_{\delta z}(\mathbf{x}^R,\omega), \quad (17)$$

which illustrates that the response $\hat{u}_z(\mathbf{x}^R,\omega)$ at a receiver $\mathbf{x}^R$ is governed by three terms.

(1) The first term $\hat{F}(\omega)$ represents the frequency content of the force transmitted by a single sleeper, when a unit axle load moves with a speed v along the track.

(2) The second term $\hat{C}(\omega)$ is determined by the composition and the speed of the train. The product $\hat{F}(\omega)\hat{C}(\omega)$ is the frequency content of the force, transmitted by a single sleeper, when a train moves with a speed v along the track.

(3) The third term $\hat{u}_{\delta z}(\mathbf{x}^R,\omega)$ represents the frequency content of the response at a receiver when each sleeper is consecutively loaded by a Dirac impulse, shifted in time by d/v. It involves a summation on an infinite number of sleepers, containing the Green's function $\hat{u}_{zz}^{G\mathrm{axi}}(r_m,z^R\omega)$. This sum can be truncated at a finite number N_s of sleepers, as the contribution for large source–receiver distances r_m is small due to the radiation and material damping in the soil. Depending on the train speed and length, the dynamic soil characteristics, and the frequency range envisaged, considerable computational effort is needed, however, to compute the Green's functions of a layered medium for a wide range of source–receiver distances and frequencies. It is therefore desirable to find an alternative for the term $\hat{u}_{\delta z}(\mathbf{x}^R,\omega)$, which is the objective of Sec. IV.

IV. THE FREE FIELD RESPONSE: ALTERNATIVE FORMULATION

The dynamic extension of the classical reciprocity theorem of Betti–Rayleigh of elastostatics is used to compute the free field vibrations due to a moving source distribution.[5,25,30] The free field vertical displacements $u_z(\mathbf{x}^R,t)$ are calculated as a convolution integral of the Green's function $u_{zz}^G(\mathbf{x}^R,\mathbf{x},t)$ and the load distribution $f_z(\mathbf{x},t)$ on the track:

$$u_z(\mathbf{x}^R,t)=\int_{-\infty}^{+\infty}\int_{-\infty}^{t}u_{zz}^G(\mathbf{x}^R,\mathbf{x},t-\tau)f_z(\mathbf{x},\tau)\,d\tau\,d\mathbf{x}. \quad (18)$$

Expression (5) is now preferred for the load distribution $f_z(\mathbf{x},t)$, as it explicitly incorporates Dirac functions that define the positions where the forces are applied in the global frame of reference and in function of the time t. The following expression is obtained:

$$u_z(\mathbf{x}^R,t)=\int_{-\infty}^{+\infty}\int_{-\infty}^{t}\int_{-\infty}^{+\infty}\sum_{k=1}^{K}u_{zz}^G(\mathbf{x}^R,\mathbf{x},t-\tau)\times\delta(x)\,\delta(y-\xi-y_k-v\tau)\,\delta(z)g_k(\xi,\tau)\,d\xi\,d\tau\,d\mathbf{x}. \quad (19)$$

Accounting for the Dirac functions and the symmetry of the Green's function with respect to source and receiver, Eq. (19) becomes:

$$u_z(\mathbf{x}^R,t)=\int_{-\infty}^{t}\int_{-\infty}^{+\infty}\sum_{k=1}^{K}u_{zz}^G(0,\xi+y_k+v\tau,0,\mathbf{x}^R,t-\tau)\times g_k(\xi,\tau)\,d\xi\,d\tau. \quad (20)$$

When the invariance of the geometry in the y direction is accounted for, Eq. (20) can be elaborated as

$$u_z(\mathbf{x}^{R'},y^R,t)=\int_{-\infty}^{t}\int_{-\infty}^{+\infty}\sum_{k=1}^{K}u_{zz}^G(\mathbf{x}^{R'},y^R-\xi-y_k-v\tau,t-\tau)\times g_k(\xi,\tau)\,d\xi\,d\tau, \quad (21)$$

where $\mathbf{x}^{R'}=\{x^R,z^R\}^T$ and it is assumed that the load is located in the origin of the coordinate system.

The representation of the vertical displacements in the wave-number-frequency domain is

$$\begin{aligned}\tilde{u}_z(\mathbf{x}^{R'},k_y,\omega)&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}u_z(\mathbf{x}^{R'},y^R,t)\exp(-i\omega t)\times\exp(+ik_y y^R)\,dt\,dy^R\\&=\tilde{u}_{zz}^G(\mathbf{x}^{R'},k_y,\omega)\int_{-\infty}^{+\infty}\sum_{k=1}^{K}\hat{g}_k(\xi,\omega-k_y v)\times\exp[ik_y(\xi+y_k)]\,d\xi. \quad (22)\end{aligned}$$

A frequency shift $k_y v$ is applied to the argument of the force distribution $\hat{g}_k(\xi,\omega-k_y v)$, where ω is the frequency at the receiver, while $\omega-k_y v$ corresponds to the frequency emitted at the source. The latter is denoted as $\tilde{\omega}$ in the following. The calculation of the Green's function $\tilde{u}_{zz}^G(\mathbf{x}^{R'},k_y,\omega)$ is explained in Appendix B. The vertical soil displacements in the frequency domain are found as the inverse Fourier transform of Eq. (22):

$$\begin{aligned}\hat{u}_z(\mathbf{x}^R,\omega)&=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\tilde{u}_z(\mathbf{x}^{R'},k_y,\omega)\exp(-ik_y y^R)\,dk_y\\&=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\tilde{u}_{zz}^G(\mathbf{x}^{R'},k_y,\omega)\int_{-\infty}^{+\infty}\sum_{k=1}^{K}\hat{g}_k(\xi,\omega-k_y v)\times\exp[-ik_y(y^R-\xi-y_k)]\,d\xi\,dk_y. \quad (23)\end{aligned}$$

A change of variables according to $k_y=(\omega-\tilde{\omega})/v$ moves the frequency shift from the force distribution $\hat{g}_k(\xi,\omega-k_y v)$ to the Green's function:

$$\hat{u}_z(\mathbf{x}^R,\omega)=\frac{1}{2\pi v}\int_{-\infty}^{+\infty}\tilde{u}_{zz}^G\left(\mathbf{x}^{R'},\frac{\omega-\tilde{\omega}}{v},\omega\right)\int_{-\infty}^{+\infty}\sum_{k=1}^{K}\hat{g}_k(\xi,\tilde{\omega})\times\exp\left[-i\left(\frac{\omega-\tilde{\omega}}{v}\right)(y^R-\xi-y_k)\right]d\xi\,d\tilde{\omega}. \quad (24)$$

Accounting for the discrete spectrum (4) of the load distribution $\hat{g}_k(\xi,\tilde{\omega})$, Eq. (24) becomes

$$\hat{u}_z(\mathbf{x}^R,\omega)=\frac{1}{2\pi v}\int_{-\infty}^{+\infty}\tilde{u}_{zz}^G\left(\mathbf{x}^{R'},\frac{\omega-\tilde{\omega}}{v},\omega\right)\int_{-\infty}^{+\infty}\sum_{k=1}^{K}\frac{T_k}{vN_{\text{eff}}^{\text{st}}}\times\frac{w(\xi)}{w_{\max}^{\text{st}}}\left[\frac{2\pi v}{d}\sum_{m=-\infty}^{\infty}\delta\left(\tilde{\omega}-m\frac{2\pi v}{d}\right)\right]\times\exp\left(i\tilde{\omega}\frac{\xi}{v}\right)\exp\left(i\tilde{\omega}\frac{y_k}{v}\right)\times\exp\left[-i\left(\frac{\omega-\tilde{\omega}}{v}\right)(y^R-\xi-y_k)\right]d\xi\,d\tilde{\omega}. \quad (25)$$

Combining the exponential functions yields

$$\hat{u}_z(\mathbf{x}^R,\omega)=\int_{-\infty}^{+\infty}\tilde{u}_{zz}^G\left(\mathbf{x}^{R'},\frac{\omega-\tilde{\omega}}{v},\omega\right)\int_{-\infty}^{+\infty}\sum_{k=1}^{K}\frac{T_k}{vN_{\text{eff}}^{\text{st}}}\times\frac{w(\xi)}{w_{\max}^{\text{st}}}\frac{1}{d}\sum_{m=-\infty}^{\infty}\delta\left(\tilde{\omega}-m\frac{2\pi v}{d}\right)\times\exp\left(i\omega\frac{\xi}{v}\right)\exp\left(i\omega\frac{y_k}{v}\right)\times\exp\left[-i\left(\frac{\omega-\tilde{\omega}}{v}\right)(y^R)\right]d\xi\,d\tilde{\omega}. \quad (26)$$

The Dirac function enforces the frequency $\tilde{\omega}$ to be equal to $m2\pi v/d$ for each value of m. Reordering of the terms finally gives

$$\hat{u}_z(\mathbf{x}^R,\omega)=\left[\frac{1}{vN_{\text{eff}}^{\text{st}}}\frac{1}{w_{\max}^{\text{st}}}\int_{-\infty}^{+\infty}w(\xi)\exp\left(i\xi\frac{\omega}{v}\right)d\xi\right]\times\left[\sum_{k=1}^{K}T_k\exp\left(i\omega\frac{y_k}{v}\right)\right]\times\left[\frac{1}{d}\sum_{m=-\infty}^{+\infty}\tilde{u}_{zz}^G\left(\mathbf{x}^{R'},\frac{\omega-m\dfrac{2\pi v}{d}}{v},\omega\right)\times\exp\left[-i\left(\frac{\omega-m\dfrac{2\pi v}{d}}{v}\right)y^R\right]\right]. \quad (27)$$

The first two bracketed terms correspond to the functions $\hat{F}(\omega)$ and $\hat{C}(\omega)$ defined in Eqs. (10) and (12); Eq. (27) therefore has the same form as Eq. (16) and the third bracketed term is the desired alternative for the function $\hat{u}_{\delta z}(\mathbf{x}^R,\omega)$. The difference with Eq. (16) is that the summation on the response contributions of the stationary sleeper forces is replaced by a summation on the contribution of moving harmonic sources at the integer multiples of the sleeper passage frequency $f_d=v/d$. The term $m=0$ corresponds to a quasistatic contribution, while $|m|=+1$ corresponds to the sleeper passage frequency. Depending on the sleeper passage frequency, the speed of the train, the dynamic soil characteristics, and the considered frequency range, the summation can be limited to a small number of harmonics, resulting in a considerable computational benefit.

The time history of the response finally follows from the inverse Fourier transformation:

$$u_z(\mathbf{x}^R,t)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\hat{u}_z(\mathbf{x}^R,\omega)\exp(+i\omega t)\,d\omega. \quad (28)$$

V. NUMERICAL EXAMPLE

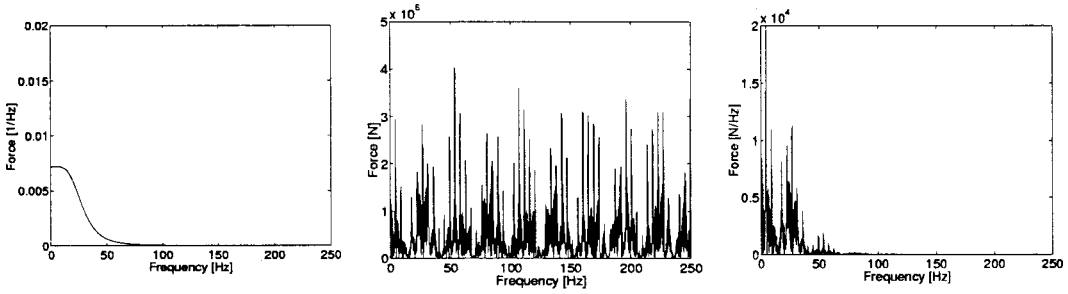
In this section, the numerical model is applied to calculate free field vibrations during the passage of a Thalys HST. The results of *in situ* vibration measurements performed during the homologation tests of the HST track between Brussels and Paris at nine speeds between 223 and 314 km/h is used as a validation tool.[21,22] Accelerations have been measured on the rail and the sleeper of each track and in the free field at distances from 4 to 72 m. In the following, only data recorded at the commercial operating speed of 300 km/h will be briefly reviewed; the reader is referred to other papers for full details on the *in situ* measurements.[21–23]

A. Load distribution due to a train passage

The HST track between Brussels and Paris is a classical ballast track. Continuously welded UIC60 rails with a mass per unit length of 60 kg/m and a moment of inertia $I=0.3038\times10^{-4}$ m^4 are fixed with Pandrol rail fasteners on precast prestressed concrete monoblock sleepers with a mass of 300 kg. Flexible rail pads are placed under the rail. The track is supported by limestone and porphyry ballast and supporting layers with a total thickness of about 1.1 m.

Calculations are made for a track with a bending stiffness $EI=12.76\times10^6$ N m^2 (both rails) and a mass per unit length $m=620.0$ kg/m (both rails and the sleepers). During the homologation tests, access to the track was limited to the time needed for the installation of the accelerometers on the rails and the sleepers. No forced vibration test on the track could be performed to measure the frequency-dependent dynamic impedance of the track. Subsequent calculations are therefore made for a track with a constant subgrade stiffness $k_s=250$ MPa of average magnitude.[31] Figure 3(a) shows the modulus of $\hat{F}(\omega)$ for a train speed $v=300$ km/h. It must be noted that, in more realistic models, the dynamic impedance of the subgrade is a complex, frequency-dependent function that couples the degrees of freedom at the interface between the track and the subgrade.[15]

Figure 4 shows the configuration of the Thalys HST, consisting of two locomotives and eight carriages; the total



a. Modulus of $\hat{F}(\omega)$. b. Modulus of $\hat{C}(\omega)$. c. Modulus of $\hat{F}(\omega)\hat{C}(\omega)$.

FIG. 3. Modulus of the functions (a) $\hat{F}(\omega)$, (b) $\hat{C}(\omega)$, and (c) $\hat{F}(\omega)\hat{C}(\omega)$ during the passage of a Thalys HST at a speed v=300.km/h.

length of the train is equal to 200.18 m. The locomotives are supported by two bogies and have four axles. The carriages adjacent to the locomotives share one bogie with the neighboring carriage, while the six other carriages share both bogies with neighboring carriages. The total number of bogies equals thirteen and, consequently, the number of axles on the train is twenty six. The carriage length L_t, the distance L_b between bogies, the axle distance L_a, the total axle mass M_t, the sprung axle mass M_s, and the unsprung axle mass M_u of all carriages are summarized in Table I.

Figure 3(b) shows the modulus of $\hat{C}(\omega)$ for a Thalys HST moving at a speed $v=300$ km/h. The quasistatic value $\hat{C}(\omega=0)$ is equal to the sum of all axle loads. As the locomotives and the adjacent carriages of the Thalys HST have a different axle composition than the six central carriages, the spectrum is quasidiscrete with peaks at the fundamental bogie passage frequency $f_b=v/L_b=4.45$ Hz and its higher order harmonics, modulated at the axle passage frequency $f_a=v/L_a=27.78$ Hz [Fig. 3(b)]. It has been observed that the measured vertical acceleration of a sleeper has a similar quasidiscrete spectrum.[23] The spectrum in Fig. 3(c) is obtained as the product of $\hat{F}(\omega)$ and $\hat{C}(\omega)$ and represents the frequency content of the force transmitted by a single sleeper during the passage of a Thalys HST. The decay of the function $\hat{F}(\omega)$ with frequency governs the frequency content of this transmitted load.

B. The dynamic soil characteristics

Cone penetration tests and triaxial tests on undisturbed samples taken from borehole experiments reveal the presence of a quaterny loam layer (0–1.5 m) on a transition layer (quaterny loam and/or an Ypresian clay, 1.5–4.0 m) on a tertiary Ypresian clay layer (4.0–12.0 m). A mean value of the soil density $\rho^s=1850$ kg/m^3 is derived from classical laboratory tests on undisturbed samples.

A spectral analysis of surface waves (SASW) has been performed to determine the dynamic soil characteristics of the site and revealed the presence of a top layer with thickness $d=1.4$ m and a shear wave velocity $C_s=80.0$ m/s and a layer ($d=1.9$ m, $C_s=133.0$ m/s) on top of a halfspace ($C_s=226.0$ m/s), in good agreement with the layering revealed by the borehole experiments. The track is constructed in an excavation with a depth of a few meters, where the soil under the ballast has been stabilized. As the SASW test has been performed on the unexcavated soil away from the track, it may be assumed that the soil under the track is stiffer than the soft shallow layer revealed by the SASW test. In the subsequent calculations, a shear wave velocity $C_s=100.0$ m/s will therefore be used for the top layer. A mean value of 0.3 is assumed for the Poisson's ratio ν^s.

Based on a simplified analysis of the transient signals recorded during the SASW test, a hysteretic material damping ratio $\beta^s=0.03$ has been derived. In practice, material damping ratios are expected to decrease with depth and may be lower than the proposed value for deeper layers.

C. The evaluation of the term $\hat{u}_{\delta z}(\mathbf{x}^R,\omega)$

The evaluation of the term $\hat{u}_{\delta z}(\mathbf{x}^R,\omega)$ in Eq. (17) has been implemented for a layered subsoil, using both alternatives presented in Eqs. (16) and (27).

Alternative 1

In the first alternative, the term $\hat{u}_{\delta z}(\mathbf{x}^R,\omega)$ is evaluated according to the bracketed term in Eq. (16), necessitating the evaluation of the Green's function $\hat{u}_{zz}^{G\text{axi}}(r_m,z^R=0,\omega)$ for a large range of source–receiver distances r_m and frequencies ω. The Green's functions are calculated with a direct stiffness formulation for wave propagation in multilayered poroelastic media,[28] using the (horizontal) stratification and the dynamic soil characteristics as derived from the SASW test. The reader is referred to Appendix A for more details on the computational method.

Figure 5 shows the modulus of the vertical Green's function $\hat{u}_{zz}^{G\text{axi}}(r,z^R=0,\omega)$ for horizontal source–receiver dis-

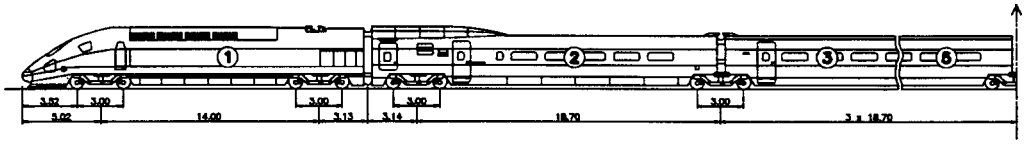


FIG. 4. Configuration of the Thalys HST.

TABLE I. Geometrical and mass characteristics of the Thalys HST.

	No. carriages	No. axles	L_t (m)	L_b (m)	L_a (m)	M_t (kg)	M_s (kg)	M_u (kg)
Locomotives	2	4	22.15	14.00	3.00	17 000	15 267	1733
Adjacent carriages	2	3	21.84	18.70	3.00	14 500	12 674	1830
Central carriages	6	2	18.70	18.70	3.00	17 000	15 170	1826

tances of 25, 50, 75, and 100 m. Radiation and material damping in the soil decrease the amplitude and frequency content of the response for increasing soil-receiver distances. The layered structure of the subsoil introduces a weak cutoff frequency at 10 Hz.

As the material damping in the soil is moderate, $N_s = 1000$ sleepers behind and ahead of the receiver line at $y^R = 0$ are accounted for. Green's functions $\hat{u}_{zz}^{Gaxi}(r, z^R = 0, \omega)$ are therefore needed for a wide range of horizontal source–receiver distances r and frequencies ω. Material and radiation damping in the soil reduce the frequency content of the response for increasing r; the calculations can therefore be optimized by making the distinction between the near and the far field. The spatial sampling of the receivers is denoted by $r_i = r_0 + (i-1)\Delta r$ for $i = 1, \ldots, N_r$, with N_r the number of receivers. In the frequency domain, the bandwidth is defined as $\Delta f = f_{\text{Nyq}}/N_f$, with f_{Nyq} the Nyquist frequency and N_f the number of frequencies. Calculations have been made in seven runs with $N_r = 60$ and $N_f = 1024$. A first run considers the near field response with small receiver spacing ($r_0 = 0.25$ m, $\Delta r = 0.25$ m) in a wide frequency range ($f_{\text{Nyq}} = 512.0$ Hz). The seventh run considers the far field response with large receiver spacing ($r_0 = 488.0$ m, $\Delta r = 8.0$ m) in the low frequency range ($f_{\text{Nyq}} = 8.0$ Hz). The parameters for the five intermediate computer runs were situated between these two extremes. The Green's functions $\hat{u}_{zz}^{Gaxi}(r_m, z^R = 0, \omega)$ are tabulated in a look-up table, as a function of r_m and ω, that can be used for interpolation.

The contribution $\hat{u}_{\delta z}(\mathbf{x}^R, \omega)$ to the response is calculated according to Eq. (16). Despite the use of a look-up table for the interpolation of the Green's functions, the computation time remains very large if compared with the alternative formulation that will be applied in the following.

Figure 6 illustrates the result for a receiver at 8 m from the track. It is obtained by a summation of $N_s = 1000$ Green's functions $\hat{u}_{zz}^{Gaxi}(r_m, z^R = 0, \omega)$, that are multiplied with a cosine function $\cos(\omega md/v)$ with a wavelength $\lambda_v = 2\pi v/\omega$, sampled at discrete locations md. After an important contribution in the low frequency range, the evaluation of the term $\cos(\omega md/v)$ for discrete values of m causes small vibration amplitudes in the frequency range between 20 and 75 Hz. Above 75 Hz, the contribution of the first sleeper passage frequency $f_d = v/d = 138.89$ Hz, smeared over a wider frequency range, can be observed.

Alternative 2

Numerical calculations are more efficient if the term $\hat{u}_{\delta z}(\mathbf{x}^R, \omega)$ is evaluated according to Eq. (27). This approach also allows for a better understanding of the observations that have been made in relation to Fig. 6.

Figure 7 shows a contour plot of the Green's function $\tilde{u}_{zz}^{G}(\mathbf{x}^{R'}, \overline{k_y}, \omega)$ for $x^R = 8$ m and $z^R = 0$ as a function of the frequency ω and the dimensionless wave number $\overline{k_y} = k_y C_s/\omega$, with C_s the shear wave velocity of the top layer. This Green's function is obtained after evaluation of Eq. (B4), as explained in Appendix B. Superimposed on the same plot are the branches of the absolute value of the dimensionless wave number $\overline{k_y} = [(\omega - \tilde{\omega})/v]C_s/\omega$ for $\tilde{\omega} = m2\pi v/d$. According to Eq. (27), the computation of the term $\hat{u}_{\delta z}(\mathbf{x}^R, \omega)$ requires the evaluation of the function $\tilde{u}_{zz}^{G}(\mathbf{x}^{R'}, \overline{k_y}, \omega)$ along these branches; the contour plot in Fig. 7 has only been made for illustrative purposes.

The quasistatic contribution ($m = 0$) corresponds to $\overline{k_y} = C_s/v$, which is a horizontal line in Fig. 7. As the speed v of the train is lower than the Rayleigh wave velocity C_R of the top layer, $\overline{k_y} = C_s/v \geqslant \overline{k_R} = C_s/C_R$. $\overline{k_R}$ is the dimensionless wave number corresponding to Rayleigh wave propagation in the top layer and is equal to 1.08 for a Poisson's ratio $\nu^s = 0.3$. The quasistatic contribution to the response is due to inhomogeneous P and S waves. It rapidly decreases for increasing frequency, as can be observed in the low frequency range of Fig. 6.

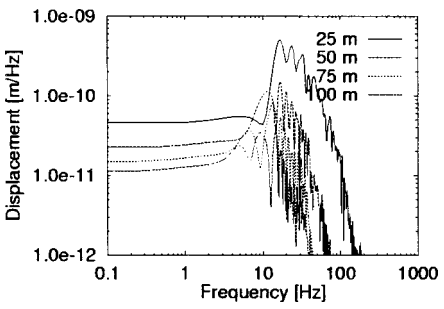


FIG. 5. Modulus of the Green's function $\hat{u}_{zz}^{Gaxi}(r, z^R = 0, \omega)$ of the layered halfspace for horizontal source–receiver distances of 25, 50, 75, and 100 m.

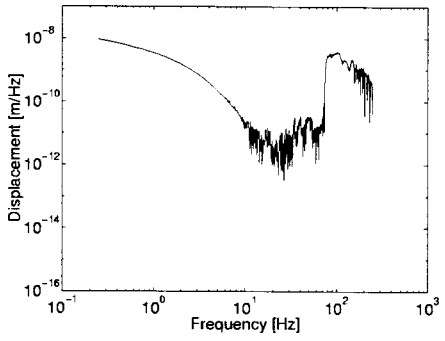


FIG. 6. Modulus of the vertical displacement $\hat{u}_{\delta z}(\mathbf{x}^R, \omega)$ as a function of ω for $x^R = 8$ m, $y^R = 0$, and $z^R = 0$.

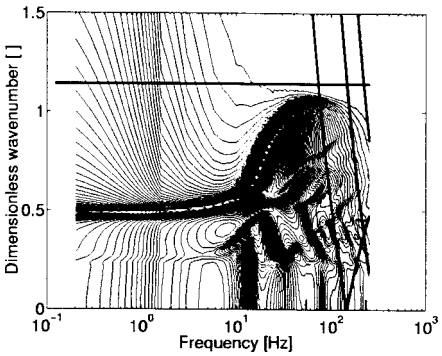


FIG. 7. Modulus of the Green's function $\tilde{u}^G_{zz}(\mathbf{x}^{R'}, \overline{k_y}, \omega)$ as a function of ω and $\overline{k_y}$ for $x^R = 8$ m and $z^R = 0$.

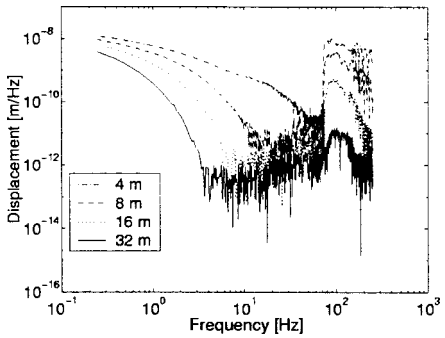


FIG. 8. Modulus of the vertical displacement $\hat{u}_{\delta z}(\mathbf{x}^R, \omega)$ as a function of ω for different x^R.

The evaluation of $\tilde{u}^G_{zz}(\mathbf{x}^{R'}, \overline{k_y}, \omega)$ along the two branches $\overline{k_y} = [(\omega - 2\pi v/d)/v]C_s/\omega$ that correspond to $m = 1$ gives important contributions to the response if $\overline{k_y}$ is smaller than $\overline{k_R}$. Elaboration of this inequality results in a frequency interval $[f_{\min}, f_{\max}] = [73\text{ Hz}, 1389\text{ Hz}]$. This explains why the spectrum in Fig. 6 shows a rather sharp cutoff (high-pass) frequency at $f_{\min} = 73$ Hz, above which the contribution of the first sleeper passage frequency starts. Similar remarks can be made for the higher harmonics m. The required number of harmonics depends on the speed of the train, the soil velocities, and the sleeper passage frequency.

The total number of function evaluations is much lower than in the first alternative based on Krylov's original formulation, reducing the total computation time by a factor of 12 in the present example.

A critical comment needs to be made on the accuracy of the results, however. Figure 8 summarizes the modulus of the vertical displacement $\hat{u}_{\delta z}(\mathbf{x}^R, \omega)$ as a function of the frequency for increasing values of x^R. Figure 8 clearly illustrates that, for increasing x^R, the contribution of the sleeper passage frequency and its higher harmonics is attenuated so that the quasistatic contribution dominates the far field response. For all receiver distances, the accuracy of the predictions in the frequency range between the quasistatic range and $f_{\min}$ is low, since Krylov's prediction model only includes the effects of quasistatic loading and sleeper passage. Other excitation mechanisms, such as parametric excitation, wheel and rail roughness, rail joints and wheel flats, that are important in the midfrequency band, are not included in the model. It is therefore expected that accurate predictions will only be made in the near field.

D. The free field response

The vertical displacement $\hat{u}_z(\mathbf{x}^R, \omega)$ at a receiver due to the passage of a Thalys HST is obtained by evaluation of Eq. (27). The three terms in Eq. (27) have been illustrated in Figs. 3(a), 3(b), and 8. The vertical soil displacements $u_z(\mathbf{x}^R, t)$ in the time domain are obtained by evaluating the inverse Fourier transform (28) with a fast Fourier transform algorithm. A fourth-order high-pass Chebyshev filter is applied to remove spurious low frequency contributions to the response.

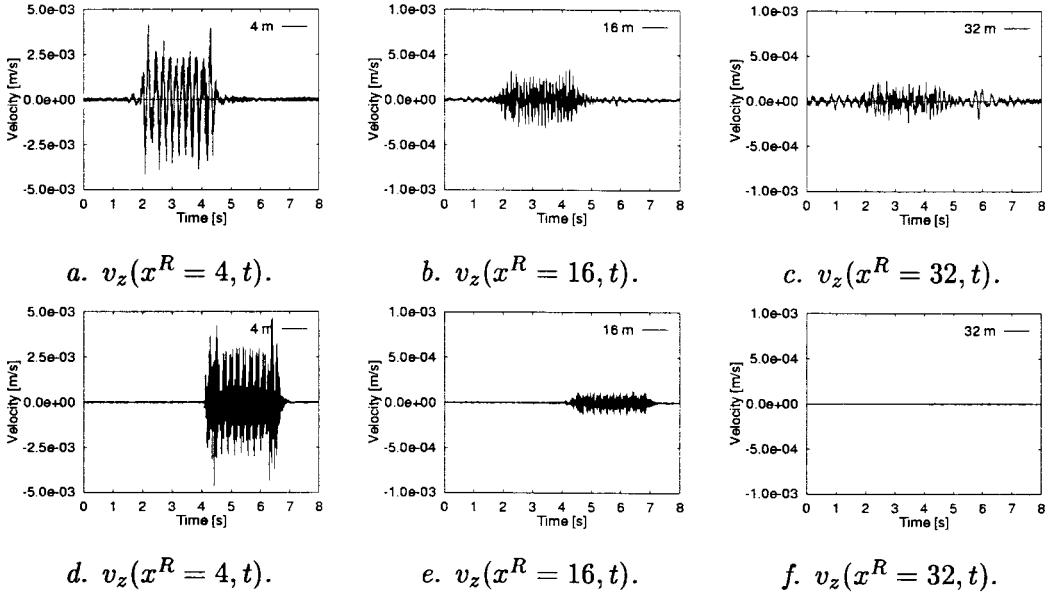


FIG. 9. Measured (top) and predicted (bottom) time history of the vertical velocity at varying distances from the track during the passage of a Thalys HST at $v = 300$ km/h.

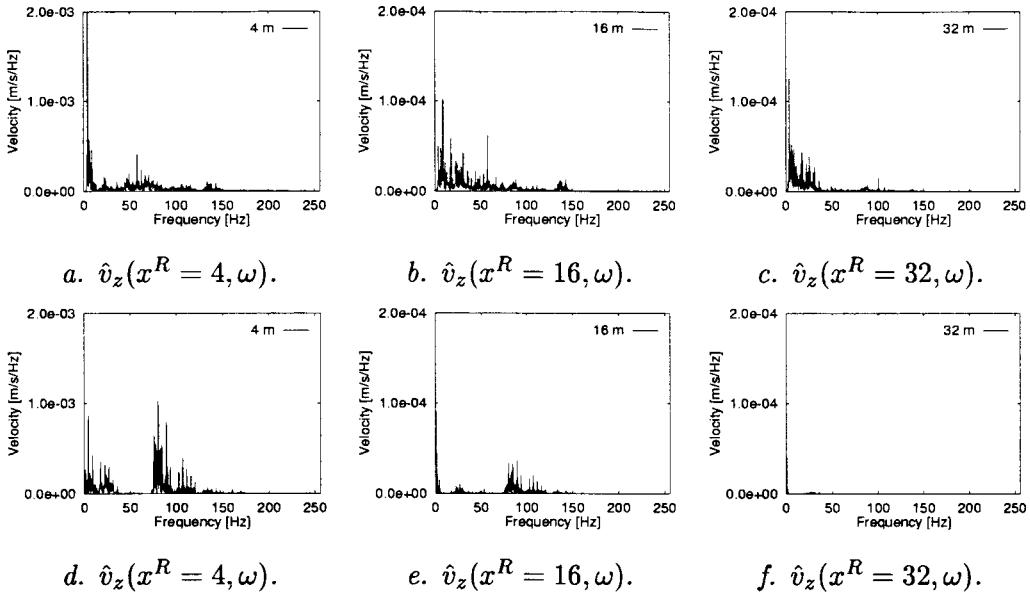


FIG. 10. Measured (top) and predicted (bottom) frequency content of the vertical velocity at varying distances from the track during the passage of a Thalys HST at $v = 300$ km/h.

Figures 9 and 10 compare the measured and predicted time history and frequency content of the vertical velocity at three receivers during the passage of a Thalys HST at 300 km/h. At 4 m from the track, the passage of the bogies can be observed, while the passage of the individual axles is no longer observable [Fig. 9(a)]. A similar remark can be made regarding the predicted velocity [Fig. 9(d)]. Whereas the peak particle velocity (PPV) is well predicted [Fig. 9(d)], the frequency content is too high [Fig. 10(d)], especially beyond $f_{\min}$. Due to the specific train composition, the observed velocity spectrum [Fig. 10(a)] is quasidiscrete, with a maximum at the fundamental bogie passage frequency $f_b = 4.45$ Hz. A similar behavior can be observed at low frequencies in the predicted spectrum [Fig. 10(d)], although the contribution at the fundamental bogie passage frequency is underestimated. This may be due to the soil stratification, which introduces a weak cutoff frequency at 10 Hz, as identified before in Fig. 5. The predicted spectrum has very low amplitudes between 40 and 75 Hz, whereas the frequency content of the response beyond 75 Hz is overestimated. This is caused by the fact that Krylov's prediction model only includes the effect of quasistatic loading and sleeper passage. The model overestimates the sleeper passage effect at high frequencies, as the sleeper forces are transmitted as point forces, an assumption that is challenged when the frequency increases and the wavelengths in the soil have the same order of magnitude as the sleeper dimensions. Furthermore, detailed modeling of the track embankment should provide better results for increasing frequency.

The previous observations become even more pronounced for the measured [Figs. 9(b) and 10(b)] and computed [Figs. 9(e) and 10(e)] response at a larger distance of 16 m from the track. The PPV is underestimated [Fig. 9(e)], while the frequency content is overestimated [Fig. 10(e)], particularly for frequencies beyond $f_{\min}$.

Next, the response in the far field at a distance of 32 m from the track is considered. The time history $v_z(x^R = 32, t)$ [Fig. 9(c)] is shown on the same vertical scale as the results at a distance of 16 m, and has a PPV of about 0.2 mm/s. This is no longer predicted by the numerical model [Fig. 9(f)], which only produces very low noise levels. The velocity spectrum $\hat{v}_z(x^R = 32, \omega)$ [Fig. 10(c)] is dominated by the bogie passage frequency and its second harmonic. Higher frequencies are attenuated by radiation and material damping in the soil; the effect of sleeper passage, for example, is no longer noticeable at 32 m from the track. The predicted velocity spectrum [Fig. 10(f)] is shown on the same vertical scale and is dominated by the quasistatic response, while higher frequency contributions are no longer present; amplitudes are underestimated, however.

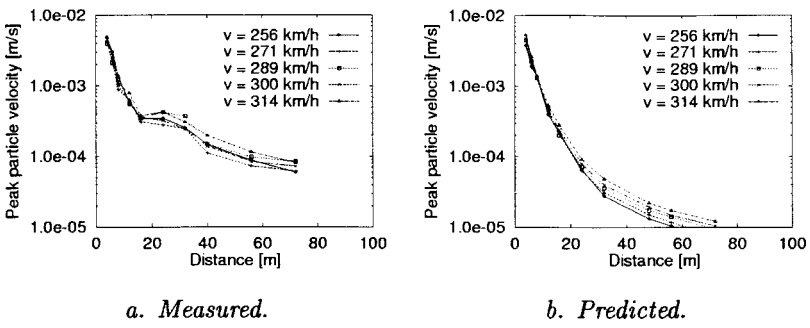


FIG. 11. (a) Measured and (b) predicted PPV as a function of distance and train speed.

E. The influence of the train speed on the peak particle velocity

Figure 11(a) finally summarizes the PPV at all points in the free field for five train passages at different speeds. The decrease of PPV with distance due to radiation and material damping in the soil is apparent. Both experimental and numerical [Fig. 11(b)] results demonstrate a rather weak dependence of PPV on train speed. It must be stressed, however, that train speeds are in the range below the critical velocity of the track–soil system (the Rayleigh wave velocity of the upper layer is about 95 m/s), while an increase of PPV may be expected when the train speed reaches this critical velocity. The PPV is well predicted in the near field (up to 16 m) and seriously underestimated at larger distances from the track; the underlying phenomena have been explained in Sec. V D.

VI. CONCLUSIONS

An efficient alternative formulation of Krylov's prediction model for train induced vibrations, based on the dynamic reciprocity theorem applied to moving sources, has been proposed. In the expression for the free field response in the frequency domain, which is a product of three terms, the original term with a summation on the sleepers is replaced by a summation on the integer multiples of the sleeper passage frequency, resulting in substantial computational benefit.

The numerical model has been validated by means of experimental data, obtained during homologation tests with the Thalys HST on the track between Brussels and Paris. The model has good predictive capabilities in the near field at low (quasistatic) and high frequencies (sleeper passage frequency and its higher harmonics), but underestimates the response in the midfrequency band. The model fails to produce good results in the far field. Apart from incomplete input data (such as dynamic track subgrade stiffness, material damping in the soil) this is mainly due to the fact that the model only incorporates quasistatic loading, which becomes dominant when the train speed reaches the critical velocity of the coupled track–soil system, a condition that has not been reached during the present measurement campaign. In the range of train speeds considered, further improvement will be obtained if other excitation mechanisms, such as dynamic loading due to rail and wheel roughness, are accounted for, and a better dynamic interaction model between the track and the soil is used. These are the objectives of ongoing research.

APPENDIX A: THE AXISYMMETRIC GREEN'S FUNCTION IN THE FREQUENCY DOMAIN

The axisymmetric Green's function $\hat{u}_{zz}^{G\mathrm{axi}}(r,z^R,\omega)$ is calculated as the following inverse Hankel transformation:

$$\hat{u}_{zz}^{G\mathrm{axi}}(r,z^R,\omega)=\int_0^{+\infty}\tilde{u}_{zz}^{G\mathrm{axi}}(k_r,z^R,\omega)k_rJ_0(k_rr)dk_r. \tag{A1}$$

The solution $\tilde{u}_{zz}^{G\mathrm{axi}}(k_r,z^R,\omega)$ is computed in the radial wave-number-frequency domain by means of an exact stiffness formulation for a multilayered halfspace.[13,27–29] In the absence of material damping, the function $\tilde{u}_{zz}^{G\mathrm{axi}}(k_r,z^R,\omega)$ becomes zero for certain values of k_r, which correspond to the surface wave poles of the integrand. The number of poles is almost proportional to the frequency and the total thickness of the layers above the halfspace. Several techniques to remove the poles from the path of integration have been reported in literature. We have added material damping as this method of pole removal corresponds to a physical reality. Moreover, the kernel function $J_0(k_rr)$ introduces rapid oscillations for large values of the horizontal source–receiver distance.

In view of the foregoing remarks, an efficient quadrature scheme is needed for the evaluation of the inverse wave number integrals. According to Xu and Mal[32] an adaptive algorithm with self-adjusting interval, concentrating abscissas around regions of sharp variations in $\tilde{u}_{zz}^{G\mathrm{axi}}(k_r,z^R,\omega)$ and taking full advantage of previously computed values of the integrand, is most useful in order to obtain an accurate evaluation of the wave number integral with a minimum number of function evaluations.

Apsel and Luco[27] have used Lagrange's five-point formula, representing the integrand locally with a quartic polynomial for variable step size. They use a hybrid quadrature scheme, in a sense that, depending on the magnitude of the arguments of the Bessel functions, a non-Filon method or a Filon method is used. We have followed Frazer and Gettrust[33] and used a generalized Filon method which makes use of an asymptotic expansion in a way that does not depend on the accuracy of the expansion.

For the numerical results presented in this paper, the value of the dimensionless quadrature step $\Delta\overline{k_r}=\Delta k_rC_s/\omega=0.001$ is specified. The choice of the quadrature step is not governed by a requested absolute and/or relative accuracy over the whole integration interval. Instead, $\Delta\overline{k_r}$ is allowed to decrease or increase by a factor of 16, depending on an integration error, estimated as the contribution of the fourth-order term to the integral in the integration interval under consideration. The upper bound restriction on $\Delta\overline{k_r}$ is relaxed for $\overline{k_r}>1.5$ beyond which the integrand is known to be monotonic. The total integration interval is truncated at $\overline{k_r}^{\max}$. A particular value of $\overline{k_r}^{\max}$ defines a decreasing spatial resolution for increasing frequency ω. In this example, the integration is continued up to $\overline{k_r}^{\max}=3200.0$, a value based on experience rather than on a criterium based on requested accuracy.

APPENDIX B: THE 3D GREEN'S FUNCTION IN THE WAVE-NUMBER-FREQUENCY DOMAIN

The Green's function $\tilde{u}_{zz}^{G}(x^R,k_y,z^R,\omega)$ in the wave-number-frequency domain is found as the inverse Fourier transformation of the Green's function $\tilde{\tilde{u}}_{zz}^{G}(k_x,k_y,z^R,\omega)$:

$$\tilde{u}_{zz}^{G}(x^R,k_y,z^R,\omega)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\tilde{\tilde{u}}_{zz}^{G}(k_x,k_y,z^R,\omega)\times\exp(-ik_x x^R)\,dk_x. \tag{B1}$$

The Green's function $\tilde{\tilde{u}}_{zz}^{G}(k_x,k_y,z^R,\omega)$ can be written immediately in terms of the axisymmetric Green's function $\tilde{u}_{zz}^{G\mathrm{axi}}(k_r,z^R,\omega)$:

$$\tilde{\tilde{u}}_{zz}^{G}(k_x,k_y,z^R,\omega)=2\pi\tilde{u}_{zz}^{G\mathrm{axi}}(\sqrt{k_x^2+k_y^2},z^R,\omega), \tag{B2}$$

where use has been made of the relation $k_r^2=k_x^2+k_y^2$ between the radial wave number k_r and the horizontal wave numbers k_x and k_y. The factor 2π follows from the relationship between the Hankel transform used in a cylindrical coordinate system and the double Fourier transform used in a Cartesian frame of reference. According to Eqs. (B1) and (B2), the Green's function $\tilde{u}_{zz}^{G}(x^R,k_y,z^R,\omega)$ becomes

$$\tilde{u}_{zz}^{G}(x^R,k_y,z^R,\omega)=\int_{-\infty}^{+\infty}\tilde{u}_{zz}^{G\mathrm{axi}}(\sqrt{k_x^2+k_y^2},z^R,\omega)\times\exp(-ik_x x^R)\,dk_x. \tag{B3}$$

Since the axisymmetric function is an even function of k_x, this can be simplified to

$$\tilde{u}_{zz}^{G}(x^R,k_y,z^R,\omega)=2\int_{0}^{+\infty}\tilde{u}_{zz}^{G\mathrm{axi}}(\sqrt{k_x^2+k_y^2},z^R,\omega)\cos(k_x x)\,dk_x. \tag{B4}$$

For each value of k_y, this inverse integral transformation is performed with a Filon method, where the kernel-free part of the integral is sampled at nonequally spaced abscissa in the integration interval $[0,k_x^{\max}]$ and approximated by a parabolic function. The integration rule is based on overlapping parabola's and it combines an integrating and a smoothing function.[34]

A few remarks regarding the choice of the quadrature step size $\Delta k_x=k_x^{i+1}-k_x^i$ and the total length $k_x^{\max}$ of the integration interval are appropriate here. As explained in Appendix A, the axisymmetric Green's function $\tilde{u}_{zz}^{G\mathrm{axi}}(k_r,z^R,\omega)$ is computed with an adaptive generalized Filon's method, using a fourth-order polynomial interpolation of the kernel-free part of the integrand, involving a variable quadrature step Δk_r. Δk_x is therefore determined as

$$\Delta k_x=\sqrt{(k_r^i)^2-k_y^2}-\sqrt{(k_r^{i+1})^2-k_y^2}, \tag{B5}$$

with k_r^i and k_r^{i+1} two successive radial wave numbers for which the axisymmetric Green's function $\tilde{u}_{zz}^{G\mathrm{axi}}(k_r,z^R,\omega)$ has been calculated. Care must be also taken regarding the choice of the upper limit $k_x^{\max}$ for the numerical integration, especially for x tending to 0, as the singular behavior of the Green's function only appears for large $k_x^{\max}$.

[1] K. Adolfsson, B. Andréasson, P.-E. Bengtson, A. Bodare, C. Madshus, R. Massarch, G. Wallmark, and P. Zackrisson, "High speed lines on soft ground. Evaluation and analyses of measurements from the West Coast Line," Technical report, Banverket, Sweden, 1999.

[2] F. C. P. de Barros and J. E. Luco, "Response of a layered viscoelastic half-space to a moving point load," Wave Motion **19**, 189–210 (1994).

[3] H. Grundmann, M. Lieb, and E. Trommer, "The response of a layered half-space to traffic loads moving along its surface," Arch. Appl. Mech. **69**, 55–67 (1999).

[4] M. Lieb and B. Sudret, "A fast algorithm for soil dynamics calculations by wavelet decomposition," Arch. Appl. Mech. **68**, 147–157 (1998).

[5] D. Aubry, D. Clouteau, and G. Bonnet, "Modelling of wave propagation due to fixed or mobile dynamic sources," in *Workshop Wave '94, Wave Propagation and Reduction of Vibrations*, edited by N. Chouw and G. Schmid (Rühr University, Bochum, Germany, 1994), pp. 109–121.

[6] G. Lombaert and G. Degrande, "Experimental validation of a numerical prediction model for free field traffic induced vibrations by *in situ* experiments," Soil Dyn. Earthquake Eng. **21**, 485–497 (2001).

[7] G. Lombaert, G. Degrande,and D. Clouteau, "Numerical modelling of free field traffic induced vibrations," Soil Dyn. Earthquake Eng. **19**, 473–488 (2000).

[8] D. Clouteau, G. Degrande, and G. Lombaert, "Numerical modelling of traffic induced vibrations," Meccanica (to be published).

[9] X. Sheng, C. J. C. Jones, and M. Petyt, "Ground vibration generated by a harmonic load acting on a railway track," J. Sound Vib. **225**, 3–28 (1999).

[10] X. Sheng, C. J. C. Jones, and M. Petyt, "Ground vibration generated by a load moving along a railway track," J. Sound Vib. **228**, 129–156 (1999).

[11] A. M. Kaynia, C. Madshus, and P. Zackrisson, "Ground vibration from high speed trains: Prediction and countermeasure," J. Geotech. Geoenviron. Eng. **126**, 531–537 (2000).

[12] C. Madshus and A. M. Kaynia, "High-speed railway lines on soft ground: Dynamic behaviour at critical train speed," J. Sound Vib. **231**, 689–701 (2000).

[13] E. Kausel and J. M. Roësset, "Stiffness matrices for layered soils," Bull. Seismol. Soc. Am. **71**, 1743–1761 (1981).

[14] K. Knothe and Y. Wu, "Receptance behaviour of railway track and subgrade," Arch. Appl. Mech. **68**, 457–470 (1998).

[15] P. Van den Broeck and G. De Roeck, "The vertical receptance of track including soil-structure interaction," in *Proceedings of the Fourth European Conference on Structural Dynamics: Eurodyn '99, Prague, Czech Republic, 7–10 June 1999*, edited by L. Frýba and J. Náprstek (Balkema, Rotterdam, 1999), pp. 837–842.

[16] V. V. Krylov, "On the theory of railway-induced ground vibrations," J. Phys. IV **4**, 769–772 (1994).

[17] V. V. Krylov, "Generation of ground vibrations by superfast trains," Appl. Acoust. **44**, 149–164 (1995).

[18] V. V. Krylov, "Vibrational impact of high-speed trains. I. Effect of track dynamics," J. Acoust. Soc. Am. **100**, 3121–3134 (1996).

[19] V. V. Krylov, "Vibrational impact of high-speed trains. I. Effect of track dynamics," J. Acoust. Soc. Am. **101**, 3810 (1997). Erratum.

[20] V. V. Krylov, "Effects of track properties on ground vibrations generated by high-speed trains," Acust. Acta Acust. **84**, 78–90 (1998).

[21] G. Degrande, "Free field vibrations during the passage of a high speed train: Experimental results and numerical predictions," in *Noise and Vibration from High-speed Trains*, edited by V. V. Krylov (Telford, London, 2001), pp. 283–312.

[22] G. Degrande and G. Lombaert, "High-speed train induced free field vibrations: *In situ* measurements and numerical modelling," in *Proceedings of the International Workshop Wave 2000, Wave Propagation, Moving Load, Vibration Reduction*, edited by N. Chouw and G. Schmid (Rühr University, Bochum, Germany, 2000), pp. 29–41.

[23] G. Degrande and L. Schillemans, "Free field vibrations during the passage of a Thalys HST at variable speed," J. Sound Vib. (to be published).

[24] V. V. Krylov and C. C. Ferguson, "Recent progress in the theory of railway-generated ground vibrations," Proc. Inst. Acoust. **17**, 55–68 (1995).

[25] W. M. Ewing, W. S. Jardetzky, and F. Press, *Elastic Waves in Layered Media* (McGraw-Hill, New York, 1957).

[26] H. Lamb, "On the propagation of tremors over the surface of an elastic solid," Philos. Trans. R. Soc. London, Ser. A **203**, 1–42 (1904).

[27] R. J. Apsel and J. E. Luco, "On the Green's functions for a layered half-space. Part II," Bull. Seismol. Soc. Am. **73**, 931–951 (1983).

[28] G. Degrande, G. De Roeck, P. Van den Broeck, and D. Smeulders, "Wave propagation in layered dry, saturated and unsaturated poroelastic media," Int. J. Solids Struct. **35**, 4753–4778 (1998). Poroelasticity Maurice A. Biot memorial issue.

[29] J. E. Luco and R. J. Apsel, "On the Green's functions for a layered half-space. I," Bull. Seismol. Soc. Am. **4**, 909–929 (1983).

[30] R. G. Payton, "An application of the dynamic Betti–Rayleigh reciprocal

theorem to moving point loads in elastic media,'' Q. Appl. Math. **21**, 299–313 (1964).
[31] K. Knothe and S. L. Grassie, ''Modelling of railway track and vehicle/track interaction at high frequencies,'' Veh. Syst. Dyn. **22**, 209–262 (1993).
[32] P.-C. Xu and A. K. Mal, ''An adaptive integration scheme for irregularly oscillatory functions,'' Wave Motion **7**, 235–243 (1985).
[33] L. N. Frazer and J. F. Gettrust, ''On a generalization of Filon's method and the computation of the oscillatory integrals of seismology,'' Geophys. J. R. Astron. Soc. **76**, 461–481 (1984).
[34] P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration*, 2nd ed. (Academic, New York, 1984).

Experimental study of structure-borne sound transmission loss of mechanical joints

Leping Feng,[a] Mingshu Liu, and Anders Nilsson
MWL, Department of Vehicle Engineering, KTH (The Royal Institute of Technology), SE-100 44 Stockholm, Sweden

(Received 22 January 2001; accepted for publication 22 May 2001)

A mechanical joint is one of the most effective ways to reduce the transmission of structure-borne sound. In order to increase the transmission loss, heavily damped joints are often used, which, in many cases, will reduce the structure integrity and hence can only be used in limited cases. In this study attention is focused on a type of resonant joint, i.e., a joint which will increase the transmission loss but will not reduce the structure integrity. The study is based on experiments in a one-dimensional structure. It is found that by adjusting the overlap of the joint, the transmission loss of 30 dB can be obtained at a certain frequency range without adding any dissipative materials. The mechanism of this high transmission loss is the cantilever-type resonance. The resonant frequency can be predicted precisely. The influence of extra dissipative material is investigated. The performance of the same joint in a finite structure is also examined by using the concept of vibrational insertion loss. When there is a certain damping in a finite system, a rather high insertion loss can still be achieved by using the above-mentioned joint, but the resonant frequency is shifted to higher end. It seems that the effective length of the cantilever is shortened by the finiteness. [DOI: 10.1121/1.1387996]

PACS numbers: 43.40.At, 43.40.Tm [PJR]

I. INTRODUCTION

A mechanical joint is one of the most common, and the most effective, ways to reduce transmitted structure-borne sound and hence the radiated noise level. Much attention has been paid recently to optimize the designs of structure connections for noise and vibration reductions. The European project *OSCAR* (Optimization of Structural Connections for Noise and Vibration Reduction), which involves 12 partners and lasts for 3 years, is an example of such an effort. In order to increase vibration transmission loss and to reduce vibration velocity level, damped joints are often used in practice, which, in many cases, are based on adding different types of viscoelastic materials and will inevitably reduce the structure integrity; hence, they can only be used in limited situations.

An investigation on the vibration transmission loss of a traditional bolted joint is carried out in this paper. In order that the joint has a high mechanical strength, no special emphasis is put on damping material. Instead, the investigation is mainly concerned with the geometry of a joint. Since no theoretical model is available for this type of traditional joint, the study described below is based on measurements concerned with bending waves only.

Two parameters are used in the investigation: vibrational transmission loss when the property of an infinite system is concerned, and "vibrational insertion loss" [1] when the system is finite. The latter is the reduction of the averaged vibration velocity level of a certain area by the insertion of the joint into a finite system and is similar to the concept of acoustical insertion loss as one uses for the description of the properties of a muffler.

[a]Author to whom correspondence should be addressed; electronic mail: fengl@fkt.kth.se

II. TRANSMISSION LOSS

A beam system is considered. At any point of the beam, there is always a bending wave, and a near-field solution, propagating in each direction. When the measurement points are far away from any discontinuities, the near-field solutions can be neglected and the transfer functions between the applied force and the vibration signals at measurement point 1 and 2 can be expressed as

$$\begin{aligned} H_1 &= H_+ + H_- \\ H_2 &= H_+ \exp(-jks) + H_- \exp(jks), \end{aligned} \tag{1}$$

where "+" and "−" denote the direction of wave propagation, and s is the distance between the two points. The term $\exp(j\omega t)$ is omitted throughout the paper. From the two measurements we can calculate the transfer functions between the applied force and the waves propagating in each direction as

$$\begin{aligned} H_+ &= \frac{H_1 \exp(jks) - H_2}{\exp(jks) - \exp(-jks)} \\ H_- &= \frac{H_2 - H_1 \exp(-jks)}{\exp(jks) - \exp(-jks)}. \end{aligned} \tag{2}$$

The formula is valid when $ks \neq n\pi$, where n is an integer.

The vibration transmission loss of a joint, as defined in Ref. 2, is measured by using the arrangement as illustrated in Fig. 1, where beam A is connected to beam B through the joint being tested. The four measurement points are symmetrically placed about the center of the joint. Using the coordinate system as illustrated in Fig. 2, the relation between the propagating waves in the two beams, in a steady state, can be expressed in matrix form as

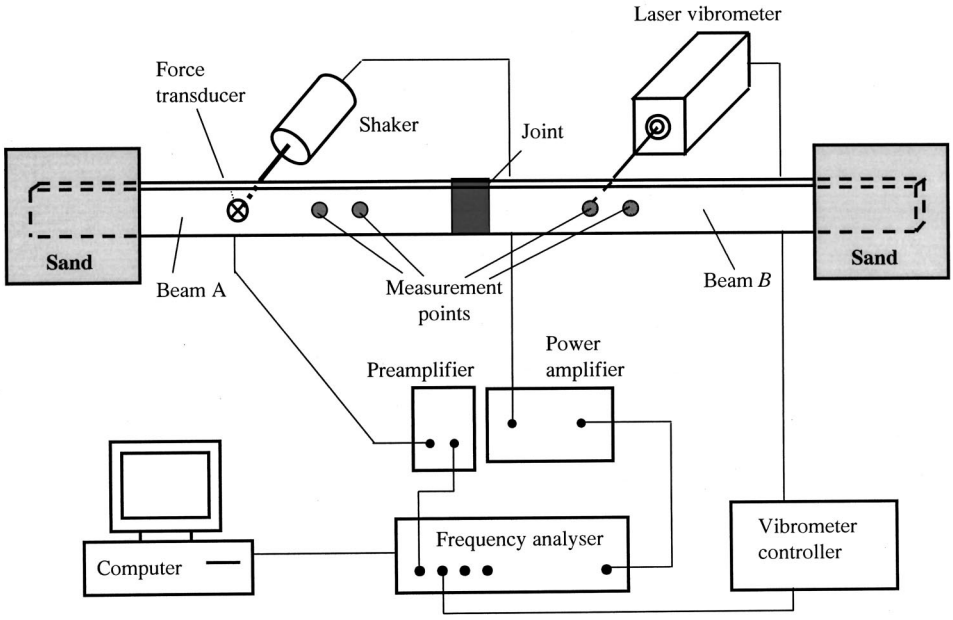


FIG. 1. Arrangement for the measurement of transmission loss.

$$\begin{bmatrix} H_{a+} \\ H_{b+} \end{bmatrix} = \begin{bmatrix} H_{a-} & H_{b-} \\ H_{b-} & H_{a-} \end{bmatrix} \begin{bmatrix} r \\ t \end{bmatrix}, \tag{3}$$

where r is the reflection coefficient and t is the transmission coefficient. Solving Eq. (3), we get

$$\begin{bmatrix} r \\ t \end{bmatrix} = \begin{bmatrix} H_{a-} & H_{b-} \\ H_{b-} & H_{a-} \end{bmatrix}^{-1} \begin{bmatrix} H_{a+} \\ H_{b+} \end{bmatrix}. \tag{4}$$

The vibration transmission loss is finally obtained as

$$R = -10^{*}\log(|t|^2). \tag{5}$$

Note that this method does not require reflection-free terminals. The arrangement of sandboxes at the terminal in the test setup (Fig. 1) is used to reduce reflections and hence to increase measurement accuracy.

The beam used in the measurement is 35 mm wide and about 2000 mm long, with the joint being tested located in the middle area. The beam is embedded into sand at both ends, each about 300 mm long, in order to reduce the reflection from terminals. The system is excited with white noise by a shaker (B&K 4810) connected to the beam via a thin steel bar and a force transducer (B&K 8200). A laser Doppler meter (Polytec Vibrometer) picks up the vibration signals.

The transmission losses for beams with thickness range from 1 to 4 mm, and the configuration of the joint being tested is shown in Fig. 3, where the result of the 2-mm beam is the average of six measurements by disassembling and reassembling the joint. Surprisingly, we get two very high peaks between 2 and 2.5 kHz for the 1-mm beam. For the 2-mm beam, we also get a fairly high peak of transmission loss at around 4 kHz. Measurements show that if the tightening torque is larger than a certain level, we always get the same result, with the standard deviation (for data in dB scale) less than 10%.

By checking the geometry of the beam, we noticed that one peak of the 1-mm beam, which is at about 2020 Hz, is due to the beamwidth (35 mm): It is just a half-wavelength of a 1-mm steel plate at that frequency. Since the beam is only connected at the center, it is very easy for the beam to resonate at half-wavelength, which, due to the interaction of the two beams at the overlap area, in turn produces a very high transmission loss. In order to suppress this type of peak, in the following measurements we mainly use two bolts in a row.

There is still one more peak for the 1-mm beam and another one for the 2-mm beam. Further investigation reveals that those peaks correspond to the first resonance of a clamped-free beam with roughly the same length of distance between the bolts and the edge of the beam. It is this resonance that blocks vibration transmission across the joint. For thicker beams, the peaks appear at higher frequencies and hence are not noticeable in the frequency range used for the analysis. In order to check that, more measurements on vibration transmission loss are performed. Some results on the 1-mm beam are shown in Fig. 4, with the corresponding specifications of the joints illustrated in the same figure. All joints are without any extra absorptive treatment.

There is basically no transmission loss at the investigated frequency range ($<$3000 Hz) when the distance between the bolts and the edge is 10 mm, which is what was expected. When this distance is increased to 20 mm, a high peak appears at around 2–2.5 kHz. The vibration transmission loss at this frequency range is about 25 to 45 dB! Increasing this distance to 30 mm we get a peak of 50 dB at about 1 kHz, and with unequal distance of 20 and 30 mm we

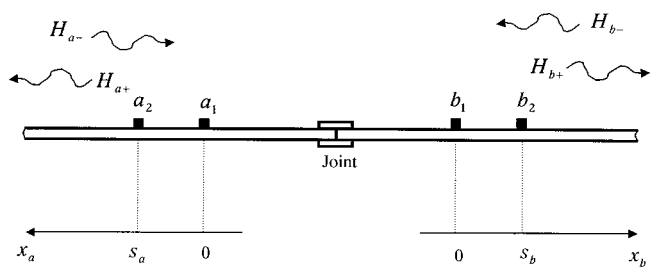


FIG. 2. The coordinate system for the calculation of transmission loss.

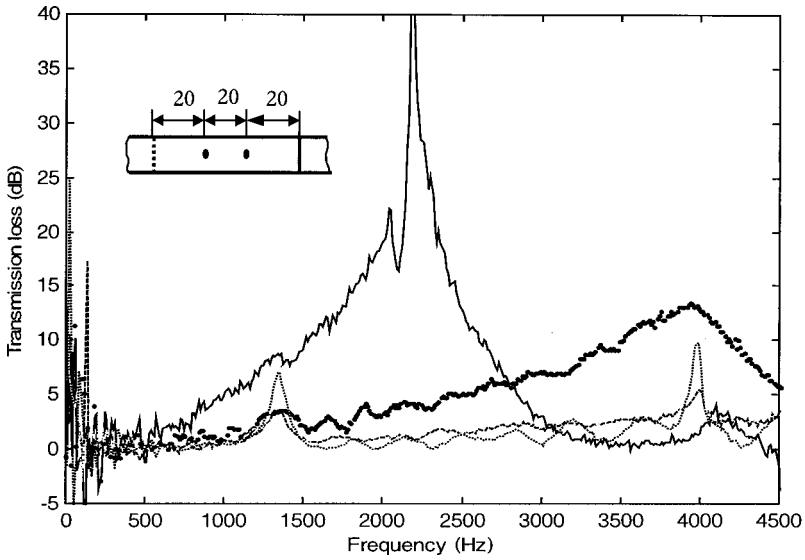


FIG. 3. Transmission loss of a joint in four different beams. Solid: h=1 mm, Dotted thick: h=2 mm, Dashed: h=3 mm, Dotted thin: h=4 mm.

get two peaks of transmission loss at the same frequencies as previous measurements, although with lower level. The frequencies of all those peaks are predictable by estimating the ''cantilever'' beam resonance.

For the case of 10-mm distance between the bolts and the edge, the peak of transmission loss should appear at the frequencies near 10 kHz, which is not detectable when the up-frequency limit is lower than that. This also explains why this cantilever resonance has seldom been reported: The length of the cantilever in a practical structure usually is around 10 mm, which puts the resonance, in most cases, out of the frequency range of interest.

Measurements are also made for 2-mm thick beam (Fig. 5) with the same kind of joints. The peaks of transmission loss appear at exactly the same frequencies as we expected. Similar to the case of the 1-mm-thick beam, unequal distance yields two lower peaks. The second peak of the combination of 20- and 50-mm distances seems higher and wider than what we expected. The reason is that this is a combination of two peaks: the first resonance of the 20-mm cantilever beam (~4196 Hz) and the second resonance of the 50-mm cantilever beam (~4272 Hz). The combination of 10- and 60-mm spacing also shows the first and second resonance of the 60-mm cantilever in the transmission loss.

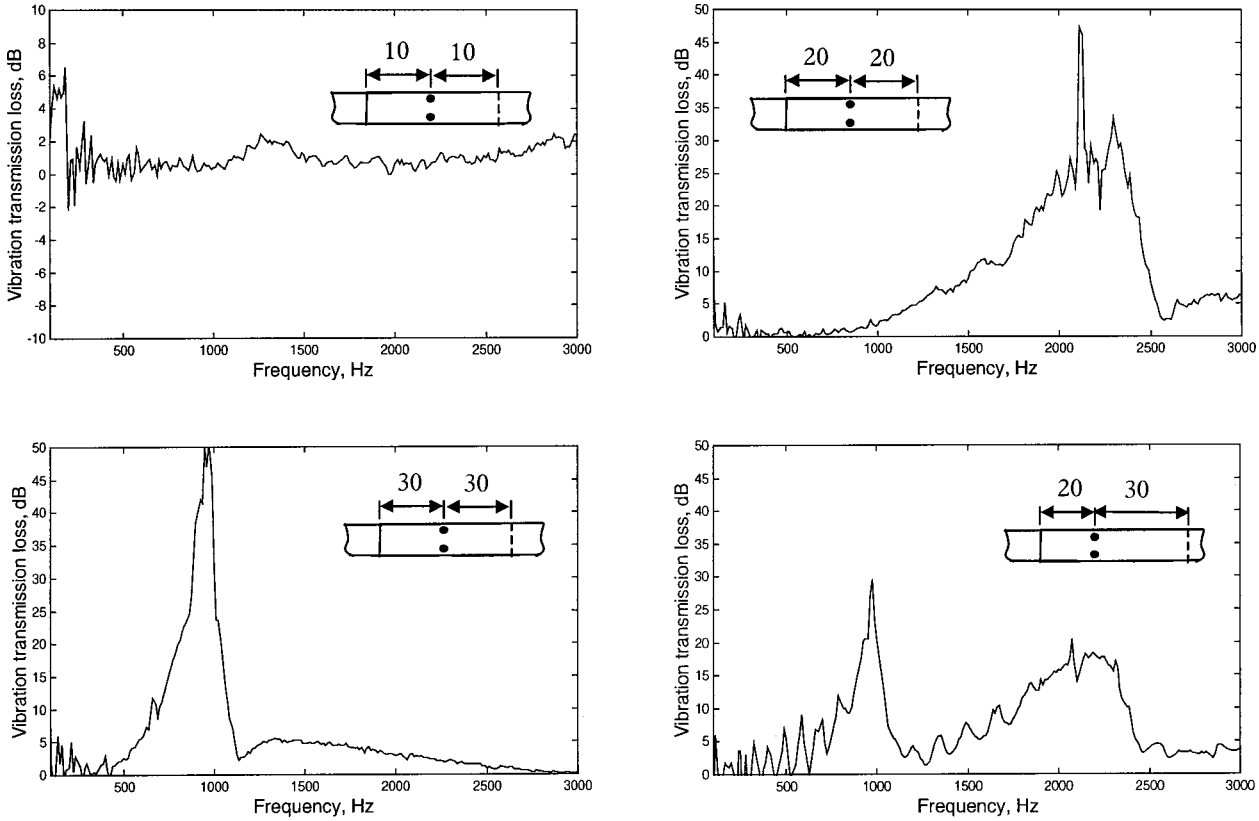


FIG. 4. Measured transmission loss of joints with 1 line of bolts. Thickness of the beam: 1 mm.

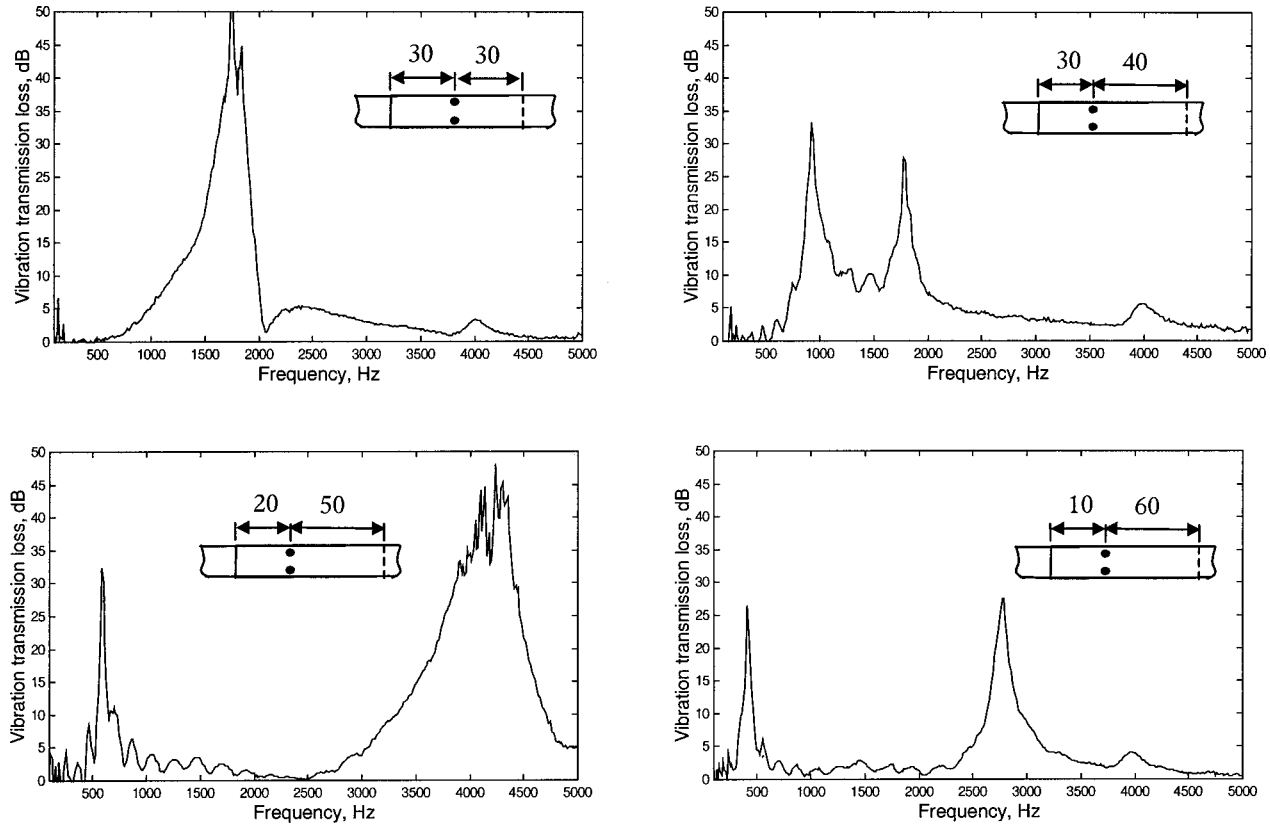


FIG. 5. More results for joints with 1 line of bolts. Thickness of the beam: 2 mm.

In order to see the mechanism of the high transmission loss, the reflection coefficient r and the quantity $1-r^2-t^2$, which can be taken as an estimate of the percentage of the energy absorbed by the joint when both sides of the joint are of the same thickness, are presented in Fig. 6. The corresponding transmission loss is shown in the lower right of Fig. 5. One may see that most of the energy is reflected back to the excitation side instead of being absorbed by the joint. The peak of the curve $1-r^2-t^2$ at around 4000 Hz, which also corresponds to a small peak of transmission loss, may be due to the friction between beams.

The influence of dissipative materials on the transmission loss is investigated in Fig. 7, where 5-mm-thick porous rubber is placed in between the two beams. As expected, a sound transmission loss of about 3–4 dB is obtained in a rather wide frequency range. A high peak of transmission loss is also seen in the figure, with the frequency much higher than the corresponding one without dissipative material, since the effective length of the cantilever is shortened due to the existence of the damping layer. If the cantilever-type resonance is desired to increase the transmission loss, one has to be careful with the dissipative materials.

A special ''washer'' is also designed and tested; the configuration and results are shown in Fig. 8. Three peaks can be seen from the figure, with one corresponding to the 20-mm cantilever and two corresponding to the 40-mm cantilever (the first and second resonance). These results indicate that in practical cases we may design special washers either to reduce the transmission loss of structure-borne sound at certain frequencies, or to increase the effective frequency band of sound reduction for the existing joints.

Figure 9 shows transmission loss of joints with two lines

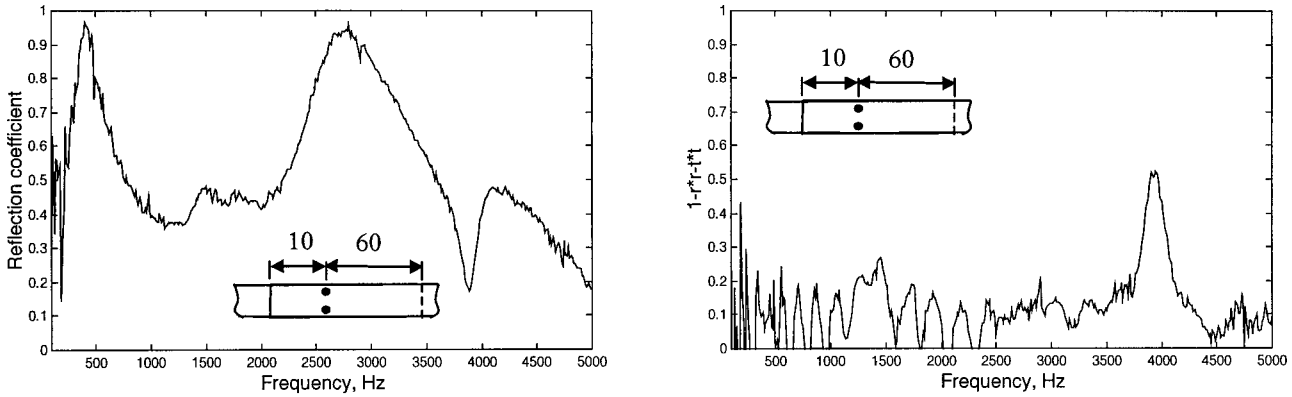
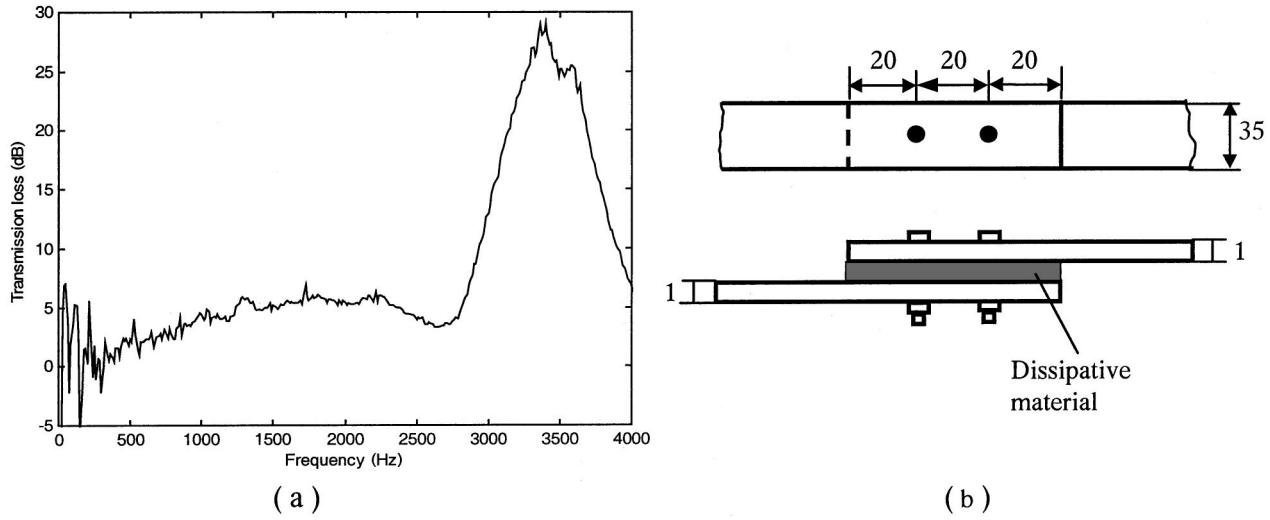


FIG. 6. Reflection coefficient and the energy absorbed at the joint. Thickness of the beam: 2 mm. The corresponding transmission loss is shown in Fig. 5.



(a) (b)

FIG. 7. Transmission loss of a joint with dissipative material between beams.

of bolts, with the specification shown in the same figure. The distance between the bolts and the edge is fixed to 10 mm in this configuration. As we have seen from the measurements shown in Fig. 4, this edge distance has virtually no influence on transmission loss in the frequency range of interest. Hence, the transmission loss obtained in Fig. 9 is mainly contributed by the part between the two lines of the bolts.

There is a blunt peak of transmission loss in this type of joint, with a height of about 4–6 dB. The position of this peak is in inverse proportion to the distance between the joints rather than to the square of the distance between the joints, indicating that the reflection due to the beam resonance is not the only reason for the transmission loss. As indicated in Refs. 2 and 3, the damping introduced by such an arrangement, mainly due to the friction of the two beams, is rather high. It is the combination of the above-indicated two effects which yields the transmission loss of this type of arrangement.

III. VIBRATIONAL INSERTION LOSS

A practical system is always finite. For many cases it is difficult to measure, and to interpret results by, transmission loss. The vibrational insertion loss is introduced in Ref. 1 for such a situation. This method employs a reference sample, which is otherwise identical to the test one but without the joint being tested. The two samples are mounted in a test rig successively and measured under exactly the same condition. By comparing the averaged vibration velocity level of the two successive measurements, the vibration reduction introduced by the insertion of the joint, or vibrational insertion loss, is obtained. As we can see from the method, the results obtained are also dependent on other structures connected to the test samples. Hence, it is very important to have the test situation, vibro-acoustically, as close to a practical situation as possible.

The same beam and joints as described previously are also tested for vibrational insertion loss in order to see the

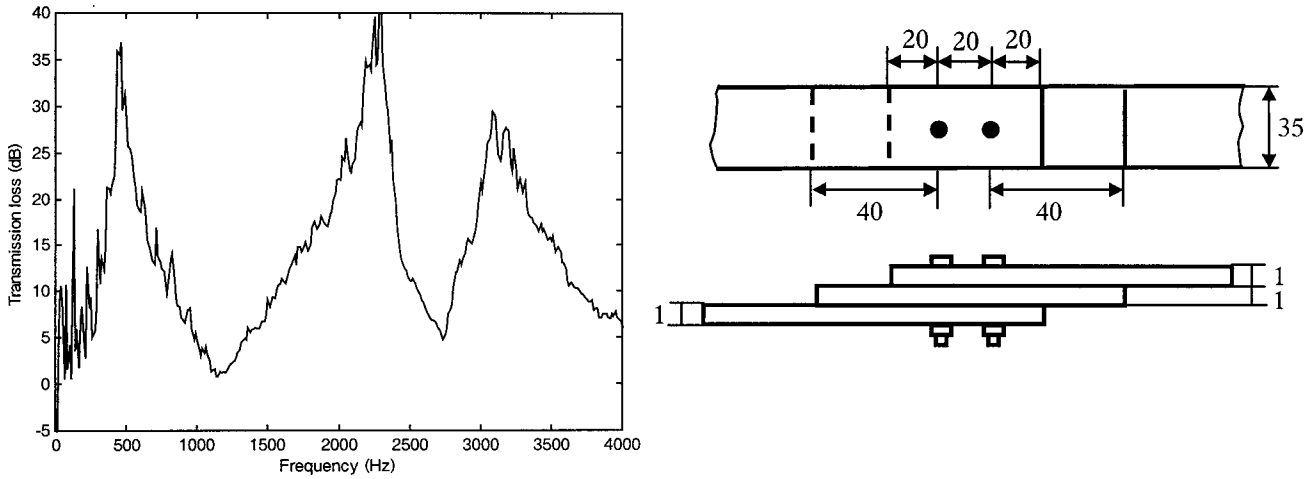


FIG. 8. Transmission loss of a joint with a special washer.

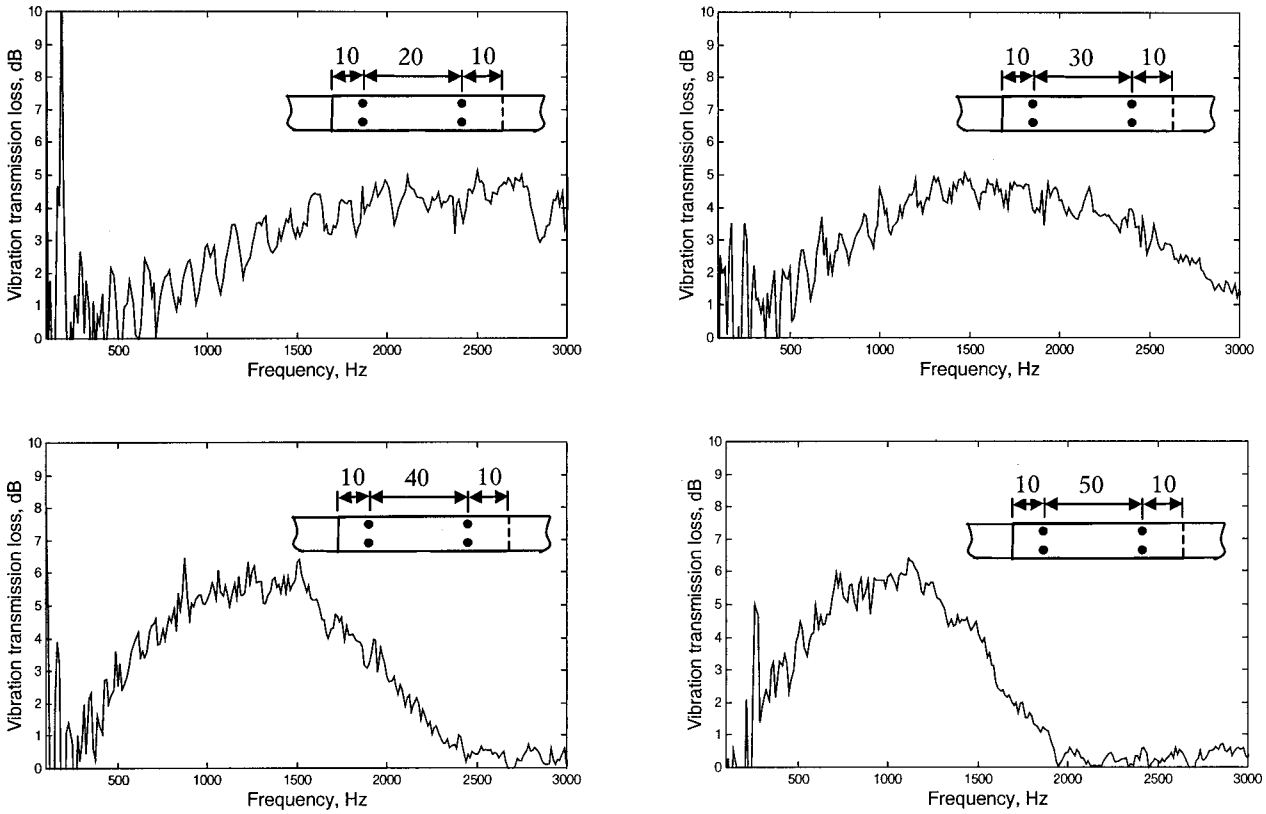


FIG. 9. Transmission loss of joints with two lines of bolts. Thickness of the beam: 1 mm.

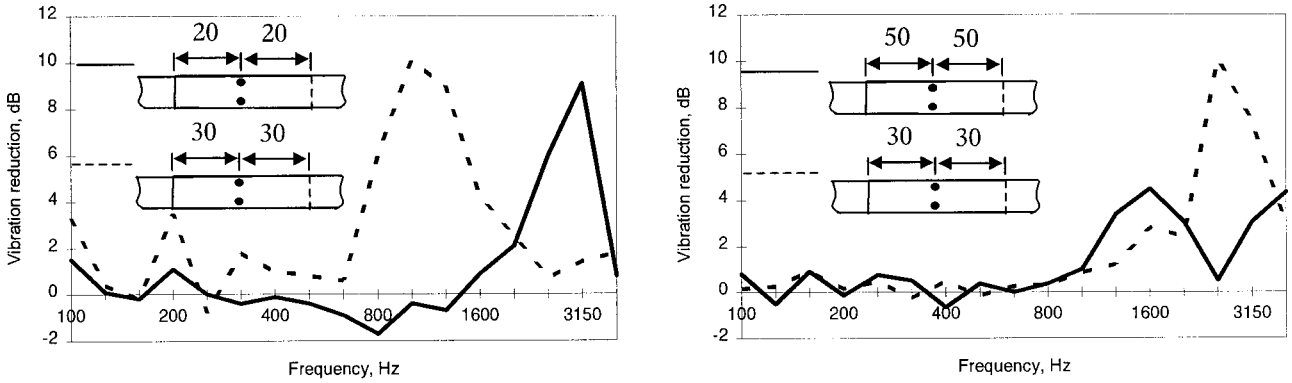


FIG. 10. Vibration insertion loss in finite beams. Left: 1-mm thick beam; Right: 2-mm thick beam.

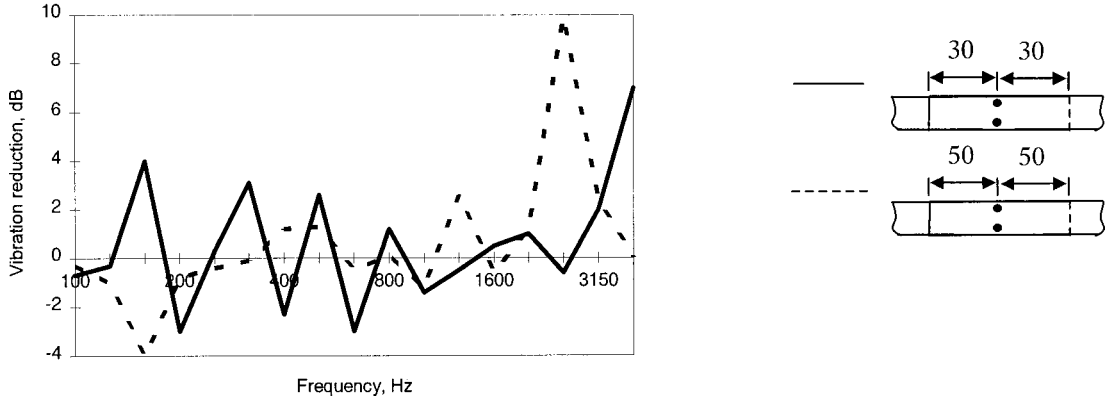


FIG. 11. Vibration insertion loss for 3-mm beam.

performance in a finite system. The two ends of the beam are clamped by two vises with 6-mm-thick rubber pads at the jaws, in order to introduce a certain damping at the ends as a practical structure does. The loss factor of the whole system in this case is about 3 percent in the frequency range of interest. In this way the samples can also be easily replaced.

In order to reduce the influence of the extra weight introduced by the bolts and the overlapped part of the beam, a joint with 10-mm distance between bolts and the edge (see the upper left of Fig. 3) is used to serve as the reference case. This joint introduces virtually no transmission loss and has a modal behavior similar to the test cases. Since the vibrational insertion loss compares spectra of two finite systems, narrow-band analysis will yield a big fluctuation in results due to small shifts of resonant peaks. Hence, the results are presented in 1/3-octave band.

The vibrational insertion losses of the joints in 1- and 2-mm beams are illustrated in Fig. 10. The peaks due to the resonance of the cantilevers are still seen in that figure, although much lower than those in transmission losses, as one can expect. The vibrational energy in this case is absorbed at the edge and at the joint itself. In order that a reflective type of joint has an effect of vibration reduction in a finite system, a certain amount of damping is necessary. Fortunately, most practical structures can fulfill this requirement. When the damping at the beam ends approaches infinity, the vibrational insertion loss approaches the transmission loss measured previously.

One may also notice that the effective length of the cantilever seems shortened in a finite system: the peaks appear at higher frequencies than those in corresponding transmission losses. However, the relative position of the peak is still in inverse proportion to the square of the relative length, indicating that the insertion loss measured is due to the resonance corresponding to the wavelength.

The vibrational insertion loss of a 3-mm beam is shown in Fig. 11, where the low-frequency fluctuation is due to the measurement error, since the beam is too short for a 3-mm-thick one if we want to avoid the near-field influence.

IV. CONCLUSIONS

The distance between the bolts and the edge of the beam is vital to the vibration transmission loss. With properly designed parameters, very high transmission loss can be expected at certain frequencies. The resonance of the cantilever is the reason for this high transmission loss, and the position of the peak is predictable. The vibration energy is mainly reflected back from the joint instead of being absorbed. It is possible to design a special washer in order to achieve a high transmission loss at the desired frequencies. For a finite system with a certain absorption, this type of joint can produce a rather high peak of vibrational insertion loss, but the effective length of the cantilever seems shortened due to the finiteness of the structure.

For a joint with two lines of bolts, a blunt peak of transmission loss can be expected. The mechanism of that is both reflection and absorption.

[1] L. Feng, ''An experimental method for vibrational insertion loss of mechanical joints,'' Acustica **87**, 191–198 (2001).
[2] L. Cremer and M. Heckl, *Structure-Borne Sound*, 2nd ed. (Springer, Berlin, 1988).
[3] A. Mazelet, ''Measurement of the reflection and transmission properties of junctions in beams,'' *Report TRITA-FKT 1997:28* (KTH, Stockholm, Sweden, 1997).

Acoustic waves generated by pulsed microwaves in viscoelastic rods: Modeling and experimental verification

Christophe Bacon, Emmanuel Guilliorit, and Bernard Hosten
Laboratoire de Mécanique Physique, Université Bordeaux 1, UMR 5469 CNRS 351, cours de la Libération, 33405 Talence Cedex, France

Dale E. Chimenti
Aerospace Engineering and Engineering Mechanics Department, and Center for NDE, Iowa State University, Ames, Iowa 50011-2271

(Received 14 December 1999; revised 8 February 2001; accepted 15 June 2001)

The acoustic wave generation in a specimen irradiated by a pulsed microwave is predicted theoretically. The specimen is a viscoelastic rod inserted into a wave guide. The model is based on Maxwell's equations, heat equation and thermoviscoelasticity theory. Computations show the presence of temperature oscillations due to the electromagnetic interferences in the irradiated rod if its electromagnetic absorption is low. An experimental method to infer indirectly the detailed behavior of microwave-generated acoustic waves in polymer rods, including the influence of electromagnetic wave reflection at the rod ends, is presented. The method consists of measuring the oscillations in the particle acceleration detected at the end of the rod that are induced by variations in the polymer rod length. The oscillations are caused by changing electromagnetic standing-wave conditions within the rod. It is found that these oscillations are in agreement in period, amplitude, and phase, with independent values of the complex dielectric constant and complex acoustic slowness of the polyvinyl chloride samples used in the study. [DOI: 10.1121/1.1391241]

PACS numbers: 43.40.Le, 43.40.Cw [RAR]

LIST OF SYMBOLS

a	lateral dimension of the rod
$a(z,t)$	axial acceleration
b	lateral dimension of the rod
C	specific heat
$\mathbf{E}(z,t)$	electric field intensity
$E^*(\omega)$	complex modulus
$\tilde{f}(z,\omega)$	Fourier transform of function $f(z,t)$ at cross-section z
$F(z,t)$	normal force
$\mathbf{H}(z,t)$	magnetic field intensity
k^*	complex electromagnetic wave number
L	rod length
m	TE mode number
m_a	accelerometer mass
n	TE mode number
$\tilde{N}(\omega)$	Fourier transforms of the strain at $z=0$ due to the wave travelling in the direction of decreasing z
P	instantaneous electromagnetic power
P_m	mean electromagnetic power
$\tilde{P}(\omega)$	Fourier transforms of the strain at $z=0$ due to the wave traveling in the direction of increasing z
$R(\omega)$	electromagnetic reflection coefficient
S	cross-sectional area of the rod
S^*	complex slowness
t	time
$T(\omega)$	electromagnetic transmission coefficient
T_0	period of the incident electromagnetic wave
$u(z,t)$	axial displacement
$v(z,t)$	axial particle velocity
z	axial coordinate
Z_a	accelerometer impedance
β	coefficient of linear thermal expansion
$\delta(\omega)$	Dirac distribution
$\varepsilon(z,t)$	longitudinal strain
ε^*	complex dielectric permittivity
μ_0	magnetic permeability
ν	frequency
ν_0	frequency of the incident electromagnetic wave
ω	angular frequency
ω_0	angular frequency of the incident electromagnetic wave
$\mathbf{\Pi}$	Poynting vector
$\theta(z,t)$	temperature rise
ρ	mass density of the rod
$\sigma(z,t)$	normal stress
τ	pulse width

I. INTRODUCTION

Acoustic wave generation by electromagnetic waves has been widely studied in the case of laser generated ultrasound. Hutchins has presented an extensive review of both experimental and theoretical works on this subject.[1] On the other hand, the literature on acoustic wave generation by thermal effects due to electromagnetic microwaves concerns essen-

tially the auditory sensations produced in men who heard a ''click'' or another sound when their head was irradiated by high-energy microwaves. A comprehensive review has been written by Chou and Guy[2] about this subject. Several mechanisms have been suggested to explain the phenomenon of microwave hearing, i.e., radiation pressure, electrostriction or thermal expansion. Now it is known that the main cause is the thermal expansion due to the microwave absorption: the microwave energy is absorbed and converted into thermal energy, which caused the irradiated material to expand in accord with its thermoelastic properties. Thus the acoustical waves are produced by a sudden heating of the irradiated material.

Microwave generation of acoustic waves was reported by Abeles[3] and later confirmed, and the results extended, by Zemel and Goldstein.[4–6] In these prior experiments the mechanism of acoustic generation was direct momentum transfer between electrons accelerated in the microwave rf field and the lattice ions in a discrete atomic solid or interfacial scattering of the electrons. Therefore, acoustic waves were produced at the same frequency as the incident microwave power (~10 GHz). The momentum transfer mechanism also dominates at lower frequencies, as demonstrated by Chimenti in *C*-band experiments on aluminum films at 400 MHz.[7]

In most studies concerned with acoustic wave generation by thermal effects caused by electromagnetic radiation (laser or microwaves), it is considered that the temperature rise decreases exponentially with increasing distance from the irradiated surface.[8–13] This statement is valid if the electromagnetic radiation absorption of the material is significant. For instance, in the case of laser-generated ultrasound, the penetration depth is generally very short compared to the thickness of the irradiated specimen. Consequently, the acoustic waves are generated in a small region very close to the irradiated surface, and the spatial temperature profile can be assumed to be exponential. This circumstance is identical for microwave irradiation only if the material absorption of microwaves is significant. For a lot of polymers, however, the microwave absorption is very low, and the penetration depth may be very large compared to the thickness of the specimen. Under these conditions, we have shown experimentally[14,15] for an irradiated polymer that the acoustic generation occurs not only near the irradiated surface, but also at the opposite surface or inside the rod. In previous papers, it was considered empirically that the spatial distribution of the temperature rise in an irradiated polymer rod was either uniform or linear[15,16] with a small variation, or exponential.[17] In these studies,[15–17] the technique was used to evaluate the Young's modulus of the irradiated material. The comparison with results obtained by an impact method on long instrumented rods showed a very good agreement.[15] However, it was not possible to predict the acoustic generation *a priori* from the microwave source power and from the electromagnetic, the thermal and the acoustic properties of the specimen.

First, the aim of the current paper is to predict the acoustic wave generation in a specimen irradiated by a pulsed microwave. The specimen is a viscoelastic rod inserted into a wave guide. If the microwave absorption is low, we will show that the temperature distribution is not exponential and that the model predicts the presence of temperature oscillations due to electromagnetic interference of waves reflected from the rod ends.

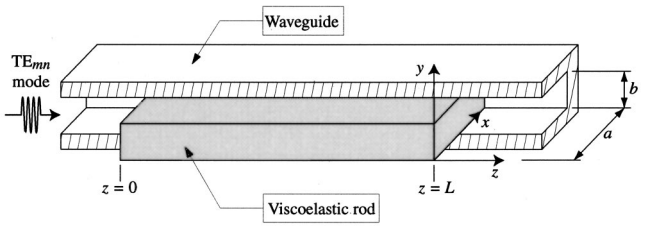


FIG. 1. Rod put in a rectangular wave guide (cut out of the wave guide).

Afterward, we will present an experimental method to test the theoretical model and to investigate indirectly the temperature oscillations by using an accelerometer mounted on a polyvinyl chloride (PVC) rod. After the presentation of the experimental arrangement, a new method will be described to determine the real part of the permittivity. Then, a complementary test is discussed that permits the evaluation of the remaining unknown characteristics of the material studied; these properties are the viscoelastic complex acoustic slowness and the imaginary part of the dielectric permittivity. The results will allow us to test the validity of the theoretical model in this case.

II. THEORY

Let us consider a rod which is straight, cylindrical, slender and made of linearly viscoelastic material. Its cross section is rectangular with lateral dimensions a and b and cross-sectional area S. Its density and length are ρ and L (see Fig. 1). The axial coordinate is z.

The rod is put in a rectangular wave guide of internal lateral dimensions a and b. There is enough play between the rod and the wave guide so that the rod can slide freely.

The rod is irradiated by a time-gated microwave at frequency ν_0 in TE$_{mn}$ electromagnetic mode. The pulse width is τ and the incident electromagnetic power (mean power) is P_m^i.

A. Determination of the temperature

To determine the temperature reached by the material we need knowledge of the instantaneous electromagnetic power.

Maxwell's equations, in the absence of currents and charges, are

$$\nabla\cdot\tilde{\mathbf{E}}=0,\quad \nabla\cdot\tilde{\mathbf{H}}=0,\quad \nabla\times\tilde{\mathbf{E}}=-\mu_0 i\omega\tilde{\mathbf{H}},$$
$$\nabla\times\tilde{\mathbf{H}}=\varepsilon^* i\omega\tilde{\mathbf{E}}, \tag{1}$$

where ω is the angular frequency related to the frequency ν by $\omega=2\pi\nu$, $\tilde{\mathbf{E}}$ the Fourier transform of the electric field intensity $\mathbf{E}$, $\tilde{\mathbf{H}}$ the Fourier transform of the magnetic field intensity $\mathbf{H}$, μ_0 the magnetic permeability and $\varepsilon^*=\varepsilon'-i\varepsilon''$ the complex dielectric permittivity of the material.

First, let us consider that the fields are due to a plane wave propagating in the positive z direction in an infinite wave guide filled with an absorbing material

$$\tilde{\mathbf{E}}^{+}(x,y,z,\omega)=\tilde{\mathbf{E}}_0^{+}(x,y,\omega)e^{-ik_z^* z}$$

and (2)

$$\tilde{\mathbf{H}}^{+}(x,y,z,\omega)=\tilde{\mathbf{H}}_0^{+}(x,y,\omega)e^{-ik_z^* z},$$

where $k_z^{*2}=\mu_0\varepsilon^*\omega^2-(k_x^2+k_y^2)=\mu_0\varepsilon^*\omega^2-k_t^2=(k_z'-ik_z'')^2$ is the complex wave number. The real part $k'(\omega)$ of k_z^* is an odd function of frequency, its imaginary part $-k''(\omega)$ is an even function.

Considering a TE mode ($E_z^+=0$) and a wave guide with a rectangular cross section, it can be demonstrated from the Maxwell's equations (1) that

$$\begin{aligned}
\tilde{E}_{0x}^{+}(x,y,\omega)&=-\frac{k_y}{k_x}\tilde{E}_0^{+}(\omega)c_x s_y,\\
\tilde{H}_{0x}^{+}(x,y,\omega)&=-\frac{k_z^*}{\omega\mu_0}\tilde{E}_0^{+}(\omega)s_x c_y,\\
\tilde{E}_{0y}^{+}(x,y,\omega)&=\tilde{E}_0^{+}(\omega)s_x c_y,\\
\tilde{H}_{0y}^{+}(x,y,\omega)&=-\frac{k_z^* k_y}{k_x\omega\mu_0}\tilde{E}_0^{+}(\omega)c_x s_y,\\
\tilde{E}_{0z}^{+}(x,y,\omega)&=0,\quad \tilde{H}_{0z}^{+}(x,y,\omega)=i\frac{k_t^2}{k_x\omega\mu_0}\tilde{E}_0^{+}(\omega)c_x c_y,
\end{aligned} \tag{3}$$

where $c_x=\cos k_x x$, $c_y=\cos k_y y$, $s_x=\sin k_x x$, $s_y=\sin k_y y$ with $k_x=m\pi/a$ and $k_y=n\pi/b$ for the TE_{mn} mode.

For a wave propagating in the negative z direction the electric and magnetic fields are deduced from Eqs. (3) by substituting k_z^* for $-k_z^*$ and $\tilde{E}_0^{+}(\omega)$ for $\tilde{E}_0^{-}(\omega)$. Consequently, for two waves propagating in both directions, the electric and magnetic fields are given by

$$\begin{aligned}
\tilde{E}_x&=-\frac{k_y}{k_x}c_x s_y(\tilde{E}_0^{+}e^{-ik_z^* z}+\tilde{E}_0^{-}e^{ik_z^* z}),\\
\tilde{H}_x&=-\frac{k_z^*}{\omega\mu_0}s_x c_y(\tilde{E}_0^{+}e^{-ik_z^* z}-\tilde{E}_0^{-}e^{ik_z^* z}),\\
\tilde{E}_y&=s_x c_y(\tilde{E}_0^{+}e^{-ik_z^* z}+\tilde{E}_0^{-}e^{ik_z^* z}),\\
\tilde{H}_y&=-\frac{k_z^* k_y}{k_x\omega\mu_0}c_x c_y(\tilde{E}_0^{+}e^{-ik_z^* z}-\tilde{E}_0^{-}e^{ik_z^* z}).
\end{aligned} \tag{4}$$

For $z<0$ the wave guide is filled with a semi-infinite medium called the incident medium (superscript i). The continuity of the tangential components of the electric and magnetic fields at the interface $z=0$ between the incident medium and the rod gives

$$\tilde{E}_0^{+}(k_z^*+k_z^{i*})+\tilde{E}_0^{-}(k_z^{i*}-k_z^*)=2k_z^{i*}\tilde{E}_0^{i+}. \tag{5}$$

For $z>L$ the wave guide is filled with a semi-infinite medium called the transmitted medium (superscript t). The continuity of the tangential components of the electric and magnetic fields at the interface $z=L$ between the rod and the transmitted medium defines a reflection coefficient $R(\omega)$ given by

$$R(\omega)=\frac{\tilde{E}_0^{-}}{\tilde{E}_0^{+}}=\frac{(k_z^*-k_z^{t*})e^{-2ik_z^* L}}{k_z^*+k_z^{t*}}. \tag{6}$$

If the transmitted-wave medium is a perfectly conducting medium, the reflection coefficient $R(\omega)$ must be replaced by

$$R(\omega)=-e^{-2ik_z^* L}. \tag{7}$$

A transmission coefficient $T(\omega)$ can be defined, from Eqs. (5) and (6),

$$\begin{aligned}
T(\omega)=\frac{\tilde{E}_0^{+}}{\tilde{E}_0^{i+}}&=[2(k_z^*+k_z^{t*})k_z^{i*}]/[(k_z^*+k_z^{i*})(k_z^*+k_z^{t*})\\
&\quad+(k_z^*-k_z^{t*})(k_z^{i*}-k_z^*)e^{-2ik_z^* L}].
\end{aligned} \tag{8}$$

Or, if the transmitted-wave medium is perfectly conducting, from Eqs. (5) and (7),

$$T(\omega)=\frac{2k_z^{i*}}{(k_z^*+k_z^{i*})-(k_z^{i*}-k_z^*)e^{-2ik_z^* L}}. \tag{9}$$

Then, Eq. (4) can be written

$$\begin{aligned}
\tilde{E}_x&=-\frac{k_y}{k_x}c_x s_y A(z,\omega)\tilde{E}_0^{i+},\\
\tilde{H}_x&=-\frac{s_x c_y}{\mu_0}\frac{k_z^* B(z,\omega)}{\omega}\tilde{E}_0^{i+},\\
\tilde{E}_y&=s_x c_y A(z,\omega)\tilde{E}_0^{i+},\quad \tilde{H}_y=-\frac{k_y c_x s_y}{k_x\mu_0}\frac{k_z^* B(z,\omega)}{\omega}\tilde{E}_0^{i+},
\end{aligned} \tag{10}$$

where the function $A(z,\omega)$ and $B(z,\omega)$ are defined by

$$A(z,\omega)=T(\omega)[e^{-ik_z^* z}+R(\omega)e^{ik_z^* z}]$$

and (11)

$$B(z,\omega)=T(\omega)[e^{-ik_z^* z}-R(\omega)e^{ik_z^* z}].$$

It can be noted that

$$\frac{\partial A}{\partial z}=-ik_z^* B\quad\text{and}\quad\frac{\partial B}{\partial z}=-ik_z^* A. \tag{12}$$

The instantaneous power is given by

$$P(z,t)=\int_S \mathbf{\Pi}\cdot\vec{e}_z\,dS, \tag{13}$$

where $\Pi=\mathbf{E}\times\mathbf{H}$ is the Poynting vector and S is the internal cross section of the wave guide. In the frequency domain the Fourier transform of the instantaneous power is written using the convolutions of the fields

$$\tilde{P}(z,\omega)=\frac{1}{2\pi}\int_S(\tilde{E}_x*\tilde{H}_y-\tilde{E}_y*\tilde{H}_x)dS, \tag{14}$$

where $\tilde{E}_x*\tilde{H}_y$ stands for the convolution of $\tilde{E}_x$ and $\tilde{H}_y$ defined by

$$\tilde{E}_x * \tilde{H}_y = \int_{-\infty}^{+\infty} \tilde{E}_x(\Omega)\tilde{H}_y(\omega-\Omega)d\Omega.$$

Introducing Eqs. (10) in Eq. (14) leads to

$$\tilde{P}(z,\omega) = \frac{1}{2\pi\mu_0} f_{xy}[A(z,\omega)\tilde{E}_0^{i+}(\omega)]^* \left[\frac{k_z^*(\omega)B(z,\omega)}{\omega}\tilde{E}_0^{i+}(\omega)\right], \tag{15}$$

where the number f_{xy} is

$$f_{xy} = \int_S \left(\frac{k_y^2}{k_x^2}c_x^2 s_y^2 + s_x^2 c_y^2\right) dS = \begin{cases} \dfrac{ab}{2} & \text{for TE}_{m0} \text{ modes,} \\ \dfrac{ab}{4}\dfrac{k_t^2}{k_x^2} & \text{for TE}_{mn} \text{ modes } (m,n \geqslant 1). \end{cases} \tag{16}$$

By using relations (12) the derivative of the power (which will be used for the calculation of the temperature), is

$$\frac{\partial \tilde{P}}{\partial z} = -\frac{i}{2\pi\mu_0} f_{xy}\left[(k_z^* B\tilde{\mathbf{E}}_0^{i+})^*\left(\frac{k_z^* B}{\omega}\tilde{\mathbf{E}}_0^{i+}\right) + (A\tilde{\mathbf{E}}_0^{i+})^*\left(\frac{k_z^{*2}A}{\omega}\tilde{\mathbf{E}}_0^{i+}\right)\right]. \tag{17}$$

For an incident sinusoidal wave at an angular frequency $\omega_0 = 2\pi\nu_0$, the field can be written

$$\tilde{\mathbf{E}}_0^{i+}(\omega) = E_0^{i+}\frac{\pi}{i}[\delta(\omega-\omega_0) - \delta(\omega+\omega_0)], \tag{18}$$

where $\delta(\omega)$ is the Dirac function. For one period T_0 it can be demonstrated that the mean values of the power and its derivative are

$$P_m(z) = \frac{1}{T_0}\int_0^{T_0} P(z,t)dt = \frac{f_{xy}}{4\mu_0\omega_0}(E_0^{i+})^2[A(z,-\omega_0)k_z^*(\omega_0)B(z,\omega_0) - A(z,\omega_0)k_z^*(-\omega_0)B(z,-\omega_0)], \tag{19}$$

$$\left.\frac{\partial P}{\partial z}\right|_m(z) = \frac{1}{T_0}\int_0^{T_0}\frac{\partial P(z,t)}{\partial z}dt = \frac{if_{xy}}{4\mu_0\omega_0}(E_0^{i+})^2 A(z,\omega_0)A(z,-\omega_0) \times[k_z^{*2}(-\omega_0) - k_z^{*2}(\omega_0)]. \tag{20}$$

The mean power transported by the incident wave in a nonabsorbing incident medium is calculated by writing $T(\omega) = 1$ and $R(\omega) = 0$ in Eqs. (11). We obtain from equation (19)

$$P_m^i = \frac{k_z^{i\prime}(\omega_0)}{2\mu_0\omega_0} f_{xy}(E_0^{i+})^2. \tag{21}$$

The microwave sinusoidal train is emitted during a very short time inducing a short thermal pulse so that heat conduction can be neglected. Then, the heat equation is

$$\rho C\frac{\partial\theta}{\partial t} = -\frac{1}{S}\frac{\partial P}{\partial z}, \tag{22}$$

where θ is the temperature rise, and C the specific heat. As a consequence the temperature rise after one period is given by

$$\theta_{T_0}(z) = -\frac{1}{\rho CS}\int_0^{T_0}\frac{\partial P(z,t)}{\partial z}dt = -\frac{2\pi}{\rho CS\omega_0}\left.\frac{\partial P}{\partial z}\right|_m(z). \tag{23}$$

Thus, by using Eqs. (20) and (21), this temperature rise can be written

$$\theta_{T_0}(z) = \frac{i\pi P_m^i}{\rho CS\omega_0 k_z^{i\prime}(\omega_0)}A(z,\omega_0)A(z,-\omega_0) \times[k_z^{*2}(\omega_0) - k_z^{*2}(-\omega_0)]. \tag{24}$$

Since the period T_0 is generally very short compared to the pulse width τ, the temperature rise is

$$\theta(z,t) = \frac{\theta_{T_0}(z)}{T_0}\tau\cdot r(t) = \frac{\theta_{T_0}(z)\omega_0}{2\pi}\tau\cdot r(t), \tag{25}$$

where $r(t)$ is a function given by

$$r(t) = \begin{cases} 0 & \text{for } t \leqslant 0, \\ t/\tau & \text{for } 0 < t < \tau, \\ 1 & \text{for } t \geqslant \tau. \end{cases} \tag{26}$$

In the following, it is convenient to write the Fourier transform of the temperature rise in the following way:

$$\tilde{\theta}(z,\omega) = \tilde{r}(\omega)K\sum_{j=1}^{4} A_j e^{a_j z}, \tag{27}$$

where $\tilde{r}(\omega)$ is the Fourier transform of the function $r(t)$ defined in Eq. (26),

$$K = \frac{2P_m^i\tau k_z'(\omega_0)k_z''(\omega_0)}{\rho CSk_z^{i\prime}(\omega_0)}T(\omega_0)T(-\omega_0) \tag{28}$$

and

$$\begin{aligned} &A_1 = 1, \quad a_1 = -i[k_z^*(\omega_0) + k_z^*(-\omega_0)] = -2k_z''(\omega_0), \\ &A_2 = R(-\omega_0), \\ &a_2 = i[k_z^*(-\omega_0) - k_z^*(\omega_0)] = -2ik_z'(\omega_0), \\ &A_3 = R(\omega_0), \quad a_3 = -a_2, \\ &A_4 = R(\omega_0)R(-\omega_0), \quad a_4 = -a_1. \end{aligned} \tag{29}$$

B. Determination of the acoustical waves

According to the classical theory of thermoviscoelasticity,[18] the normal stress $\sigma(z,t)$ and the longitudinal strain $\varepsilon(z,t)$ are connected through the convolution relation

$$\sigma(z,t)=\int_{-\infty}^{t}E(t-\tau)\frac{\partial}{\partial\tau}[\varepsilon(z,\tau)-\beta\theta(z,\tau)]d\tau$$

$$=\frac{\partial E(t)}{\partial t}*[\varepsilon(z,t)-\beta\theta(z,t)], \tag{30}$$

where β is the coefficient of linear thermal expansion. In the frequency domain, this relation becomes

$$\tilde{\sigma}(z,\omega)=E^*(\omega)\tilde{\varepsilon}(z,\omega)-E^*(\omega)\beta\tilde{\theta}(z,\omega), \tag{31}$$

where $E^*(\omega)$ is the complex Young's modulus of the material, $\tilde{\sigma}(z,\omega)$, $\tilde{\varepsilon}(z,\omega)$, and $\tilde{\theta}(z,\omega)$ are the Fourier transforms of the normal stress, the longitudinal strain, and the temperature rise above the initial temperature, respectively.

If the lateral dimensions of the rod are much shorter than the wavelength, the radial inertia can be neglected. Then, the Fourier transforms, $\tilde{\sigma}(z,\omega)$ and $\tilde{u}(z,\omega)$, of the normal stress $\sigma(z,t)$ and the axial displacement $u(z,t)$, respectively, are related by

$$\frac{\partial}{\partial z}\tilde{\sigma}(z,\omega)=-\rho\omega^2\tilde{u}(z,\omega). \tag{32}$$

To facilitate the analysis the temperature distribution will be assumed to be independent of the mechanical state of the material. From Eqs. (31) and (32) it can be demonstrated that the Fourier transforms of the axial displacement $\tilde{u}(z,\omega)$, the axial particle velocity $\tilde{v}(z,\omega)$, and the axial acceleration $\tilde{a}(z,\omega)$ can be calculated if the longitudinal strain and the temperature rise are known, i.e.,

$$\tilde{u}(z,\omega)=\frac{1}{\omega^2S^{*2}}\left[\beta\frac{\partial\tilde{\theta}(z,\omega)}{\partial z}-\frac{\partial\tilde{\varepsilon}(z,\omega)}{\partial z}\right],$$
$$\tilde{v}(z,\omega)=i\omega\tilde{u}(z,\omega),\quad \tilde{a}(z,\omega)=-\omega^2\tilde{u}(z,\omega), \tag{33}$$

where $S^*=S'-iS''$ is the complex slowness of the material given by $S^*=\sqrt{\rho/E^*}$. Assuming the small-strain hypothesis, the longitudinal strain is related to the axial displacement by $\varepsilon(z,t)=\partial u(z,t)/\partial z$. Consequently, Eq. (32) becomes

$$\frac{\partial^2\tilde{\varepsilon}(z,\omega)}{\partial z^2}+\omega^2S^{*2}\tilde{\varepsilon}(z,\omega)-\beta\frac{\partial^2\tilde{\theta}(z,\omega)}{\partial z^2}=0. \tag{34}$$

From Eq. (27) this last equation can be written

$$\frac{\partial^2\tilde{\varepsilon}(z,\omega)}{\partial z^2}+\omega^2S^{*2}\tilde{\varepsilon}(z,\omega)-\beta\tilde{r}(\omega)K\sum_{j=1}^{4}a_j^2A_je^{a_jz}=0. \tag{35}$$

The general solution of this equation is

$$\tilde{\varepsilon}(z,\omega)=\tilde{P}(\omega)e^{-i\omega S^*z}+\tilde{N}(\omega)e^{i\omega S^*z}$$
$$+\beta\tilde{r}(\omega)K\sum_{j=1}^{4}\frac{a_j^2A_j}{a_j^2+\omega^2S^{*2}}e^{a_jz}, \tag{36}$$

with $\tilde{P}(\omega)$ and $\tilde{N}(\omega)$ being two functions depending on the boundary conditions at the ends of the viscoelastic rod. From Eqs. (31) and (33), we can write the normal stress and the acceleration in function of $\tilde{P}(\omega)$ and $\tilde{N}(\omega)$. In the four expressions $\tilde{\sigma}(0,\omega)$, $\tilde{\sigma}(L,\omega)$, $\tilde{a}(0,\omega)$, and $\tilde{a}(L,\omega)$, we eliminate $\tilde{P}(\omega)$ and $\tilde{N}(\omega)$ to obtain relations between stresses and accelerations at the ends of the rod. Therefore, these equations are independent of the boundary conditions

$$\begin{bmatrix}\tilde{\sigma}(L,\omega)\\ \tilde{a}(L,\omega)\end{bmatrix}=P\begin{bmatrix}\tilde{\sigma}(0,\omega)\\ \tilde{a}(0,\omega)\end{bmatrix}+G, \tag{37}$$

where

$$P=\begin{bmatrix}\cos\omega S^*L & \dfrac{\rho}{\omega S^*}\sin\omega S^*L\\ -\dfrac{\omega S^*}{\rho}\sin\omega S^*L & \cos\omega S^*L\end{bmatrix},$$

$$G=\beta\tilde{r}K\omega^2\left[\frac{\rho}{\omega S^*}\sum_{j=1}^{4}\frac{A_j}{a_j^2+\omega^2S^{*2}}(\omega S^*\cos\omega S^*L+a_j\sin\omega S^*L-\omega S^*e^{a_jL})\right.$$
$$\left.\times\sum_{j=1}^{4}\frac{A_j}{a_j^2+\omega^2S^{*2}}(-\omega S^*\sin\omega S^*L+a_j\cos\omega S^*L-a_je^{a_jL})\right]. \tag{38}$$

It is noteworthy that the matrix P is related to the wave propagation while the vector G is related to the wave generation due to the thermal expansion. With this form, the acceleration at $z=L$ can be obtained from the boundary conditions. The rod end at $z=0$ is free. Consequently, the normal stress at $z=0$ must be zero, i.e., $\tilde{\sigma}(0,\omega)=0$.

The rod is instrumented with an accelerometer at the end $z=L$. The accelerometer has a mass m_a. Thus, the normal force at this end is related to the axial velocity by

$$\tilde{F}(L,\omega)=-Z_a\overline{\nu}(L,\omega)\quad\text{or}\quad S\tilde{\sigma}(L,\omega)=\frac{iZ_a}{\omega}\tilde{a}(L,\omega), \tag{39}$$

where Z_a is the mechanical impedance of the accelerometer. At low frequencies the accelerometer can be assumed to be a rigid mass. Then, its mechanical impedance is given by $Z_a=im_a\omega$. With the help of these boundary conditions and Eq. (38), the acceleration at $z=L$ is

$$\tilde{a}(L,\omega)=\frac{P_{22}G_1-P_{12}G_2}{\dfrac{iZ_a}{S\omega}P_{22}-P_{12}}. \tag{40}$$

III. NUMERICAL RESULTS

As examples of applications, the accelerations and the displacements at the end of an irradiated rod are predicted in two different cases: for a very long rod and for a short one.

For both rods, the same mechanical characteristics are used in order to focus on the effect of the electromagnetic wave interference. These characteristics are close to the characteristics of a polymeric material (such as PVC). To simplify, it is assumed that the complex slowness is constant in the considered frequency range ($S'=600\,\mu$s/m, $S''=10\,\mu$s/m). Actually, this fact is impossible since the complex slowness must depend on the frequency owing to the dispersion due to the radial inertia and the dispersion due to the viscoelastic effect.[19] This hypothesis is an approximation. The density, the specific heat and the coefficient of linear

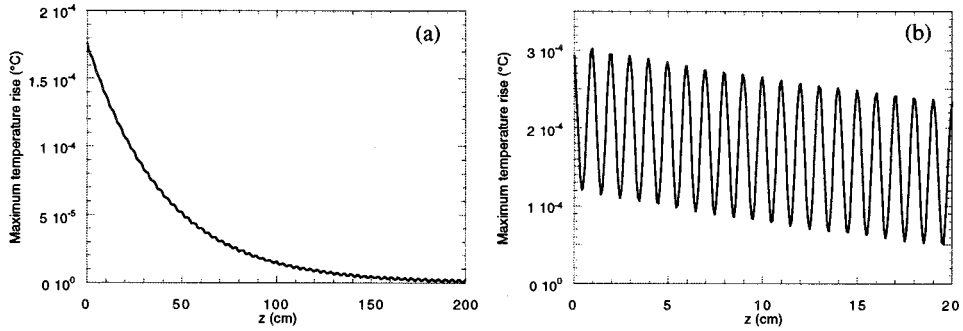


FIG. 2. Maximum temperature rises for the long rod [(a) $L=2$ m] and the short rod [(b) $L=20$ cm].

thermal expansion used for the calculations equal 1380 kg/m^3, 960 J/kg/°C, and 80.10^{-6}/°C, respectively. The lengths of the irradiated specimens are 20 cm and 2 m. The mass of the accelerometer is neglected so that Z_a is zero. The characteristics of the microwave generator are the same as the ones of the generator used for the experimental verification (see Sec. IV A).

The complex electromagnetic permittivities are identical for the both irradiated rods. The real and imaginary relative permittivities ε_r' and ε_r'' are 3 and 0.02, respectively.

The maximum temperature rises predicted from Eq. (25) are plotted in Fig. 2 for both rods.

It can be seen that, as it could be expected, the temperature variation is exponential if the rod is very long [see Fig. 2(a)]. However, in the case of the short irradiated rod [see Fig. 2(b)], some oscillations occur. These oscillations are generated by the interferences of the electromagnetic waves which are reflected on the opposite ends of the irradiated rod. In the case of the long rod, the microwaves are almost completely absorbed before reflecting at the end of the rod.

The accelerations at the end $z=L$ predicted from Eq. (40) are plotted in Fig. 3. We can notice in Fig. 3(a) that the larger peak (noted G_0) corresponds to the generation of the wave at the end $z=0$. This wave propagates in the rod and is reflected at the end $z=0$ and leads to the second peak (R_0), and so on. Before the peak G_0, the acceleration is caused by the generation in the bulk of the rod. In Fig. 3(b), the same peaks are visible (G_0 and R_0) but a series of other peaks appears. The peak G_L corresponds to the wave generated at the end $z=L$, that is reflected at the end $z=L$ and leads to the peak R_L. It is noticeable that the bulk generation is now well influenced by the interferences of the electromagnetic waves inside the rod and produces high frequency acoustic oscillations. They are caused by the oscillations of the temperature rise which were predicted in Fig. 2(b). The peak at high frequencies in the acceleration spectrum is due to the same phenomenon [see Fig. 4(b)].

An important remark is that the penetration depth effect that was assumed in Ref. 14 (Sec. D) is not predicted by the present model. The previous measurements in Ref. 14 were probably affected by the temperature rise of the material submitted to very high power microwaves. Hence the sound speed in the material changed, that can explain the variation of the arrival time.

The following section will be essentially dedicated to the experimental verification of this new model. The presence of the temperature oscillations due to the electromagnetic interferences will be particularly investigated.

IV. EXPERIMENTAL VERIFICATION

The aim of this section is to test experimentally the new model. The model predicts the presence of temperature oscillations due to electromagnetic interference, in the case of low microwave absorption, of waves reflected from the rod ends. The main problems with experimental detection of this effect are that the particle displacements associated with these temperature oscillations are extremely small (<1 nm) and that they occur in a frequency range (around 200 kHz for polymers) well beyond the sensitivity of the mechanical accelerometer used in prior studies. The upper frequency bandwidth of an accelerometer is typically limited to 15–20 kHz. One possible detection mechanism that was considered was the use of a capacitive transducer to measure directly the temperature-induced displacements. Unless such a capacitive transducer has a large bandwidth and a very high sensitivity,[20,21] however, the estimated displacements would be too small to be accurately characterized in this way. We present another indirect method in this section.

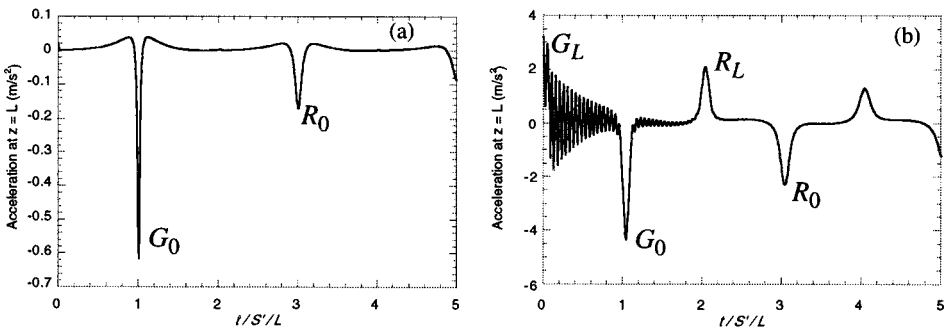


FIG. 3. Accelerations at $z=L$ for the long rod (a) and for the short rod (b) versus nondimensional time.

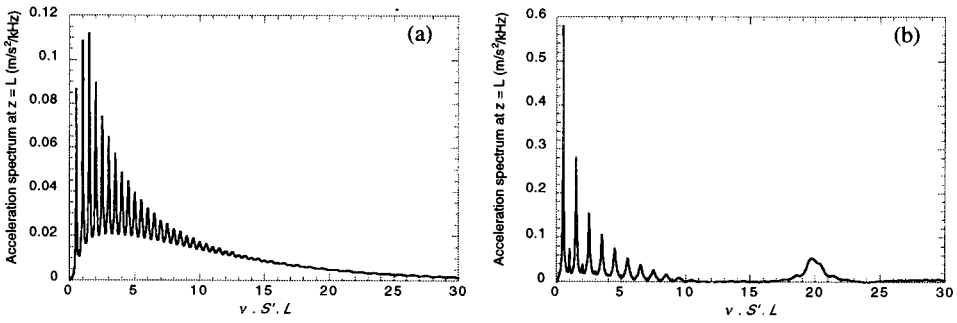


FIG. 4. Acceleration spectra for the long rod (a) and for the short rod (b) versus nondimensional frequency.

A. Experimental setup

The microwave tonebursts are produced by a time-gated electromagnetic generator at 9.41 GHz, as shown in Fig. 5. The maximum power of the incident electromagnetic wave is 5 kW, and the pulse width is 8.75 μs. The wave is delivered by a standard wave guide plumbing (of cross section 23 $\times$10 mm) that permits only the fundamental wave guide mode TE_{10} to propagate. To obtain the maximum power from the generator, a hybrid tee impedance adaptor is inserted into the circuit and adjusted to achieve a minimum voltage standing wave ratio (VSWR), as measured by a wattmeter.

The PVC rod sample is held vertically inside the wave guide by means of a piece of foam put at the end of the wave guide. The cross section of the rod is almost identical to the cross section of the wave guide, but there is enough play between them so that the specimen can slide freely. A piezoelectric miniature accelerometer (Brüel & Kjaer 4374) is attached to the end of the rod by means of a very thin layer of ultrasonic couplant. The accelerometer's mass m is 0.65 g and its charge sensitivity is 0.129 pC/m s^{-2}. Its dynamic frequency range lies between about 1 Hz and 20 kHz. The signal from the accelerometer is amplified by a charge amplifier (Brüel & Kjaer 2525), and the amplified acceleration signals are recorded on a digital oscilloscope (Lecroy 9310) that is, in turn, linked to a computerized data acquisition system. The signals are averaged in the scope for 50 sweeps in order to eliminate the electronic noise and the environmental vibrations. Then, the signals are transferred to a computer for further processing.

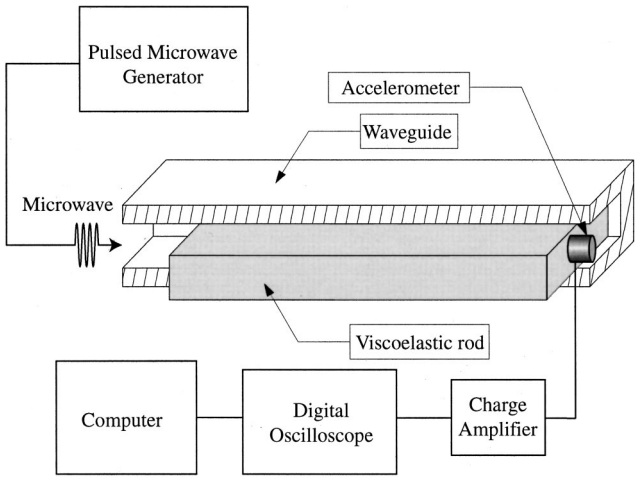


FIG. 5. Experimental setup (cut out of the wave guide).

B. Determination of the real part of the permittivity

For the case of a short rod of PVC irradiated by microwaves, the model presented in the preceding sections predicts spatial oscillations of temperature in the rod caused by the interferences of the electromagnetic waves reflected from opposite ends of the irradiated rod. This phenomenon is manifested as a peak at high frequencies in the acceleration spectrum, and the experimental characterization of this peak would allow one to determine the electromagnetic properties of the specimen if the phase velocity in the rod is known at any frequency. Although the frequency of the peak is too high to be measured with an accelerometer, the acoustic effect of the oscillations described above can be investigated at lower frequencies by another method. The idea is based on the fact that the period of the temperature oscillations does not depend on the length of the irradiated rod. Rather, the amplitude of the acceleration generated is strongly influenced by the temperature at its two ends because the effect is produced by standing electromagnetic waves. As a consequence, a variation of the length of the rod should lead to oscillations of the amplitudes of the acceleration signals related to the temperature oscillations in the rod.

In order to illustrate this phenomenon, four simulations are performed in Fig. 6 for a polymeric material (such as PVC). For the sake of simplicity, it is assumed that the complex slowness is constant in the considered frequency range ($S'=600\ \mu$s/m, $S''=10\ \mu$s/m, corresponding to a Young's modulus of $E'=3.83$ GPa and $E''=0.13$ GPa). The real and imaginary relative permittivities are $\varepsilon_r'=3$ and $\varepsilon_r''=0.02$, respectively. The density, the specific heat and the coefficient of linear thermal expansion used for calculations are 1380 kg m^{-3}, 960 J kg^{-1} °C^{-1} and 80×10^{-6} °C^{-1}, respectively. The length of the material varies from 2.5 to 25 cm, while its lateral dimensions are a constant 23×10 mm.

The amplitude $a_{\max}(L)$ of the acceleration peak identified in Fig. 6(a) is plotted versus rod length in Fig. 6(b). We identify λ_1 as the period corresponding to the oscillations in the curve of the amplitude of the first acceleration spectrum peak versus rod length [Fig. 6(a)]. Figure 6(c) presents the maximum temperature rise versus coordinate distance along the rod. The period of the spatial temperature rise is λ_2. Finally, λ_1 and λ_2 are compared in Fig. 6(d).

As we can see in Fig. 7, the period λ_2 of the temperature

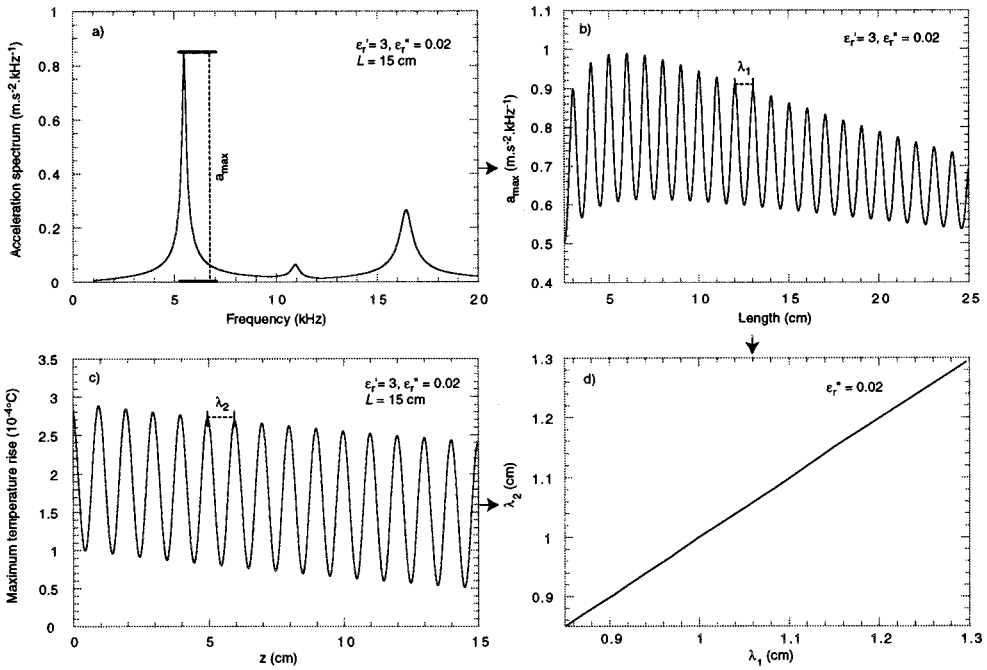


FIG. 6. (a) Acceleration spectrum. (b) Amplitude of the first peak versus rod length. (c) Maximum temperature rise. (d) Comparison between λ_1 and λ_2.

oscillations is nearly independent of the imaginary part of the permittivity ε_r'' (for the case of low absorption). For this reason a measurement of the period λ_2 permits the direct determination of the real part ε_r' of the material permittivity.

In order to measure the real part ε_r' of the relative permittivity of PVC, the acceleration caused by the acoustic waves generated in rods for several lengths must be acquired. The same PVC rod was used for all measurements after progressively reducing its length by cutting one edge. Figure 8 shows the values of the amplitudes of the first peak in the acceleration spectra versus rod length. The experimental data correspond to 58 different lengths of PVC rods from 7.6 to 11 cm. The error bars for each measurement correspond to the dispersion obtained for five measurements. For the theoretical curve the characteristics of the PVC found in the next section are used. Indeed, although the thermal period λ_2 depends only on the value of ε_r', the amplitude $a_{\max}(L)$ of the first peak in the acceleration spectrum versus the rod length depends on the other electromagnetic, thermal, and mechanical characteristics.

As expected, oscillations corresponding to the interferences of reflected electromagnetic wave appear in the measured data. To achieve the closest agreement with experiment, we adjust the value of ε_r' in successive comparisons to the experimental results. By successive approximation and exploitation of error we estimate that, for the case shown in Fig. 8, the optimal value is $\varepsilon_r' = 2.67 \pm 0.03$. The accuracy of the fitted value may be attributed to the high sensitivity of ε_r' with respect to the oscillation period of the acceleration spectrum maxima in Fig. 6(a).

C. Determination of the other mechanical and electromagnetic properties

To test the theoretical model thoroughly, the amplitudes of the theoretical and the experimental curves plotted in Fig.

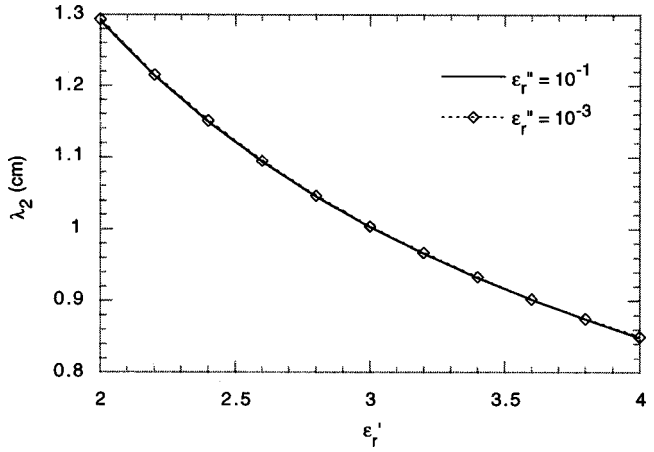


FIG. 7. Period of the temperature oscillations versus real part of the permittivity for two different values of the imaginary part.

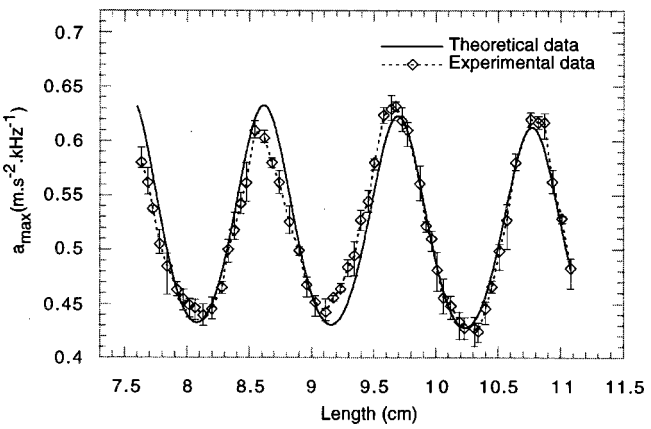


FIG. 8. Maxima of the first peaks of acceleration spectra versus rod length (for PVC).

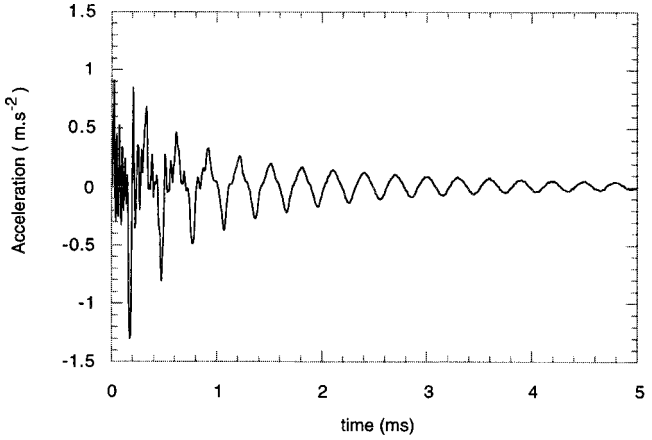


FIG. 9. Experimental acceleration signal.

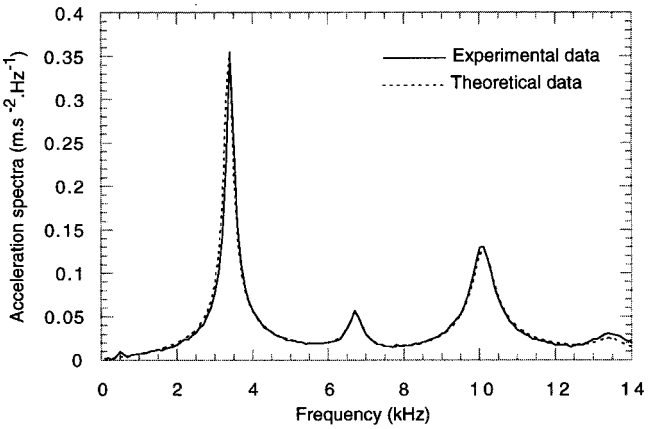


FIG. 10. Experimental and theoretical acceleration spectra (PVC, $L = 23.3$ cm).

8 must be compared. For now, even though the complete theoretical curve is plotted in Fig. 8, comparison between the theoretical and the experimental values of the period λ_2 involves the behavior only of ε_r'. The comparison of the amplitudes and the phases of the curves in Fig. 8, as well as the periodicities, introduces additionally the behavior of the imaginary part of the permittivity and that of the viscoelastic characteristics of the solid (the coefficient of thermal expansion, the density, the specific heat, and the microwave generator power are determined independently by means of other tests). To obtain values of these additional parameters, complementary tests are performed on a PVC rod having a length of 23.3 cm. The longer rod allows us to observe more peaks within the frequency bandwidth of the accelerometer and consequently to achieve better accuracy. The acceleration signal of the acoustic waves generated inside the PVC is given in Fig. 9. The high frequency oscillations at times less than 1 ms are the result of spurious accelerometer resonances.

To compare experimental and theoretical spectra, a minimization procedure is used. The calculations are carried out using three adjustable parameters (the real and imaginary parts of the complex acoustic slowness and the imaginary part of the dielectric constant). The generator power is a known quantity, and the real part of the permittivity is determined by the tests described above. The minimization procedure allows us to find an optimal set of parameters $\mathbf{p}$ that provides a best fit of the theoretical spectrum to the experimental one. The function we minimize is given by

$$J(\mathbf{p}) = \sum_{\nu} \; ||\tilde{a}_{\exp}(\nu)| - |\tilde{a}(\nu,\mathbf{p})||, \tag{41}$$

where $|\tilde{a}_{\exp}(\nu)|$ and $|\tilde{a}(\nu,\mathbf{p})|$ are the moduli of the Fourier transforms of the experimental and the theoretical accelerations, respectively, and ν is the frequency. The set of adjustable parameters is $\mathbf{p}=(S',S'',\varepsilon_r'')$. The model acceleration spectrum is varied by adjusting $\mathbf{p}$ to obtain a minimization of the function $J(\mathbf{p})$ above. It can be seen in Fig. 10 that the fitted spectrum is quite close to the experimental one.

The minimization method permits us to estimate the complex slowness and the imaginary part of the PVC dielectric permittivity. This fitting procedure is performed on the PVC rod of length 23.3 cm and for each specimen in which we have varied the length from 7.6 to 11 cm. Although the shorter rods evidence fewer peaks in their acceleration spectra within the accelerometer bandwidth, we present here the results of the minimization calculated for these shorter specimens because a higher number of data points (see Fig. 8) increases the accuracy of the method. We have verified that the results obtained for the longer rod are in good agreement with the set of parameters shown in Table I. The theoretical curve $a_{\max}(L)$ obtained with these parameters are plotted in Fig. 8 and compared with the experimental one. This figure shows a good agreement for any rod length, and we note that the phases of the experimental and the theoretical curves are also very close.

V. CONCLUSIONS

A model based on Maxwell's equations, heat equation, and thermoviscoelasticity theory has been established to compute the acoustic field generated in a rod irradiated by pulsed microwaves and embedded in a rectangular wave guide. Compared to the previous model presented in Ref. 15, the temperature oscillations and their acoustic effect can be investigated. The amplitude of the acceleration can be predicted *a priori*.

Moreover, we have presented a method to infer indirectly the detailed behavior of microwave-generated acoustic waves in polymer rods. Because the displacements associated with these wave motions are extremely small (<1 nm) and occur in a frequency range beyond the sensitivity of mechanical accelerometers, we have chosen to examine the behavior by employing an indirect inference method. We have measured the oscillations induced in the acceleration at lower frequencies by variations in the length of the polymer rod. It has been found that these oscillations are in agreement in period, amplitude, and phase, with independent values of the complex dielectric constant and complex acoustic slow-

TABLE I. Results obtained for PVC.

S' (μs m^{-1})	S'' (μs m^{-1})	ε_r'	$\varepsilon_r''(10^{-2})$
626 ± 8	14.9 ± 1.1	2.67 ± 0.03	1.89 ± 0.19

ness of the PVC samples used in the study. This finding demonstrates the completeness of the theoretical calculations and tends to confirm their validity. Combined with our earlier experimental investigations on detection of microwave-generated sound using gas-coupled acoustic techniques, this new means of sound wave generation is ready for technological exploitation in several potential applications.

ACKNOWLEDGMENTS

A part of this work was carried out during a visit by Dale E. Chimenti funded by the University Bordeaux 1. The authors are indebted to the *Conseil Régional d'Aquitaine* and the *Délégation Régionale à la Recherche et à la Technologie d'Aquitaine* for their financial support.

[1] D. A. Hutchins, *Ultrasonic Generation by Pulsed Laser*, edited by W. P. Mason and R. N. Thurston (Physical Acoustics, New York, 1988), Vol. XVII.
[2] C. K. Chou and A. W. Guy, "Auditory perception of radio-frequency electromagnetic fields," J. Acoust. Soc. Am. **71**, 1321–1334 (1982).
[3] B. Abeles, Phys. Rev. Lett. **19**, 1181 (1967).
[4] A. Zemel and Y. Goldstein, "Conversion of electromagnetic into acoustic energy via indium films," Phys. Rev. B **7**, 191–200 (1973).
[5] A. Zemel and Y. Goldstein, "Conversion of electromagnetic into acoustic energy using Sn, Au and Cu films at a low temperatures," Phys. Rev. B **9**, 1499–1505 (1974).
[6] Y. Goldstein, S. Barzilai, and A. Zemel, "Electron mean-free-path dependence of electromagnetic phonon generation," Phys. Rev. Lett. **32**, 463–466 (1974).
[7] D. E. Chimenti, "Nonlocal electromagnetic acoustic wave generation in aluminum films," Phys. Rev. B **13**, 4245–4251 (1976).
[8] J. E. Michaels, "Thermally induced elastic wave propagation in slender bars," *Proceedings of the 3rd U.S. National Congress on Applied Mechanics* (American Society of Mechanical Engineers, New York, 1958), p. 209.
[9] R. M. White, "Generation of elastic waves by transient surface heating," J. Appl. Phys. **34**, 3559–3567 (1963).
[10] L. S. Gournay, "Conversion of electromagnetic to acoustic energy by surface heating," J. Acoust. Soc. Am. **40**, 1322–1330 (1966).
[11] D. E. Borth and C. A. Cain, "Theoretical analysis of acoustic signal generation in materials irradiated with microwave energy," IEEE Trans. Microwave Theory Tech. **MTT-25**, 944–953 (1977).
[12] L. M. Lyamshev and B. I. Chelnokov, "Sound generation in a solid by penetrating radiation," Sov. Phys. Acoust. **29**, 220–225 (1983).
[13] F. Enguehard and L. Bertrand, "Effects of optical penetration and laser pulse duration on laser generated longitudinal acoustic waves," J. Appl. Phys. **82**, 1532–1538 (1997).
[14] B. Hosten and P. A. Bernard, "Ultrasonic wave generation by time-gated microwaves," J. Acoust. Soc. Am. **104**, 860–866 (1998).
[15] C. Bacon, B. Hosten, and P. A. Bernard, "Acoustic wave generation in viscoelastic rods by time-gated microwaves," J. Acoust. Soc. Am. **106**, 195–201 (1999).
[16] E. Guilliorit, C. Bacon, and B. Hosten, "Génération d'ondes acoustiques par micro-ondes impulsionnelles," *14ème Congrès Français de Mécanique*, No. 91 in CD-Rom, Toulouse, France, 1999.
[17] B. Hosten and C. Bacon, "Measurement of complex young moduli of composite materials by time-gated microwaves," in *Review of Progress in Quantitative Non Destructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 2000), Vol. 19, pp. 1113–1120.
[18] R. M. Cristensen, *Theory of Viscoelasticity: An Introduction* (Academic, New York, 1971).
[19] C. Bacon, "An experimental method for considering dispersion and attenuation in a viscoelastic Hopkinson bar," Exp. Mech. **38**, 242–249 (1998).
[20] M. N. Islam, C. Edwards, and S. B. Palmer, "Detection of laser generated Lamb waves in non-metals by capacitance transducers," Nondestr. Test. Eval. **11**, 9–20 (1994).
[21] D. Hutchins and J. D. Macphail, "A new design of capacitance transducer for ultrasonic displacement detection," J. Phys. E **18**, 69–72 (1985).

Modifying modal characteristics of sound fields by state feedback control

Toshiya Samejima
Department of Acoustic Design, Kyushu Institute of Design, 4-9-1, Shiobaru, Minami-ku, Fukuoka 815-8540, Japan

(Received 27 July 2000; revised 28 March 2001; accepted 25 May 2001)

State feedback control of a sound field is investigated for improving the irregular distribution of the natural frequencies at low frequencies. A state-space description of a sound field is derived from an inhomogeneous wave equation using the finite element method. The feedback control system is realized as a feedback filter between sensor microphones and control sources using a state estimator for a linear dynamic system. Pole allocation is employed for calculating the state feedback gain vector, such that the roots of the closed-loop system have the desired modal distribution. Computer simulations are performed to demonstrate the control achieved by distributing the lowest undamped natural frequencies in a uniform manner. As a result, the transfer functions with the state feedback control are modified so as to have peaks at identical intervals for the frequency range of interest. A control experiment in an enclosure is carried out with the same objective of control as modeled in the simulations. The experiment verifies that the desired modal distribution can be achieved by introducing the proposed state feedback control into a sound field. [DOI: 10.1121/1.1387096]

PACS numbers: 43.50.Ki, 43.55.Br, 43.55.Ka [MRS]

I. INTRODUCTION

Active sound field control, which involves the use of electroacoustic transducers to drive the output of an acoustic plant to a desired output, is useful in low-frequency applications because passive methods such as installing sound absorbers are not effective at low frequencies. Active sound field control technology has been developed and studied by a number of researchers, and has primarily involved the use of methods based on inverse filtering of the transfer functions from sound sources to receiving points in a sound field and simulating the desired transfer functions as strictly as possible. However, in addition to dealing with transfer functions from sound sources to receiving points, it is necessary to regard a full sound field as a plant and to develop a method to alter plant dynamics.

The full sound field can be treated as a plant by expressing a sound field mathematically with a wave equation. Nelson *et al.* developed feedforward control of a sound field using frequency response functions based on a modal analysis of the wave equation so as to minimize acoustic potential energy.[1–3] Ise *et al.* formulated a control method that accounted for actuator dynamics using the boundary element method for the same control objective.[4] However, both of these control methods require the reference signal of the primary source to be available and the sound field to be harmonic. Thus, the effectiveness of such control cannot be guaranteed when the primary source cannot be measured or is uncorrelated. In addition to this, these methods are not intended to alter plant dynamics.

The dynamics between a sound source and a receiving point in a sound field are significantly affected by the poles of the transfer function between them. By introducing a feedback loop into the system under study, the poles of the transfer function can be controlled, thereby controlling the modal response of the plant. Olson's and May's ''electronic sound absorber'' [5,6] is one of the most notable works that involve the application of feedback control to active sound field control. In their system, the sound pressure at the microphone collocated with a loudspeaker is fed back through an electroacoustic transducer. By choosing an appropriate feedback gain, the local sound pressure near the microphone can be reduced. Other investigations based on Olson's and May's device have been published since then, and have primarily been concerned with the classical control theory based on a transfer function analysis and were not intended to provide control over the acoustical characteristics of a sound field, i.e., to treat a full sound field as a plant. Clark *et al.*[7] proposed a feedback control system based on Olson's and May's device and direct rate feedback control for control of the modal damping ratio of acoustic modes. Using the mode theory, they described the stability of the sound field into which their control system was introduced and showed the simulation results for globally damping the acoustic response of an enclosure.

Several studies on the application of the modern control theory based on the state-space method for active sound field control have been published. Dohner *et al.*[8] and Bai *et al.*[9] developed an active noise control system using linear quadratic optimal control. In their system, the full sound field was regarded as a plant, however it was assumed that the modal parameters for describing the sound field in state space were theoretically or experimentally known. Dohner *et al.*[10] also derived a method that was not based on the mode theory for constructing a state equation for a sound field but accounted for every mode. However, their analysis was applied to a one-dimensional problem for simplicity.

In the present study, a state-space description of a sound field is derived from an inhomogeneous wave equation using

the finite element method. This allows the sound field to be treated without relying on any kind of shape that may have unknown modal parameters and also relieves us of the need to identify the modal parameters of the sound field experimentally. Controlling the poles of the transfer function for the sound field using state feedback control is investigated for improving the irregular distribution of the natural frequencies of a sound field at low frequencies.

Generally, in a small room and at low frequencies, the natural frequencies of the sound field are widely spaced. Specifically, if the room is symmetrical such as cubical or spherical, the natural frequencies begin to degenerate, resulting in an irregular acoustic response of the room, with intermittent large gaps between resonances. However, if the shape of the room is designed appropriately, the natural frequencies do not degenerate at all, and the natural frequencies are more evenly spaced. Thus, in room acoustics, the distribution of the natural frequencies as well as the damping of the sound field should be designed appropriately for the fidelity of transmission.[11]

When feedback control is employed for active noise control for an enclosed sound field, the real parts of the poles of the transfer function are modified such that the damping is increased, attenuating the acoustic response in the vicinity of the resonance frequencies. In order to distribute the natural frequencies appropriately, the imaginary parts of the poles of the transfer function should be modified. At present, there are very few papers discussing this problem based on active sound field control.

The optimum design for room shape for improving the distribution of the natural frequencies of the sound field has been discussed in several papers.[12,13] However, the desired distribution cannot always be realized by simply modifying room shape. If the distribution can be actively controlled with electroacoustic transducers, it is expected that the acoustic response can be made more regular than by modifying room shape alone. Therefore, it is noteworthy to control the distribution of the natural frequencies of a sound field by active control.

II. STATE FEEDBACK CONTROL SYSTEM

A. State-space description of a sound field

Here, the procedure for constructing a state equation for a sound field using the finite element method for describing a sound field, i.e., a control plant, in state space is presented.

The inhomogeneous wave equation related to velocity potential ϕ is expressed as follows:

$$\nabla^2\phi - \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = -q, \tag{1}$$

where c is the speed of sound in the fluid, and q is the distribution of the strength of sound sources. Applying the finite element method to the above equation with area Ω and boundary Γ, the following can be obtained:[14,15]

$$\mathbf{M}\ddot{\boldsymbol{\phi}} + \mathbf{D}\dot{\boldsymbol{\phi}} + \mathbf{K}\boldsymbol{\phi} = \mathbf{f}v, \tag{2}$$

with

$$M_{ij} = \frac{1}{c^2}\int\int\int_\Omega N_i N_j\, dV, \quad D_{ij} = -\rho\int\int_\Gamma \frac{1}{z} N_i N_j\, dS,$$

$$K_{ij} = \int\int\int_\Omega \operatorname{grad} N_i \cdot \operatorname{grad} N_j\, dV, \quad f_i = \int\int\int_\Omega N_i\, dV,$$

where $\boldsymbol{\phi}$ is a vector with components of velocity potential ϕ at each node, v is the strength of the sound source, ρ is the density of the fluid, z is the normal acoustic impedance of the boundary surface, and N_i is the interpolation function of the ith node. Equation (2) can be transformed to a state-space description as follows:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{v}, \quad y = \mathbf{c}\mathbf{x}, \tag{3}$$

with

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{M}^{-1}\mathbf{K} & -\mathbf{M}^{-1}\mathbf{D} \end{bmatrix}, \quad \mathbf{B} = [\mathbf{b}_1 \mathbf{b}_2 \cdots \mathbf{b}_M],$$

$$\mathbf{b}_i = \begin{Bmatrix} \mathbf{0} \\ \mathbf{M}^{-1}\mathbf{f}_i \end{Bmatrix} \quad (i = 1,2,...,M),$$

$$\mathbf{x} = \begin{Bmatrix} \boldsymbol{\phi} \\ \dot{\boldsymbol{\phi}} \end{Bmatrix}, \quad \mathbf{v} = \begin{Bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_M \end{Bmatrix},$$

$$\mathbf{c} = \{ \quad \mathbf{0} \quad 0\cdots 0 \quad \rho \quad 0\cdots 0 \},$$

where $\mathbf{I}$ is the identity matrix, and M is the number of sound sources; $\mathbf{c}$ of the output equation is formulated such that the output y of the system becomes the sound pressure $p = \rho\partial\phi/\partial t$ at a point in the sound field.

B. Realization of feedback control system

To control a system through state feedback, all state variables of the system must be directly observable. The state variables of the state equation for a sound field derived in the preceding section are the velocity potential and its time deferential at each node. It is not practical to observe all these quantities, thus they must be estimated from some observed variables, i.e., the outputs of the system. In this section, the method for realizing the feedback control system as a feedback filter is described, using the theory of the observer.

By using the principle of superposition, the state equation for a sound field having control sources in addition to a primary source can be expressed as follows:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}_p v_p + \mathbf{B}_s\mathbf{v}_s + \mathbf{w}_1, \quad \mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{w}_2, \tag{4}$$

where v_p is the strength of the primary source, $\mathbf{v}_s$ is a vector with components of the strength of each control source, $\mathbf{w}_1$ is

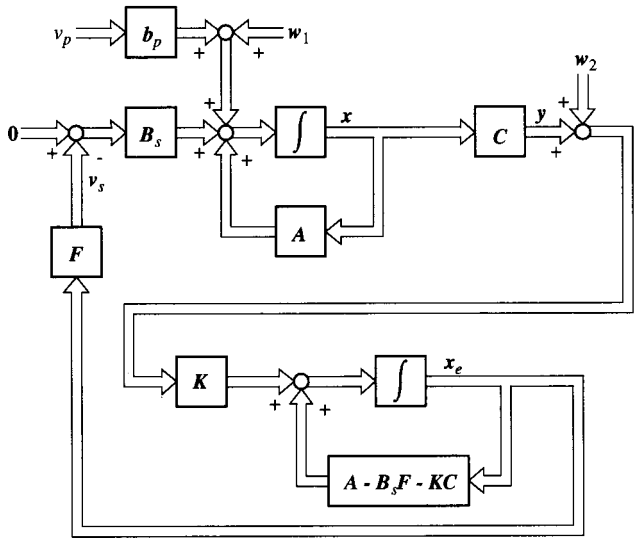


FIG. 1. Block diagram of acoustic system with feedback controller.

the process noise vector, and $\mathbf{w}_2$ is the observation noise vector. By carrying out appropriate state feedback; $\mathbf{v}_s = -\mathbf{F}\mathbf{x}$, Eq. (4) can be modified into the following state-space description for a closed loop system:

$$\dot{\mathbf{x}} = (\mathbf{A} - \mathbf{B}_s\mathbf{F})\mathbf{x} + \mathbf{b}_p v_p + \mathbf{w}_1, \quad \mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{w}_2. \tag{5}$$

When we use the Kalman filter to estimate the state variables $\mathbf{x}$ from the outputs $\mathbf{y}$ of the system, the state estimate $\mathbf{x}_e$ is given by

$$\dot{\mathbf{x}}_e = (\mathbf{A} - \mathbf{B}_s\mathbf{F} - \mathbf{K}\mathbf{C})\mathbf{x}_e + \mathbf{K}\mathbf{y}, \tag{6}$$

where

$$\mathbf{K} = \mathbf{P}\mathbf{C}^T\mathbf{R}^{-1} \tag{7}$$

and $\mathbf{P}$ is the solution of the Riccati equation;

$$\mathbf{0} = (\mathbf{A} - \mathbf{B}_s\mathbf{F})\mathbf{P} + \mathbf{P}(\mathbf{A} - \mathbf{B}_s\mathbf{F})^T + \mathbf{Q} - \mathbf{P}\mathbf{C}^T\mathbf{R}^{-1}\mathbf{C}\mathbf{P}, \tag{8}$$

where $\mathbf{Q}$ and $\mathbf{R}$ are the covariance matrices of $\mathbf{w}_1$ and $\mathbf{w}_2$, respectively. By carrying out $\mathbf{v}_s = -\mathbf{F}\mathbf{x}_e$ instead of $\mathbf{v}_s = -\mathbf{F}\mathbf{x}$, the transfer function matrix $\mathbf{G}(s)$ from $\mathbf{y}$ to the control source commands $\mathbf{v}_s$ can be expressed as follows:

$$\begin{aligned}\mathbf{V}_s(s) &= -\mathbf{F}(s\mathbf{I} - \mathbf{A} + \mathbf{B}_s\mathbf{F} + \mathbf{K}\mathbf{C})^{-1}\mathbf{K}\mathbf{Y}(s) \\ &= \mathbf{G}(s)\mathbf{Y}(s),\end{aligned} \tag{9}$$

where the Laplace transforms of $\mathbf{y}$ and $\mathbf{v}_s$ are $\mathbf{Y}(s)$ and $\mathbf{V}_s(s)$, respectively. Figure 1 shows a block diagram of the acoustic system to which the feedback controller described above is applied. The sound field control system can be arranged as shown in Fig. 2, where the use of one sensor microphone is assumed for simplicity. The state equation of the composite system for the feedback controlled sound field and the Kalman filter can be given by

$$\begin{aligned}&\begin{Bmatrix}\dot{\mathbf{x}} \\ \dot{\mathbf{x}}_e\end{Bmatrix} = \begin{bmatrix}\mathbf{A} & -\mathbf{B}_s\mathbf{F} \\ \mathbf{K}\mathbf{C} & \mathbf{A} - \mathbf{B}_s\mathbf{F} - \mathbf{K}\mathbf{C}\end{bmatrix}\begin{Bmatrix}\mathbf{x} \\ \mathbf{x}_e\end{Bmatrix} + \begin{Bmatrix}\mathbf{b}_p \\ \mathbf{0}\end{Bmatrix} v_p, \\ &y = \{\mathbf{c} \;\; \mathbf{0}\}\begin{Bmatrix}\mathbf{x} \\ \mathbf{x}_e\end{Bmatrix}.\end{aligned} \tag{10}$$

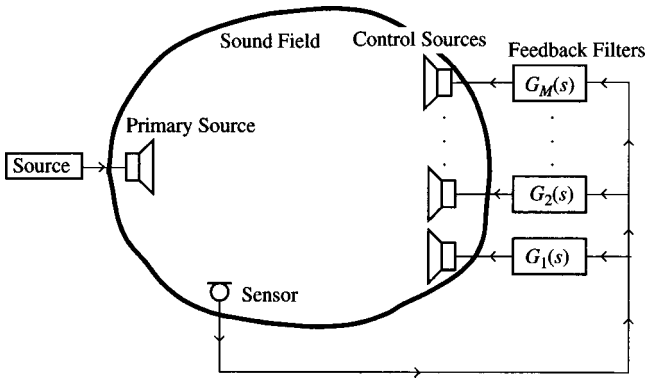


FIG. 2. Arrangement of sound field control system.

C. Control system design

If pole allocation is employed, the state feedback gain vector $\mathbf{F}$ can be found such that the eigenvalues of $\mathbf{A} - \mathbf{B}_s\mathbf{F}$ are placed arbitrarily. This means that each of the natural frequencies of a sound field can be shifted to a desired frequency, thus allowing any distribution of natural frequencies to be realized.

There are several algorithms to solve pole allocation problems. For example, by the Ackerman's algorithm, which is applicable to single control input systems, $\mathbf{F}$ can be calculated as follows:

$$\begin{aligned}\mathbf{F} = &\{0 \;\; 0 \;\; \cdots \;\; 1\} \\ &\times \mathbf{M}_c^{-1}(\mathbf{A}^N + c_{N-1}\mathbf{A}^{N-1} + c_{N-2}\mathbf{A}^{N-2} + \cdots + c_0\mathbf{I}),\end{aligned} \tag{11}$$

where $\mathbf{M}_c$ is the controllable matrix of the system and c_n $(n = 0,1,...,N-1)$ are the coefficients of the characteristic equation having the desired poles $(\lambda_1, \lambda_2, \ldots, \lambda_N)$:

$$\begin{aligned}&(s - \lambda_1)(s - \lambda_2)\cdots(s - \lambda_N) \\ &\quad = s^N + c_{N-1}s^{N-1} + \cdots + c_1 s + c_0.\end{aligned} \tag{12}$$

III. NUMERICAL STUDIES ON PROPOSED CONTROL METHOD

A. Control of a rectangular enclosure

A computer simulation was performed to demonstrate control of the natural frequencies of an enclosure. A lightly damped, two-dimensional model of a rectangular enclosure as illustrated in Fig. 3 was selected for the simulation. The

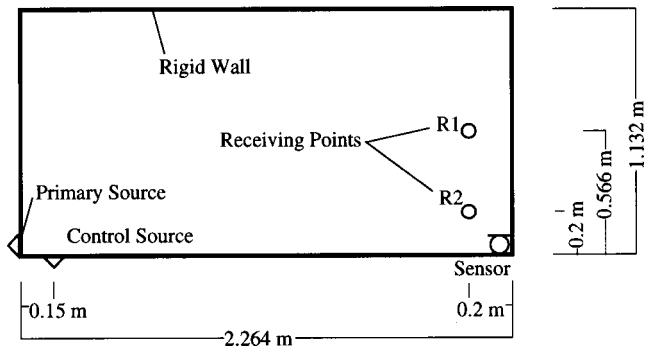


FIG. 3. Two-dimensional model of rectangular enclosure for computer simulation.

TABLE I. Desired poles set for pole allocation: rectangular room.

Mode number	Poles without control	Desired poles
1(1,0)	$-4.7890 \pm j\ 478.88$	$-10 \pm j\ 478.88$
2(0,1)	$-9.5943 \pm j\ 959.39$	$-10 \pm j\ 687.88$
3(2,0)	$-9.6131 \pm j\ 961.26$	$-10 \pm j\ 896.88$
4(1,1)	$-10.772 \pm j\ 1077.2$	$-10 \pm j\ 1105.9$
5(2,1)	$-13.725 \pm j\ 1372.5$	$-10 \pm j\ 1314.9$
6(3,0)	$-14.519 \pm j\ 1451.8$	$-10 \pm j\ 1523.9$
7(3,1)	$-17.640 \pm j\ 1763.9$	$-10 \pm j\ 1732.9$
8(0,2)	$-19.414 \pm j\ 1941.3$	$-10 \pm j\ 1941.3$

enclosure dimensions were $2.264\,\mathrm{m} \times 1.132\,\mathrm{m}$, chosen so as to raise the degeneration of its natural frequencies. The primary and control sources were modeled as piston sources having uniform velocity distributions over their surfaces. The finite elements used to discretize the sound field were first-order, two-dimensional, isoparametric triangles. There were 960 such elements, assembled in 525 degrees of freedom. For calculating the damping matrix **D**, it was assumed that the matrix could be transformed into a diagonal matrix using the undamped modal matrix and also assumed that the damping ratios were 0.01 for all modes. Then, the state equation for the sound field was constructed by the method described in Sec. II A.

The objective of control was defined as achieving a uniform distribution of undamped natural frequencies from the first mode to the eighth mode. Thus, the desired poles were set as shown in Table I to calculate the state feedback gain vector. For the higher modes not listed in Table I, the real parts of the desired poles were set to -10, and the imaginary parts were set equal to those without control. In this case, the spacing between adjacent undamped natural frequencies was 33.3 Hz in the frequency range of interest. To calculate the Kalman filter, the covariance matrix **Q** of the process noise vector $\mathbf{w}_1$ was set to $10^{-4}\mathbf{I}$ and the variance r of the observation noise w_2 was set to 1, where the covariance matrix **R** of the observation noise vector $\mathbf{w}_2$ in Eq. (8) was altered to r because only one sensor microphone was used. The impulse responses at the receiving points were calculated from the state equation of the composite system for the feedback controlled sound field and the Kalman filter expressed in Eq. (10).

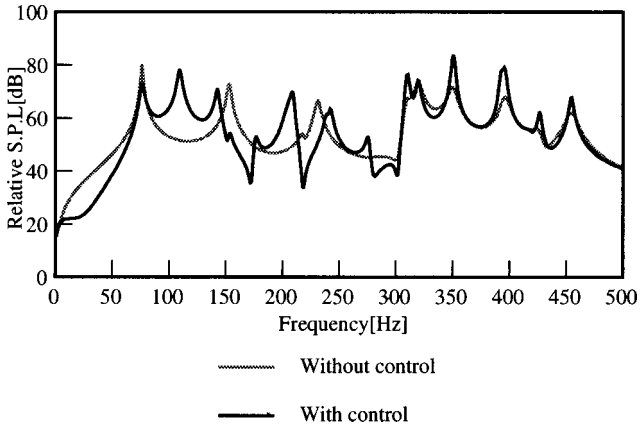


FIG. 4. Transfer functions calculated at receiving point R1 in rectangular enclosure.

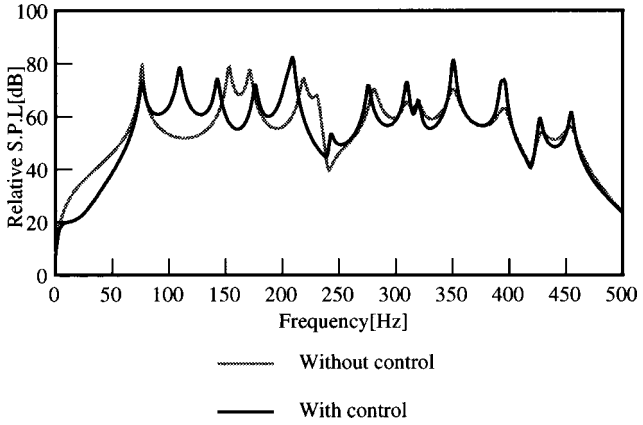


FIG. 5. Transfer functions calculated at receiving point R2 in rectangular enclosure.

Figures 4 and 5 show the transfer functions at receiving points R1 and R2, respectively. In the transfer functions, without state feedback control both at R1 and R2, peaks due to the acoustic resonance of the enclosure are observed at irregular intervals. In particular, the resonance peaks of the first (1,0) mode and second (0,1) mode are widely spaced because the second (0,1) and third (2,0) modes occur close together. Introducing the proposed control, the natural frequencies are moved such that the peaks are observed at the same intervals from the first mode (76.2 Hz) to the eighth mode (309.0 Hz). These results indicate that the acoustic response is expected to become more regular by applying the proposed state feedback control to a sound field.

The proposed active control of a sound field is based on self-tuning via a minor loop and changing the pole placement of the acoustic system. As the poles of an acoustic system are common to any transfer function in that system, effective control, similar to that at R1 and R2, can be obtained at any other receiving point in the sound field.

B. Control of a nonrectangular enclosure

The feedback control method described in this paper has the ability to control domains with arbitrary shapes. A com-

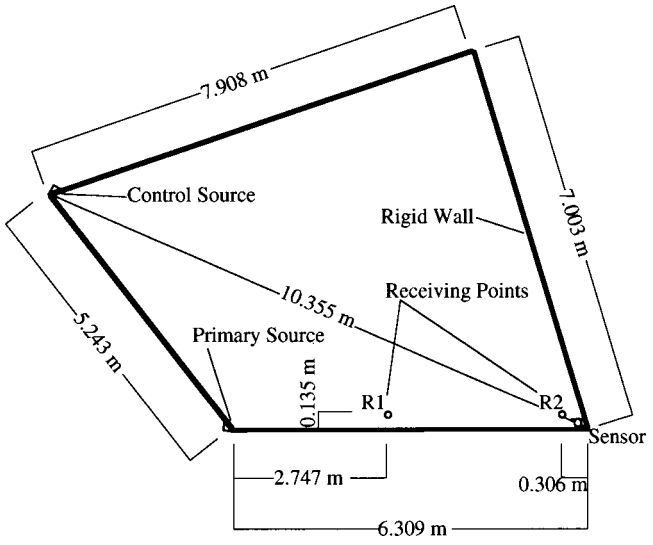


FIG. 6. Two-dimensional model of nonrectangular enclosure for computer simulation.

TABLE II. Desired poles set for pole allocation: nonrectangular room.

Mode number	Poles without control	Desired poles
1	$-1.4745 \pm j\ 147.44$	$-3 \pm j\ 147.44$
2	$-1.8412 \pm j\ 184.12$	$-3 \pm j\ 190.44$
3	$-2.4332 \pm j\ 243.31$	$-3 \pm j\ 233.44$
4	$-3.1110 \pm j\ 311.08$	$-3 \pm j\ 276.44$
5	$-3.4320 \pm j\ 343.18$	$-3 \pm j\ 319.44$
6	$-3.6326 \pm j\ 363.25$	$-3 \pm j\ 362.44$
7	$-4.3640 \pm j\ 436.38$	$-3 \pm j\ 405.44$
8	$-4.4912 \pm j\ 449.10$	$-3 \pm j\ 449.10$

puter simulation was also performed for a two-dimensional nonrectangular enclosure, as illustrated in Fig. 6. This shape is the floor plan of the three-dimensional room that was identified by Milner *et al.*[13] as the optimal room shape with respect to the regularity of the modal distribution. In this simulation, there were 1922 elements with 1024 degrees of freedom. The desired poles were set as shown in Table II. The real parts of the desired poles not listed in Table II were set to -3, and the imaginary parts were set equal to those without control. The spacing between adjacent undamped natural frequencies was 6.8 Hz. The other numerical conditions were the same as in the rectangular case.

Figures 7 and 8 show the transfer functions at receiving points R1 and R2, respectively. Without control, the large gaps between the resonance peaks that were observed in the rectangular case are not seen due to the larger volume and asymmetry of this enclosure. However, the gaps between the third and fourth modes, and between the sixth and seventh modes, are larger than the other gaps. By applying the proposed control, the natural frequencies move such that they are distributed in a uniform manner. These results indicate that the distribution of natural frequencies of an asymmetrical room can also be controlled using the proposed control method.

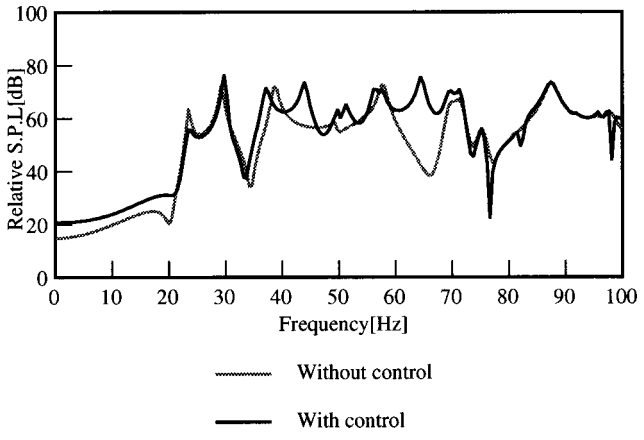


FIG. 7. Transfer functions calculated at receiving point R1 in nonrectangular enclosure.

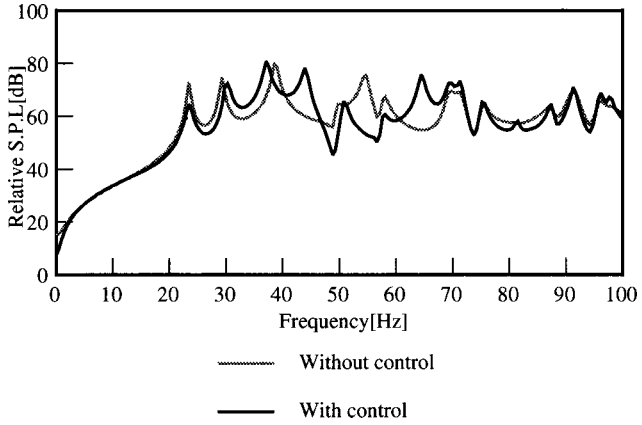


FIG. 8. Transfer functions calculated at receiving point R2 in nonrectangular enclosure.

IV. CONTROL EXPERIMENTS IN AN ENCLOSURE

A. Experimental arrangement

Figure 9 shows the configuration of the experimental system. The experimental enclosure was constructed of acrylic plates. The procedure used to construct the state equation for the sound field was similar to that used in the computer simulations, except for certain numerical conditions; in the experimental system, there were 972 elements, 532 degrees of freedom, and the damping ratios was 0.03. The objective of the control was also identical to that of the computer simulations; the desired poles were set as shown in Table III from the first mode to the eighth mode. The spacing between adjacent undamped natural frequencies was 38.4 Hz.

In this experiment, real time feedback control was not carried out but was simulated using a control command calculated in advance. The control command $\mathbf{v}_s$ was calculated as follows.

The state variables $\mathbf{z}$ of a sound field with the state feedback follow the differential equation:

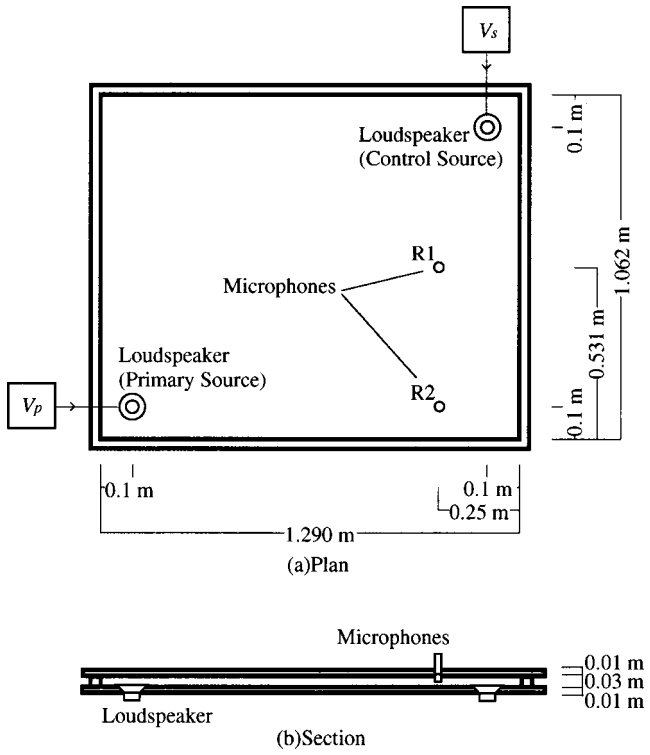


FIG. 9. Schematic of experimental enclosure and control system.

TABLE III. Desired poles set for pole allocation: experiment.

Mode number	Poles without control	Desired poles
1(1,0)	$-8.4009 \pm j\ 840.05$	$-10 \pm j\ 840.05$
2(0,1)	$-10.211 \pm j\ 1021.0$	$-10 \pm j\ 1081.1$
3(1,1)	$-13.258 \pm j\ 1325.8$	$-10 \pm j\ 1322.1$
4(2,0)	$-16.830 \pm j\ 1682.9$	$-10 \pm j\ 1563.1$
5(2,1)	$-19.780 \pm j\ 1977.9$	$-10 \pm j\ 1804.1$
6(0,2)	$-20.533 \pm j\ 2053.2$	$-10 \pm j\ 2045.1$
7(1,2)	$-22.259 \pm j\ 2225.8$	$-10 \pm j\ 2286.1$
8(3,0)	$-25.319 \pm j\ 2531.7$	$-10 \pm j\ 2527.1$

$$\mathbf{z} = (\mathbf{A} - \mathbf{B}_s\mathbf{F})\mathbf{z} + \mathbf{b}_p v_p . \tag{13}$$

The control command

$$\mathbf{v}_s = -\mathbf{F}\mathbf{z} \tag{14}$$

is equivalent to the control command obtained when all state variables are measured by sensors and full state feedback is carried out. In this experiment, a time-stretched pulse was chosen as the input to the primary source v_p and the input to the control source $\mathbf{v}_s$ was calculated by Eqs. (13) and (14). The primary source and the control source were then driven simultaneously.

B. Results

Figures 10 and 11 show the transfer functions at microphones R1 and R2, respectively. With state feedback control, peaks are observed near the undamped natural frequencies set for the design of the feedback controller, except for the third mode at R1. R1 is positioned on a nodal plane of the third mode of the sound field without control, and for the design of the feedback controller, the value of the undamped natural frequency of the third mode is set at nearly the same value. Thus, a zero of the transfer function for the nodal plane remains at the same position even with control and the peak due to the third mode cannot be observed.

V. CONCLUSIONS

In an effort to realize state feedback control of a sound field with any shape, a method was given that links the finite element method with the modern control theory based on the state-space method. The objective of control was to distribute the lowest undamped natural frequencies in a uniform manner so as to realize more regular acoustic responses. Pole allocation was employed to calculate the state feedback gain vector so that the roots of the closed loop system had the desired modal distribution.

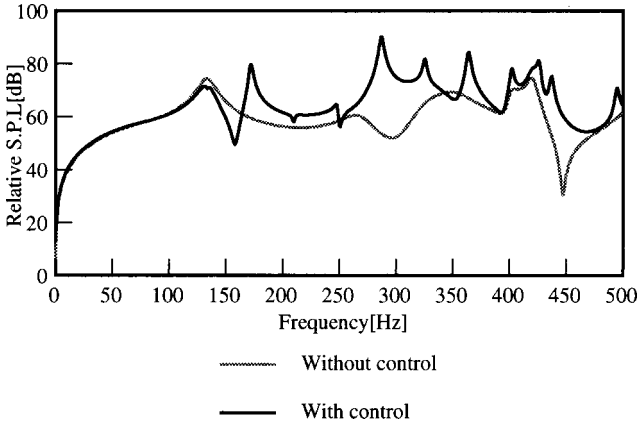


FIG. 10. Transfer functions measured at microphone R1.

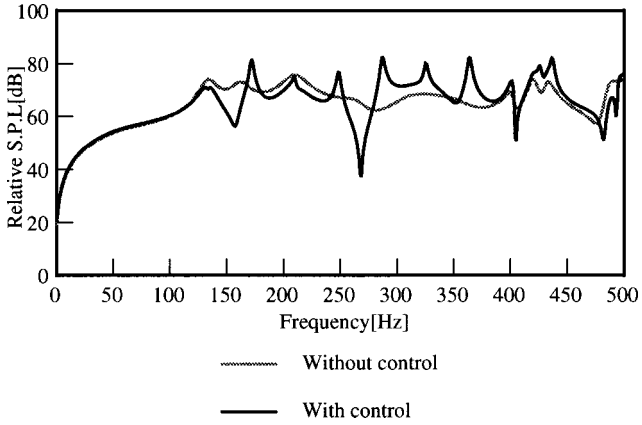


FIG. 11. Transfer functions measured at microphone R2.

Computer simulations were performed, and demonstrated that the positions of the peaks of the transfer functions changed according to state feedback control and that the peaks were shifted such that they occurred at equal intervals over the frequency range of interest, even in an irregularly-shape enclosure. A practical control system was then built for experimental verification of the results obtained in the computer simulations. The measured transfer functions under state feedback control contained peaks near the undamped natural frequencies set for the design of the feedback controller. The results of the computer simulations and experiment suggest that desired modal distributions and regularity of acoustic responses can be achieved by applying the proposed state feedback control to a sound field.

[1] P. A. Nelson, A. R. D. Curtis, S. J. Elliott, and A. J. Bullmore, "The active minimization of harmonic enclosed sound fields, part I: theory," J. Sound Vib. **117**, 1–13 (1987).

[2] A. J. Bullmore, P. A. Nelson, A. R. D. Curtis, and S. J. Elliott, "The active minimization of harmonic enclosed sound fields, part II: a computer simulation," J. Sound Vib. **117**, 15–33 (1987).

[3] S. J. Elliott, A. R. D. Curtis, A. J. Bullmore, and P. A. Nelson, "The active minimization of harmonic enclosed sound fields, part III: experimental verification," J. Sound Vib. **117**, 35–58 (1987).

[4] S. Ise, S. Sakamoto, and H. Tachibana, "Analysis of active mode control in a room by the boundary element method," J. Acoust. Soc. Jpn. **51**, 25–33 (1995) (in Japanese).

[5] H. F. Olson and E. G. May, "Electronic sound absorber," J. Acoust. Soc. Am. **25**, 1130–1136 (1953).

[6] H. F. Olson, "Electronic control of noise, vibration, and reverberation," J. Acoust. Soc. Am. **28**, 966–972 (1956).

[7] R. L. Clark and D. G. Cole, "Active damping of enclosed sound fields through direct rate feedback control," J. Acoust. Soc. Am. **97**, 1710–1716 (1995).

[8] J. L. Dohner and R. Shoureshi, "Modal control of acoustic plants," ASME J. Vib., Acoust., Stress, Reliab. Des. **111**, 326–330 (1989).

[9] M. R. Bai, C. Shieh, and D. Lee, "Development of a modal LQG active noise canceler," Proceedings of Active 95, 911–920 (1995).

[10] J. L. Dohner and R. Shoureshi, "A method for active noise control using a source-point model," J. Acoust. Soc. Am. **86**, 1053–1059 (1989).

[11] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), pp. 576–599.

[12] M. M. Louden, "Dimension-ratios of rectangular rooms with good distribution of Eigentones," Acustica **24**, 101–104 (1971).

[13] J. R. Milner and R. J. Bernhard, "An investigation of the modal charac-

teristics of nonrectangular reverberation rooms,'' J. Acoust. Soc. Am. **85**, 772–779 (1989).
[14] Y. Kagawa, T. Yamabuchi, and A. Mori, ''Finite element simulation of an axisymmetric acoustic transmission system with a sound absorbing wall,'' J. Sound Vib. **53**, 357–374 (1977).
[15] A. Craggs, ''The application of acoustic and absorption finite elements to sound fields in small enclosures,'' in *Finite Element Applications in Acoustics*, edited by M. M. Kamal and J. A. Wolf, Jr. (ASME, New York, 1981), pp. 1–19.

Evaluation of decay times in coupled spaces: Bayesian parameter estimation[a),b)]

Ning Xiang[c)]
National Center for Physical Acoustics and Department of Electrical Engineering, University of Mississippi, 1 Coliseum Drive, University, Mississippi 38677

Paul M. Goggans
Department of Electrical Engineering, University of Mississippi, 1 Coliseum Drive, University, Mississippi 38677

(Received 15 March 2001; accepted for publication 29 May 2001)

Determination of sound decay times in coupled spaces often demands considerable effort. Based on Schroeder's backward integration of room impulse responses, it is often difficult to distinguish different portions of multirate sound energy decay functions. A model-based parameter estimation method, using Bayesian probabilistic inference, proves to be a powerful tool for evaluating decay times. A decay model due to one of the authors [N. Xiang, J. Acoust. Soc. Am. **98**, 2112–2121 (1995)] is extended to multirate decay functions. Following a summary of Bayesian model-based parameter estimation, the present paper discusses estimates in terms of both synthesized and measured decay functions. No careful estimation of initial values is required, in contrast to gradient-based approaches. The resulting robust algorithmic estimation of more than one decay time, from experimentally measured decay functions, is clearly superior to the existing nonlinear regression approach. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1390334]

PACS numbers: 43.55.Br, 43.55.Ka, 43.60.–c [JDQ]

I. INTRODUCTION

Acoustics in coupled spaces have long been studied in the context of architectural acoustics. The orchestra pit, the stagehouse, and balconies in an opera house/theater coupled to the main floor are common examples of spaces for performing art. In certain conditions, the sound energy in these coupled spaces decays with two or more distinct exponential rate constants. A growing number of halls with secondary hard chambers have recently been received positively.[1] These hard chambers are partially coupled to the primary space, simultaneously achieving clarity and reverberance. The opening and closing of combinations of these secondary chambers has become an important tool for generating the range of acoustic conditions needed for the widely varying music performed today. To obtain better understanding and control of acoustics in coupled spaces it is vital to have an efficient technique of determining decay times. In the 1930s Eyring[2] investigated coupled rooms theoretically, and observed experimentally that sound energy in two coupled rooms decays at two different rates in certain conditions. A systematic discussion of sound energy decay in a pair of coupled rooms is given in Ref. 3. However, decay time measurements in these rooms are often considered difficult. In the evaluation of decay times great care is needed to distinguish different portions of simple exponentials having different decay rates. This is traditionally accomplished by visual inspection, since no algorithmic approach has been available. The present study provides a more efficient determination of decay times in coupled rooms, using Schroeder's backward integration method.[4] To accomplish this, the decay function model established recently in Ref. 5 is extended to multirate decay functions. Bayesian parameter estimation is then proposed to yield a robust algorithm for the evaluation of decay times.

The present study demonstrates that a model-based analysis using Bayesian probability theory is well suited to determining decay times from measurements in coupled spaces. Bayesian probability theory, which includes Bayes' theorem, includes all valid rules of statistics for relating and manipulating probabilities; interpreted as logic, it is a quantitative theory of inference. Bayesian probability theory has recently found increasing applications in science and engineering, and is well described as *probability theory as logic*.[6]

Decay rates in a given model can also be estimated by the least square (LS) of residuals method. Previous work[5] has used LS to estimate a single reverberation time from Schroeder's decay functions. The LS method is a gradient-based recursive approach, which requires an initial estimate of the parameter values. In the estimation of a single reverberation time, the initial parameter values are not critical.[5] However, with an increasing number of decay rates, convergence of the iterative procedure is guaranteed only when the initial parameter values lie in a small subspace around the global extreme in parameter space. This is a serious weakness of the method.

The Bayesian approach has several advantages. In parameter estimation, Bayesian probability theory generates a posterior probability density over the parameter space. If a single best value is required then the maximum of that den-

[a)]Parts of this work were presented at the 136th and the 141st meeting of the Acoustical Society of America [J. Acoust. Soc. Am. **104**, 1763(A) (1998) and **109**, 2283 (2001)].
[b)]Dedicated to Professor Manfred R. Schroeder on the occasion of his 75th birthday.
[c)]Author to whom correspondence should be addressed; electronic mail: nxiang@olemiss.edu

sity can be found. A number of efficient optimization methods exist that do not get stuck at local extreme and which approach the global extreme without need for initial estimates of parameters.

Once the model for the process has been specified in terms of the parameters to be estimated, any further relevant available information can be routinely incorporated using Bayes' theorem. Bayesian analysis ensures that the resulting posterior distribution is based on precisely the information put into it—not more and not less—and utilizes that information optimally.

A great advantage of the Bayesian formalism is its ability to handle parameters, having a bearing on the parameters of interest but which are not of interest themselves, known as nuisance parameters. Bayesian analysis allows them to be incorporated and then averaged (marginalized) over. The result is a posterior distribution for the desired parameters alone, as required, yet which takes account of the influence of the nuisance parameters. Marginalization is a rigorous consequence of the laws of probability and it reduces the dimensionality of the parameter space.

Systematic development of a Bayesian formalism for estimation of decay rates therefore involves the explicit incorporation of information into the model, followed by marginalization. This analysis has been developed before in another context,[7] but has not been applied to architectural acoustics using Schroeder's decay functions.

A further advantage is that the Bayesian formalism can estimate, from the data, the number of decay rates present in a given Schroeder decay function. The Bayesian literature refers to this as the problem of model comparison and selection. It is, however, beyond the scope of the present work, and will be addressed in a separate paper. We concentrate here on the estimation of decay rates.

Section II, begins by extending an existing model for Schroeder's decay functions, based on earlier work on the decay of sound energy in coupled spaces. Section III outlines the Bayesian approach to parameter estimation. Both synthesized and measured decay functions support performance testing of the implemented approach (Sec. IV). The detailed derivations involved in the Bayesian formalism are given in the Appendix.

II. MODELING MULTIRATE DECAYS

Eyring[2] pointed out that curves on a logarithmic scale of the decay of sound energy are not in general linear for coupled rooms having different natural reverberation times, or even for a single room with nonuniformly distributed absorption and no diffusing scheme. This section develops an analytical description of Schroeder's decay functions for these conditions.

A. Energy decay in two coupled rooms

Consider a room coupled to another room through an open interface, as illustrated in Fig. 1. The room containing the sound source will be designated as the *primary room* (PR), and the other as the *secondary room* (SR). In the simplified analysis of Ref. 3, the sound energy in each room is assumed to be uniform and diffuse. The energy decay in these rooms is then given[3] as

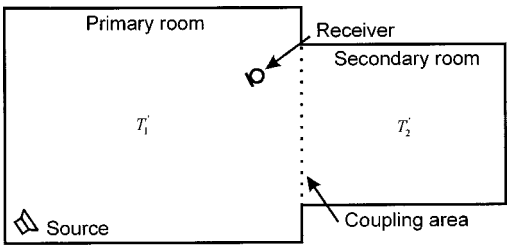


FIG. 1. Sketch of two coupled rooms. The room containing the source is designated as the primary room, and the other as the secondary room. By acoustically closing the coupling area, the sound decay in each room can be characterized by its natural reverberation time T_1' or T_2'. The present paper concentrates on the situation where the receiver is together with the source in the primary room and $T_1' < T_2'$.

$$E_I(t_k) = E_{I1} e^{-B_1 t_k} + E_{I2} e^{-B_2 t_k}, \tag{1}$$

$$E_{II}(t_k) = E_{II1} e^{-B_1 t_k} + E_{II2} e^{-B_2 t_k}, \tag{2}$$

where $E_I(t_k)$ and $E_{II}(t_k)$ are the energy decay functions in PR and SR respectively, E_{I1}, E_{I2} and E_{II1}, E_{II2} are the initial values of each decay mode, which depend on the coupling area, the source energy, and the location of the sound source and receiver. B_1 and B_2 are decay constants in PR and SR, which can be written in terms of decay times $T_i = 13.8/B_i$. In this paper time will be treated as a discrete variable t_k. In Fig. 1 the natural (decoupled) reverberation time of each room is denoted by T_1' or T_2', and the corresponding values in the coupled state, if identifiable, are denoted T_1 and T_2.

When a receiver is located in the same room as the source, Eq. (1) signifies the sound energy decay for this source-receiver arrangement. The energy decay then has two-rate character, and this is particularly clear when $T_1' < T_2'$. This situation is often of practical interest, and is emphasized below.

B. Decay function model for coupled rooms

For the experimental determination of decay rates in coupled rooms, the widely used maximal-length sequence correlation technique (see, for example, Ref. 8) provides a room impulse response (RIR) $h(t_k)$ defined between the sound source and the receiver. RIRs measured experimentally are invariably contaminated with background noise (BN). Following Ref. 5 the BN is supposed to be additive, with mean-square value $\overline{n^2}$. By applying Schroeder's backward integration[4] to the squared RIR, containing a noiseless RIR $h_f(t_k)$ and BN (see Refs. 5 and 9 for more detail), the normalized decay function d_k is given as

$$d_k \approx \frac{1}{N} \sum_{s=t_k}^{L} h_f^2(s) + \frac{\overline{n^2}}{N} (L - t_k), \tag{3}$$

where

$$N = \sum_{s=0}^{L} h^2(s) \tag{4}$$

and k and s are integers. Here L denotes the upper limit of the Schroeder integration due to the finite length of the RIR.

The first term on the right-hand side in Eq. (3), corresponding to the contribution from the noise-free RIR $h_f(t_k)$, describes the decay process of the sound energy. For coupled rooms as in Refs. 2 and 3, a multirate decay function generally results. This can be modeled as

$$F(\mathbf{A},\mathbf{B},t_k)=\sum_{j=1}^{m} A_j G_j(B_j,t_k), \quad 0\leqslant t_k \leqslant L, \tag{5}$$

where

$$G_j(B_j,t_k)=\begin{cases}\exp(-B_j\cdot t_k) & \text{for } j=1,2,...,m-1,\\ L-t_k & \text{for } j=m,\end{cases} \tag{6}$$

and $\mathbf{A}=\{A_1,A_2,\ldots,A_m\}$, $\mathbf{B}=\{B_1,B_2,\ldots,B_m\}$, with $B_i=13.8/T_i$, and $B_m=0$, $A_m=\overline{n^2}/N$. Here T_i denotes the ith decay time to be estimated, and m is the number of terms in the model. The A_j in Eqs. (5) and (6) are said to be linear parameters, and the B_j are nonlinear parameters. In the Bayesian literature, Eq. (5) is known as the general linear model.

Figure 2 shows decay curves based on Eqs. (5) and (6). A scaled time variable is used in the figure for simplicity. The BN is taken to be approximately -45 dB, and $m=3$. A single-rate Schroeder decay curve with reverberation time $T=0.5$ is also shown for comparison. In Fig. 2(a) the first decay mode, having $T_1=0.5$, begins at 0 dB. The second, with a rate $T_2=1.0$, is taken to begin at -5 dB, -10 dB, and -15 dB. As shown in Fig. 2(a), the lower the level at which the second mode becomes significant, the closer the double-rate decay curve approaches the single-rate curve. In Fig. 2(b) the second decay mode becomes significant at -10 dB with rates $T_2=1.0$, $T_2=1.25$, and $T_2=1.5$, respectively. The latter part of these decay curves, which first falls with a different decay rate between 0.3 and 0.8 than the initial part and then rapidly towards the upper limit of the integration, is a consequence of Schroeder's backward integration when RIRs contaminated with BN. It corresponds to the last term in Eq. (5), decreasing linearly with time. These simulations reveal difficulties in distinguishing double-rate decay, particularly in the deformation in the later part of the Schroeder decay curves. This deformation depends on the initial value and the decay time of the second decay mode, and on the mean value of the BN and the upper limit of integration. Visual inspection cannot always identify different decay modes in a multirate decay curve, especially when corrupted with fluctuations in real data.

The fitting of sums of decaying exponentials to experimental data is a problem that has been studied and described by Lanczos[10] as a very ill posed problem. An efficient algorithm for decay rate estimation from multirate Schroeder decay functions is needed. The next sections set up a formalism based on Bayesian model-based parameter estimation for solving this problem.

III. BAYESIAN DECAY TIME ESTIMATION

The decay model approximates Schroeder's decay function

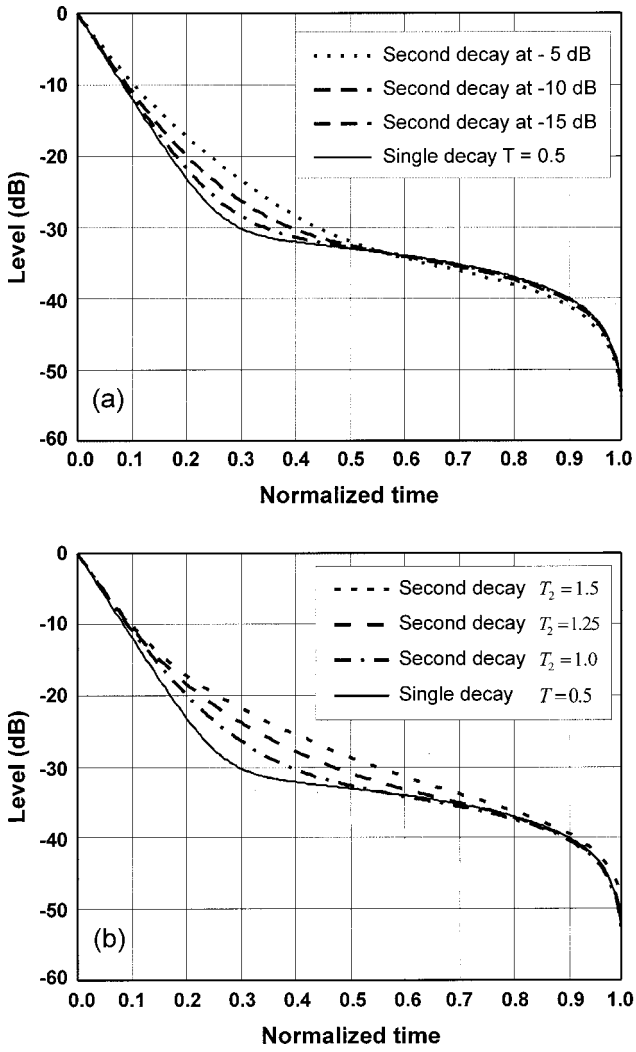


FIG. 2. Double-rate Schroeder decay curves compared with a single-rate curve. A dimensionless, scaled time axis is used. The first rate at fixed $T_1=0.5$ begins at 0 dB. (a) The second decay, with a fixed time of $T_2=1.0$, begins at -5 dB, -10 dB, and -15 dB, respectively. (b) The second decay begins at a fixed level of -10 dB with a decay time of $T_2=1.0$, $T_2=1.25$, and $T_2=1.5$, respectively.

$$d_k=F(\mathbf{A},\mathbf{B},t_k)+\epsilon_k, \quad 1\leqslant k\leqslant K, \tag{7}$$

in such a way that $F(\mathbf{A},\mathbf{B},t_k)$ in Eq. (5) or (6) models Schroeder's decay function with error ϵ_k. This error is defined as the difference between the measured data and the model and is often referred to as the residual error. It includes measurement errors, fluctuations in the decay function data [see Figs. 4(b) and 5(c)] and modeling errors. The information I that specifies the problem includes the hypothesis that the data consist of a systematic part $F(\cdot)$ and an additive error part ϵ_k, the error statistics (see Sec. II B), and the model specified in Eq. (5).

A. Posterior probability of parameters

Given Schroeder's decay function data $\mathbf{D}=\{d_1,d_2,\ldots,d_K\}$ obtained from experimental measurements, and the relevant background information I, the joint probability of all the model parameters $p(\mathbf{A},\mathbf{B}|\mathbf{D},I)$ expresses the

probability that the given model in Eq. (5) accurately describes the physical situation. According to Bayes' theorem it can be written as

$$p(\mathbf{A},\mathbf{B}|\mathbf{D},I)=\frac{p(\mathbf{A},\mathbf{B}|I)p(\mathbf{D}|\mathbf{A},\mathbf{B},I)}{p(\mathbf{D}|I)}, \tag{8}$$

where $p(\mathbf{A},\mathbf{B}|I)$ is the joint probability density of **A** and **B** given the background information I encapsulating what is known about these parameters before observing the data; it is often called the *prior* (probability). This prior specifies the degree to which the prior information I implies the model is correct; it can also be interpreted as the degree we should believe the model is correct based on the information I. The probability $p(\mathbf{D}|\mathbf{A},\mathbf{B},I)$ is called the likelihood distribution for the data, if the parameter values in the model were known. It is vital in finding the probability $p(\mathbf{A},\mathbf{B}|\mathbf{D},I)$ and is studied in depth below. The probability $p(\mathbf{A},\mathbf{B}|\mathbf{D},I)$ is called the posterior probability in Bayesian terminology, because it applies after the data have been taken into account. The probability $p(\mathbf{D}|I)$ is the probability of the data **D** given only the background information I. It is referred to as the *evidence* and is of interest when comparing models. Here it acts merely as a constant that normalizes the product of the prior and the likelihood to give the joint posterior distribution for **A** and **B**.

The prior $p(\mathbf{A},\mathbf{B}|I)$ and the likelihood $p(\mathbf{D}|\mathbf{A},\mathbf{B},I)$ must be assigned *a priori*.[11] For the present problem the prior information has so little bearing on the parameter values that $p(\mathbf{A},\mathbf{B}|I)$ varies little in the region in which the likelihood distribution places most of its weight, as a relatively sharp peak. In consequence a uniform prior can safely be assigned for $p(\mathbf{A},\mathbf{B}|I)$. This constant cancels in the normalization procedure giving the posterior as proportional to the product of the prior and likelihood. In principle, Bayesian analysis has the capability to take account of any prior information by incorporating it into the calculation.

B. Likelihood distribution

The likelihood is the probability density that a particular data set should be observed supposing that the model parameters are known. From Eq. (7) it is just the probability of the residual error.[7] Supposing that the model specified by Eq. (5) properly describes the situation, the only available information about the error ϵ_k is that it corresponds to a finite but unknown amount of power. The error samples ϵ_k (for $1\leqslant k\leqslant K$) are logically independent of each other, and the finite power implies a finite variance σ^2. We may also take the error to have zero mean amplitude, since any other value would be part of the model. In the absence of any other information regarding the error statistics, the *principle of maximum entropy* now assigns a Gaussian probability density function for ϵ_k, with (unspecified) variance σ^2:[12,13]

$$p(\mathbf{D}|\mathbf{A},\mathbf{B},\sigma,I)=(\sqrt{2\pi}\sigma)^{-K}\times\exp\left\{\frac{-1}{2\sigma^2}\sum_{k=1}^{K}[d_k-F(\mathbf{A},\mathbf{B},t_k)]^2\right\}, \tag{9}$$

where K is the number of data points. The logical independence of ϵ_k has been used to multiply the probabilities for each data point according to the product rule. The likelihood distribution is written $p(\mathbf{D}|\mathbf{A},\mathbf{B},\sigma,I)$ as a reminder that the variance σ^2 is also unknown. Note that this assignment is not the same as if the error (noise) is taken to be Gaussian white.[7] The maximum-entropy assignment of a Gaussian probability density function follows from the mean and variance of the error and the fact that no further information about ϵ_k is available.

Introduce the matrix notation

$$g_{ij}=\sum_{k=1}^{K}G_i(B_i,t_k)G_j(B_j,t_k),\quad i,j=1,...,m \tag{10}$$

and let e_{ji} represent the ith component of the jth eigenvector of **G**, λ_j be the jth eigenvalue of **G**, so that the model function in Eq. (5) can be written[7] as

$$F(\boldsymbol{\alpha},\mathbf{B},t_k)=\sum_{i=1}^{m}\alpha_iQ_i(B_i,t_k), \tag{11}$$

where $\boldsymbol{\alpha}=\{\alpha_1,\alpha_2,\ldots,\alpha_m\}$ and

$$Q_j(B_j,t_k)=\frac{1}{\sqrt{\lambda_j}}\sum_{i=1}^{m}e_{ji}G_i(B_i,t_k), \tag{12}$$

$$\alpha_j=\sqrt{\lambda_j}\sum_{i=1}^{m}A_ie_{ji}\quad\text{and}\quad A_j=\sum_{i=1}^{m}\frac{\alpha_ie_{ij}}{\sqrt{\lambda_i}}. \tag{13}$$

$Q_j(B_j,t_k)$ in Eq. (12) fulfills

$$\sum_{k=1}^{K}Q_i(B_i,t_k)Q_j(B_j,t_k)=\delta_{ij}. \tag{14}$$

Now substitute the model function in Eq. (11) into the likelihood function of Eq. (9). By taking a uniform prior for $p(\boldsymbol{\alpha},\mathbf{B}|I)$ and suppressing the normalization constant $p(\mathbf{D}|I)$, the posterior probability of the model parameters in Eq. (8) given the variance of ϵ_k is

$$p(\boldsymbol{\alpha},\mathbf{B}|\mathbf{D},\sigma,I)\propto(\sqrt{2\pi}\sigma)^{-K}\exp\left\{-\frac{-K}{2\sigma^2}\times\left[\overline{d^2}-\frac{2}{K}\sum_{j=1}^{m}\alpha_jq_j+\frac{1}{K}\sum_{j=1}^{m}\alpha_j^2\right]\right\} \tag{15}$$

with

$$\overline{d^2}=\frac{1}{K}\sum_{k=1}^{K}d_k^2, \tag{16}$$

$$q_j=\sum_{k=1}^{K}d_kQ_j(B_j,t_k),\quad 1\leqslant j\leqslant m. \tag{17}$$

C. Removal of linear parameters

Architectural acoustics is principally concerned with the decay constants **B**. The $\boldsymbol{\alpha}$, are at the moment, considered as nuisance parameters. It is assumed that no information is available about these parameters, and they can be removed by performing a probabilistic averaging process over them,

known as *marginalization*. This is a rigorous consequence of the sum and product rules of probability. The α_j's are assigned a uniform prior, following which the marginalization process corresponds to the integral

$$p(\mathbf{B}|\sigma,\mathbf{D},I)=\int_{-\infty}^{\infty} p(\boldsymbol{\alpha},\mathbf{B}|\mathbf{D},\sigma,I)\,d\boldsymbol{\alpha}. \tag{18}$$

In the present case Eq. (18) can be integrated analytically. (see the Appendix) to give the posterior probability density for the desired decay rates **B**, without the unwanted $\boldsymbol{\alpha}$, as

$$p(\mathbf{B}|\sigma,\mathbf{D},I)\propto\sigma^{-K+m}\exp\left\{-\frac{K\overline{d^2}-m\overline{q^2}}{2\sigma^2}\right\}, \tag{19}$$

where

$$\overline{q^2}=\frac{1}{m}\sum_{j=1}^{m} q_j^2. \tag{20}$$

By assigning a Jeffreys prior $1/\sigma$ (Ref. 14) to the error parameter σ, the dependence of the likelihood function on σ can be removed by a further marginalization integral (see the Appendix), giving

$$p(\mathbf{B}|\mathbf{D},I)\propto\left[1-\frac{m\overline{q^2}}{K\overline{d^2}}\right]^{(m-K)/2}. \tag{21}$$

The posterior density function in Eq. (21) is often called the Student *t*-distribution (STD). It depends only on the decay parameters **B** given the data **D** and the model. It has been shown[15] that the STD in Eq. (21) becomes singular at some point in the parameter space if and only if the model can be fitted to the data exactly. This does not imply a failure of Bayesian probability theory—it is still the correct probability assignment indicating an infinitely greater probability for the parameter values than for any others.[15]

The linear parameters **A** (through $\boldsymbol{\alpha}$) have been treated as nuisance parameters above, and removed by marginalization. If desired they can also be estimated, and this is now shown.

D. Estimation of linear parameters

The expectation values $\langle\alpha_j\rangle$ are given by the first moment

$$\langle\alpha_j\rangle=\frac{\int_{-\infty}^{+\infty}\alpha_j p(\boldsymbol{\alpha},\mathbf{B}|\mathbf{D},\sigma,I)\,d\alpha_1\cdots d\alpha_m}{\int_{-\infty}^{+\infty} p(\boldsymbol{\alpha},\mathbf{B}|\mathbf{D},\sigma,I)\,d\alpha_1\cdots d\alpha_m} \tag{22}$$

under the assumption that **B** were known. On substituting the likelihood function $p(\boldsymbol{\alpha},\mathbf{B}|\mathbf{D},\sigma,I)$ given by Eq. (15) into Eq. (22), it emerges that[15]

$$\langle\alpha_j\rangle=q_j. \tag{23}$$

According to Eq. (13) the expectation values of $\langle A_j\rangle$ can be obtained as

$$\langle A_j\rangle=\sum_{i=1}^{m}\frac{q_i e_{ij}}{\sqrt{\lambda_i}}. \tag{24}$$

The expectation $\langle A_j\rangle$ is dependent on the **B**, since all of the q_j, the eigenvectors e_{ij} and the eigenvalues λ_j are functions of the parameters **B**. For a sufficiently large quantity of data the probability $p(\mathbf{B}|\mathbf{D},I)$ in Eq. (21) is so sharp as to be effectively a delta function (see Figs. 4 and 5), so that the estimate of the parameters **B** given by this peak can simply be substituted into Eq. (24).

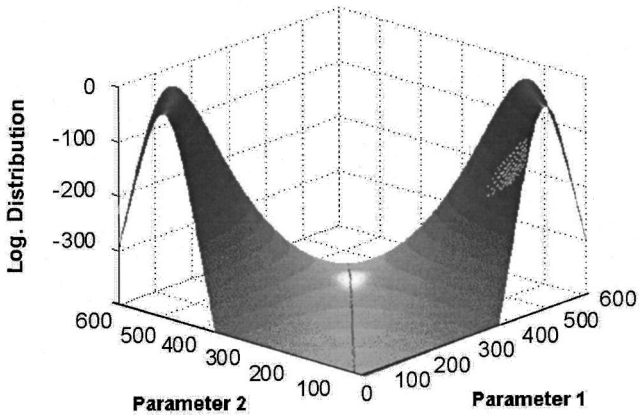


FIG. 3. Normalized logarithmic posterior probability density function of two decay times (rates). In this example, the plotted range of the two parameters is divided into 600 steps.

E. Search for nonlinear parameters

It is conventional to work with the logarithm of Student *t*-distribution. Because of its sharpness here, a given parameter set **B** effectively corresponds to a single value. Figure 3 shows the STD evaluated over a two-dimensional parameter space using a simulated Schroeder decay function. The data are simulated using Eq. (5) with $m=3$ and background noise of approximately -50 dB. As shown in Fig. 3 the STD is multimodal, since the decay constants are interchangeable such that $B_1=\beta$, $B_2=\chi$ is equivalent to $B_1=\chi$, $B_2=\beta$. In general there are $(m-1)!$ identical extreme. Any of these will serve when seeking the global maximum.

If the maximum of Eq. (21) cannot be found analytically, a search algorithm can be used. Exhaustive searching over the multidimensional parameter space is in general not computationally feasible when the dimension of the parameter space exceeds one. The example in Fig. 3 requires a great number of STD calculations over a grid of 600×600 in the parameter space. The STD is sampled efficiently over the parameter space using Markov Chain Monte Carlo (MCMC) methods.[16–18] Of these, the Gibbs sampler[17,18] is a highly flexible technique and is used in the present work. The basis of Gibbs sampling is to reduce the problem of drawing samples from a multivariate density into one of drawing successive samples from densities of lesser dimensionality.

To speed up the search, the random walk can be restricted to a reasonable value range when estimating decay times. In particular, a rough estimate of the first decay rate can be found from the early part of the Schroeder decay curve, say between -5 and -10 dB. A small range can then be specified around the resulting estimate, significantly shortening the search time.

In summary, this section has derived the Bayesian posterior probability. For the linear model of Eq. (5) it is possible to reduce the posterior probability to a compact form:

Student *t*-distribution [STD, Eq. (21)]. The present analysis can be applied not only to decay rate estimation but to any parameter estimation problem in which the data can be modeled by Eq. (5) (in its general form) and the number of data K acquired from experiments is clearly larger than the number of the model parameters m. The Student *t*-distribution has to be calculated over the parameter space. The decay times are estimated by searching a nonlinear parameter set associated with the global maximum of the STD. Gibbs sampling has been employed. This reduces the problem of drawing samples from a multivariate density to one of drawing successive samples from densities of lesser dimensionality. An initial estimate of the first decay rate can be made from the early part of the Schroeder decay function, speeding up the search.

IV. EXPERIMENTAL RESULTS

Room impulse responses (RIRs) were obtained from measurements in real halls and in scaled-down coupled rooms. Schroeder integration was applied to all RIRs after filtering by an octave bandpass filter at a central frequency of 1 kHz.

A. Decay function with one or two decay rates

Figure 4(a) shows the Student *t*-distribution for a single decay (reverberation) time based on a Schroeder decay function measured experimentally in a real hall. The signal-to-noise (S/N) ratio of the measured RIR amounts to 50 dB. In this case, a single-rate decay model ($m=2$) is used in evaluating the STD. After estimating the reverberation time as $T=1.898$ s, the expectation value of the linear parameters is calculated as in Sec. III D. Based on the estimates of both the nonlinear parameter (the reverberation time) and the linear parameters, the estimated model function in Eq. (5) or Eq. (11) is illustrated in Fig. 4(b), together with the Schroeder decay function obtained from the measured RIR.

Figure 5 similarly shows the posterior probability density (STD) for decay times based on a double-rate Schroeder decay function. The Schroeder decay function is obtained from measurements in two coupled scaled-down room models, with a scaling factor 1:8. The primary room was arranged to have a shorter natural reverberation time than the secondary room. Diffusers were placed over most of the surfaces in the two model rooms, to achieve as diffuse a sound field as possible within the frequency range of interest. The S/N ratio of the room impulse response is approximately 53 dB. In Fig. 5(a), with the second decay time constant optimized (at 0.74 s), the normalized probability density of the first decay time constant is shown by the thick dashed line. A sharp peak at $T_1=0.23$ s is seen. Conversely, with the first decay time constant optimized at 0.23 s, the normalized probability density for the second decay time constant (the thin continuous line) shows a sharp peak at $T_2=0.74$ s. Figure 5(b) shows the normalized probability density over parameters B_1 and B_2. The estimate of the linear parameters in Eq. (24) shows that the second decay mode begins at 8.8 dB lower than the first mode [see Fig. 6(b)]. Figure 5(c) shows the decay model function, along with the Schroeder decay function for comparison. It confirms the validity of the decay

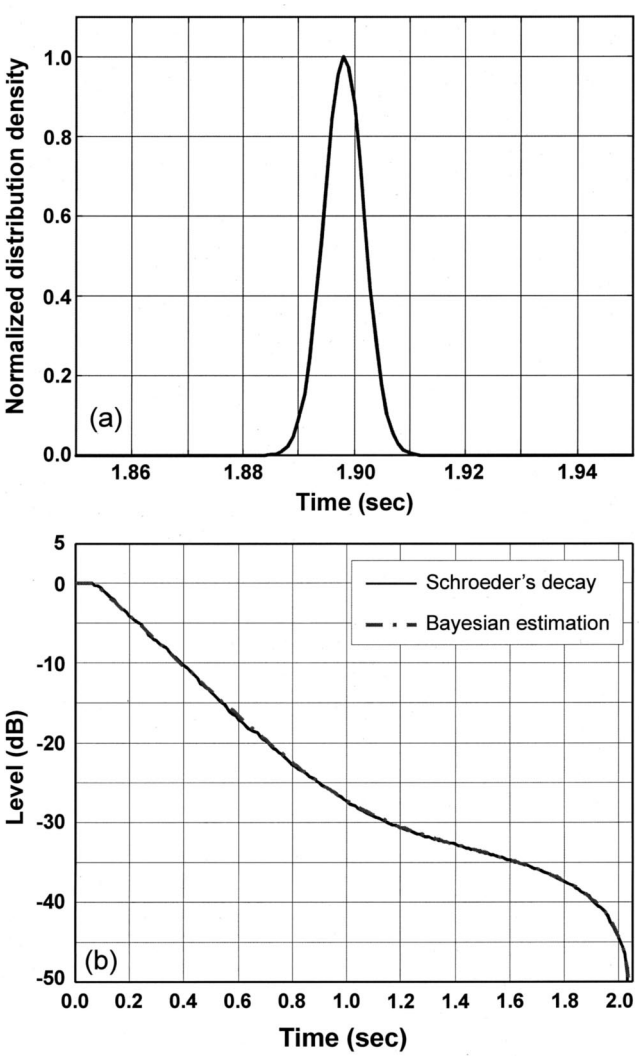
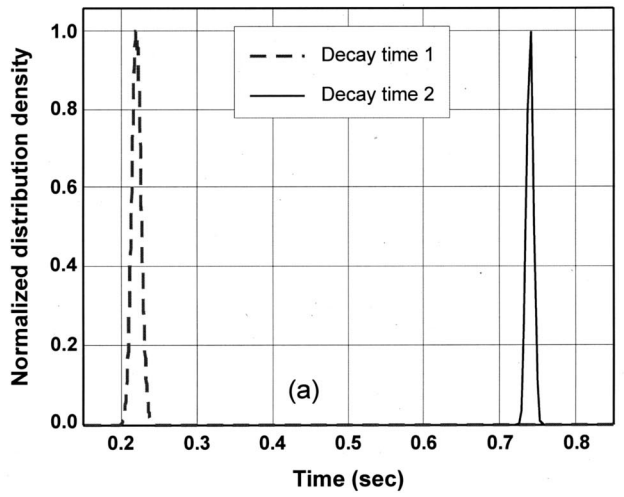


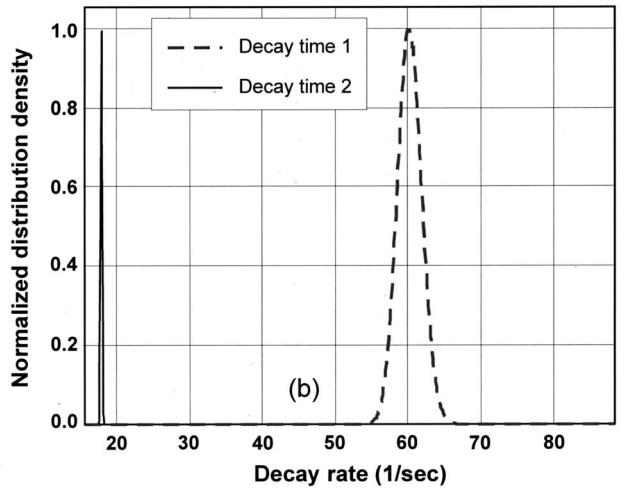
FIG. 4. (a) Posterior probability density function of a single decay time evaluated from a Schroeder decay function based on the room impulse response measured in a church. The Student *t*-distribution is evaluated over a one-dimensional parameter space. The posterior probability density peaks at a reverberation time of 1.898 s. (b) Comparison between the Schroeder decay function based on the measured room impulse response and the decay model function according to Bayesian reverberation time estimation. The signal-to-noise ratio of the measured room impulse response is 50 dB.

model function in Eq. (5). The decay model function can be reconstructed using either Eq. (5) or Eq. (11).

Given the S/N ratio 53 dB of the RIR, fluctuations on the measured Schroeder decay curve as shown in Fig. 5(c) are due predominantly to individual wall reflections/echoes in nonideal diffuse sound field. While the characteristic curvature in the middle part of the decay function [between 0.4 s and 1.0 s in Fig. 5(c)] is a consequence of the background noise in the RIR.[5] Different S/N ratios result in different levels of the curvature in the middle part as shown in Fig. 7. This curvature is not confused with the curvature due to multirate decays. Even in a single-rate decay function, the characteristic curvature is present as shown in Figs. 2 and 4(b).

The results in Figs. 3 and 5(b) show that contours of the STD are narrow and are skewed in the parameter space **B**. A simple reparametrization from B_i to T_i, where $B_i=13.8/T_i$, gives a more "orthogonalized" density[19] and is highly recommended for Gibbs sampling from the Student





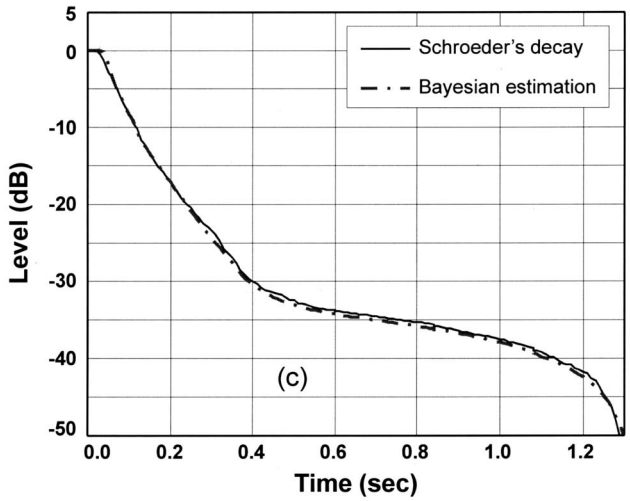
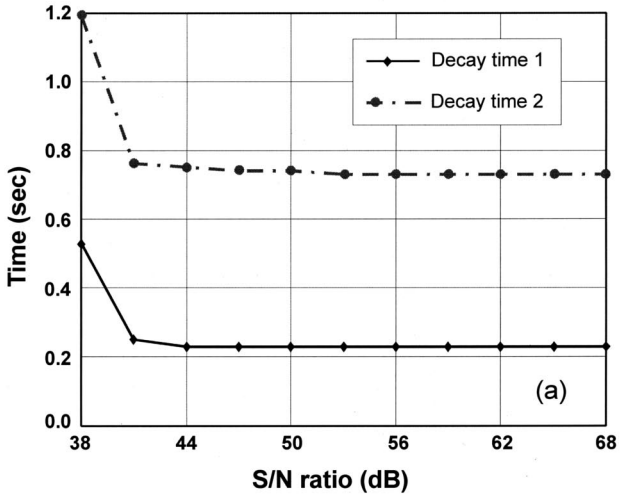


FIG. 5. (a) Normalized probability density functions of two decay times (T_1, T_2) evaluated from a Schroeder decay function measured in two scaled-down coupled model rooms. A two-rate decay model is used in decay time estimation. The probability density function of T_1 is evaluated given the optimal value of T_2 and visa versa. (b) Normalized probability density functions of two decay rates (B_1, B_2) with $B_i = 13.8/T_i$. The probability density function of B_1 is evaluated given the optimal value of B_2 and visa versa. (c) Comparison between the Schroeder decay function based on a measured room impulse response and the decay model function. The signal-to-noise ratio of the RIR amounts to 53 dB. The second decay mode begins at approximately 8.8 dB below the first $[10 \log(A_1/A_2) \approx 8.8 \text{ dB}]$.



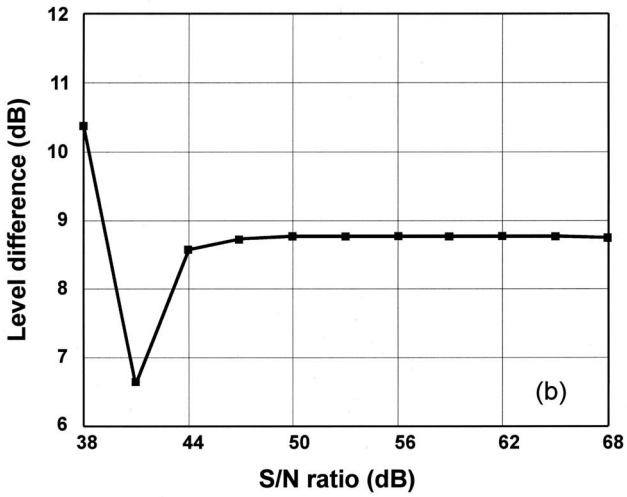


FIG. 6. Effect of the signal-to-noise ratio on the decay parameter estimation in two scaled-down coupled model rooms. The upper limit of the Schroeder integration is set at 1.5 s. (a) Dependence of two decay times on the signal-to-noise ratio. (b) Dependence on the signal-to-noise ratio of the amplitude ratio for the two decay processes $[10 \log(A_1/A_2)$ in dB].

t-distribution $p(\mathbf{T}|\mathbf{D},I)$ (with $\mathbf{T}=\{T_1, T_2, \cdots, T_{m-1}\}$) to speed up the search.

So far we have estimated the decay times when their number has been known. In practice, the number of decay rates may not be known prior to the estimation. Bayesian probability theory can also give a quantitative estimate, from the data, of how many rates are present in a Schroeder decay function. To do this the Bayesian evidence $p(\mathbf{D}|I)$ in Eq. (8) must be evaluated. The analytical solution for the Bayesian evidence in the present problem, and the experimental results, will be addressed in a subsequent work.

B. Dependence on signal-to-noise ratio

The level of background noise (BN) in measured room impulse responses (RIRs) influences the Schroeder decay function, in view of the upper limit of the integration. To study the effect of this factor, pseudorandom noise was added at different levels to the RIR used in the case studied above, as in Ref. 5. In this way the S/N ratios of RIRs can be systematically adjusted. After filtering using an octave band-

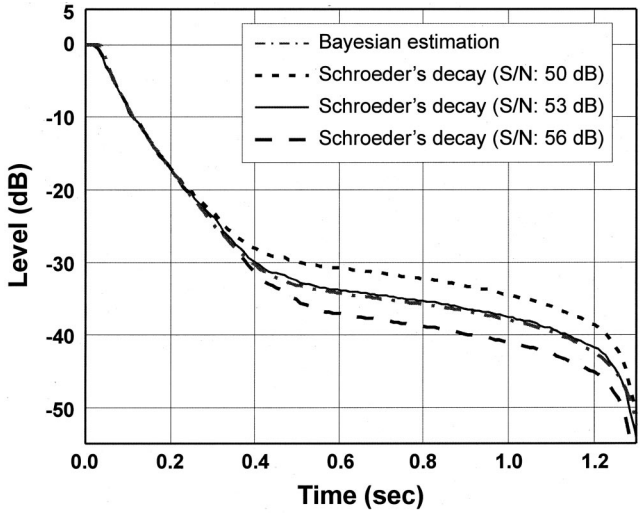


FIG. 7. Schroeder decay functions obtained from room impulse responses with a S/N ratio of 50, 53, and 56 dB. The upper limit of Schroeder integration is set at 1.5 s. Bayesian decay time estimation yields similar model decay functions for the three differing S/N ratios. A single model function is shown which concurs well with the Schroeder decay function for 53 dB. Discrepancy in the later part of the Schroeder decay function does not significantly influence the decay time estimate.

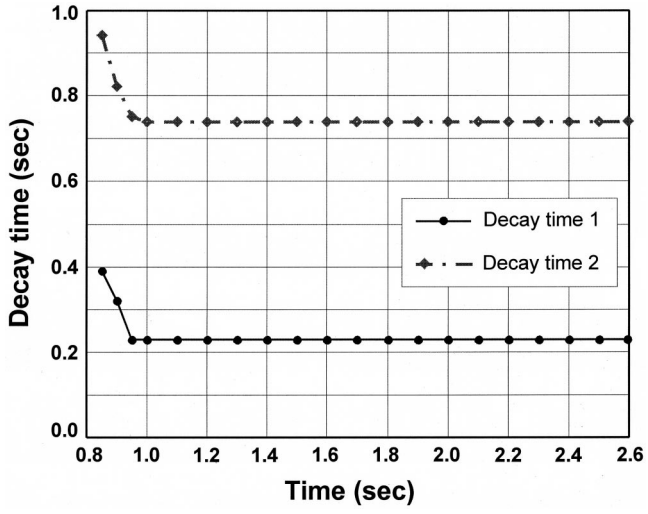


FIG. 8. Dependence of two decay times on the upper limit of Schroeder integration for a room impulse response measured in a scaled-down model of two coupled rooms. The signal-to-noise ratio of the room impulse response is 53 dB.

pass filter with a central frequency of 1 kHz, the S/N ratio of the RIRs is found to fall from 65 dB down to 38 dB in 3 dB steps.

Figure 6 depicts the dependence of Bayesian decay rate estimates on the S/N ratio. For a given arrangement of sound source and receiver, two decay rates and the ratio of the first two linear parameters are estimated under different S/N ratios of RIRs with an upper limit of integration $L=1.5$ s. Figure 6 indicates that the S/N ratio of the RIR under test is not critical to decay time estimation when it is greater than 41 dB. This safety limit of the S/N ratio is liable to vary according with the experimental and theoretical details, especially when the second decay process begins at different levels and there are differing decay rates.

In Sec. III D the linear parameters were deduced from the decay rates. The reliable estimation of two decay times [Fig. 6(a)] and of the quantity $10\log(A_1/A_2)$ [Fig. 6(b)] for all S/N ratios above 41 dB also implies reliable estimation of the parameter A_3. Table I lists the values of these estimates.

TABLE I. Dependence of decay time estimates on signal-to-noise ratio. A double-rate decay model is used. In this example, Bayesian decay time estimation yields a reliable value when the signal-to-noise ratio is higher than 41 dB.

S/N ratio (dB)	T_1 (s)	T_2 (s)	A_1/A_2 (dB)	A_3
38	0.53	1.2	10.4	5.94E−7
41	0.25	0.77	6.62	5.71E−7
44	0.23	0.76	8.57	5.31E−7
47	0.23	0.75	8.72	5.08E−7
50	0.23	0.74	8.76	4.96E−7
53	0.23	0.74	8.76	4.90E−7
56	0.23	0.74	8.77	4.86E−7
59	0.23	0.74	8.77	4.84E−7
62	0.23	0.74	8.76	4.84E−7
65	0.23	0.74	8.76	4.83E−7

The decay model function taken with the values of these estimates should also be reliably independent of the S/N ratio. When comparing these Schroeder decay functions at differing S/N ratios with the corresponding decay model functions, a discrepancy is observed in the later part of the decay curves. This is shown in Fig. 7 for S/N ratios of 50, 53, and 56 dB. In Fig. 7 only the Schroeder decay curve with S/N ratio of 53 dB agrees well with the decay model curve. All model curves associated with different S/N ratios effectively fall together. A closer look at the decay model of Eq. (5) reveals that the number of linear parameters is greater than the number of nonlinear parameters. The final linear parameter A_m is not accompanied by any nonlinear parameter. Fortunately this last linear parameter A_m is a nuisance parameter in architectural acoustics. Estimation of the decay time(s) is the primary aim, not comparison between the real Schroeder decay curve and the decay model curve. Also, the absolute values of the first two linear parameters are of less significance than their ratio. Bayesian estimation using a large number of experimentally measured Schroeder decay functions indicates that most of discrepancies in the later part of decay functions can be reduced by adding (or subtracting) additional noise at an appropriate level to the measured RIRs.

C. Dependence on the upper limit of integration

The effect of the upper limit of integration L on decay time estimates has also been investigated, at a given S/N ratio of 53 dB. Figure 8 illustrates the dependence of estimates on L. Figure 8 clearly shows that the estimated decay time changes little over a large range of L, provided that L is sufficiently large. In this example, any upper limit greater than 1 s will result in a reliable estimate.

In summary, this section has verified that the decay function models derived in Sec. III are realistic models for estimating two decay rates from experimental data. Sections IV B and IV C confirm that both the upper limit of integra-

tion and the S/N ratio are not critical to the Bayesian decay time estimate provided that the former is sufficiently large and the latter is not less than a critical value.

V. SUMMARY AND CONCLUSIONS

Determination of decay times in coupled spaces has long been a challenge to acousticians. This paper has applied Schroeder's backward integration method to the evaluation of decay times in two coupled spaces, via the room impulse responses. Under certain conditions, the decay functions in two coupled spaces are multirate in character. However, real Schroeder decay function data are invariably influenced by background noise of the room impulse responses, and deformation of the later parts of Schroeder decay functions also makes it difficult to identify different decay modes of decay functions by inspection. A decay model derived from a recent work[5] has been extended for modeling Schroeder's decay functions in coupled spaces. Model-based parameter estimation within the Bayesian probabilistic framework is then well suited to evaluate decay times in these coupled spaces. Bayesian parameter estimation has been outlined, and the posterior probability density is in the form of Student *t*-distribution. In the resulting algorithm, numerical sampling is done by Gibbs sampling. By employing both simulated and experimentally measured Schroeder decay functions, the performance of the Bayesian algorithm for estimating decay times can be tested. Multiple decay times can be reliably estimated with no need for careful choice of initial values, in contrast to gradient-based algorithms. Comparison of decay time estimates from real and model decay functions confirm the validity of the extended decay model. The results are found not to depend significantly on the signal-to-noise ratio of the room impulse responses or the upper limit of the integration.

This paper has shown how to estimate a given number of decay modes from the Bayesian viewpoint. Bayesian model comparison and selection is also of practical importance to find the most likely number of decay modes from data in coupled spaces, and this will be treated in future work.

ACKNOWLEDGMENTS

The authors thank Professor D.Y. Maa and Professor M.R. Schroeder for their encouragement and constructive criticism, and Dr. M. Vorlaender, Dr. R. Martin, Dr. W. T. Chu, and Dr. J. M. Sabatier for valuable suggestions. Thanks are also due to Donghua Li, who contributed to the collection of data from scaled-down models as part of his Master's degree thesis.

APPENDIX

Removal of linear parameters

To remove the linear parameters one begins by substituting Eq. (15) into Eq. (18), to give

$$p(\mathbf{B}|\sigma,\mathbf{D},I)=(\sqrt{2\pi}\sigma)^{-K}\int_{-\infty}^{\infty}\exp\left(-\frac{R}{2\sigma^2}\right)d\alpha_1\cdots d\alpha_m, \tag{A1}$$

where

$$R=K\overline{d^2}-2\sum_{j=1}^{m}\alpha_j q_j+\sum_{j=1}^{m}\alpha_j^2$$

$$=K\overline{d^2}-m\overline{q^2}+\sum_{j=1}^{m}(\alpha_j-q_j)^2 \tag{A2}$$

and $\overline{q^2}$ is given by Eq. (20). Equation (A1) now becomes

$$p(\mathbf{B}|\sigma,\mathbf{D},I)=(\sqrt{2\pi}\sigma)^{-K}\exp\left[\frac{K\overline{d^2}-m\overline{q^2}}{2\sigma^2}\right]\cdot W, \tag{A3}$$

where

$$W=\prod_{j=1}^{m}\int_{-\infty}^{\infty}\exp\left[-\frac{(\alpha_j-q_j)^2}{2\sigma^2}\right]d\alpha_j, \tag{A4}$$

with the change of variable $\alpha_j-q_j=u_j$ so that $d\alpha_j=du_j$, Eq. (A4) reads

$$W=\prod_{j=1}^{m}\left[2\int_0^{\infty}\exp\left(-\frac{u_j^2}{2\sigma^2}\right)du_j\right]=(\sqrt{2\pi}\sigma)^m. \tag{A5}$$

Substitution of Eq. (A5) into Eq. (A3) gives Eq. (19).

Removal of the nuisance parameter σ

By integrating Eq. (18) with the Jeffreys[14] prior $1/\sigma$ over σ it reads

$$p(\mathbf{B}|,\mathbf{D},I)\propto\int_0^{\infty}\sigma^{-K+m-1}\exp\left\{-\frac{K\overline{d^2}-m\overline{q^2}}{2\sigma^2}\right\}d\sigma. \tag{A6}$$

The integral (A6) is of the form

$$\int_0^{\infty}u^{2n}\exp\{-au^2\}du=\frac{1\cdot3\cdot5\cdots(2n-1)}{2^{n+1}a^n}\sqrt{\frac{\pi}{a}}$$

with the change of variable $u=1/\sigma$; $d\sigma=-\sigma^2\,du$ and replacing the integral limits with $\sigma\to0$; $u\to\infty$ and $\sigma\to\infty$; $u\to0$,

$$n=\frac{K-m-1}{2}\quad\text{and}\quad a=\frac{K\overline{d^2}}{2}\left(1-\frac{m\overline{q^2}}{K\overline{d^2}}\right).$$

If $p(\mathbf{B}|\mathbf{D},I)$ is regarded as a function of the parameter $\mathbf{B}$ given the particular data set $\mathbf{D}$, $(K\overline{d^2}/2)^{(m-K)/2}$ can be treated as a constant, so that

$$p(\mathbf{B}|,\mathbf{D},I)\propto\frac{1\cdot3\cdot(K-m)}{2^{(K-m+3)/2}}\sqrt{\pi}\cdot\left[\frac{K\overline{d^2}}{2}\left(1-\frac{m\overline{q^2}}{K\overline{d^2}}\right)\right]^{(m-K)/2}. \tag{A7}$$

By disregarding irrelevant constants, this corresponds to the Student *t*-distribution given in Eq. (21).

[1] R. Johnson, E. Hahle, and R. Essert, "Variable coupled cubage for music performance," Proc. MCHA'95, 1995, Kirishima, Japan.

[2] F. Eyring, "Reverberation time measurements in coupled rooms," J. Acoust. Soc. Am. **3**, 181–206 (1931).

[3] L. Cremer, H. A. Müller, and T. J. Schultz, *Principles and Applications of Room Acoustics* (Applied Science Publishers, London, UK, 1982).

[4] M. R. Schroeder, "New method of measuring reverberation time," J. Acoust. Soc. Am. **37**, 409–412 (1965).

[5] N. Xiang, ''Evaluation of reverberation times using a nonlinear regression approach,'' J. Acoust. Soc. Am. **98**, 2112–2121 (1995).
[6] E. T. Jaynes, ''Probability theory as logic,'' in *Maximum Entropy and Bayesian Methods*, edited by P. F. Fougere (Kluwer Academic, Dordrecht, The Netherlands, 1990), pp. 1–16.
[7] L. Bretthorst, ''Bayesian Analysis. I. Parameter Estimation Using Quadrature NMR Models,'' J. Magn. Reson. **88**, 533–551 (1990).
[8] N. Xiang, ''Using M-sequences for determining the impulse responses of LTI-systems,'' Signal Process. **28**, 139–152 (1992).
[9] W. T. Chu, ''Comparison of reverberation measurements using Schroeder's impulse method and decay-curve averaging method,'' J. Acoust. Soc. Am. **63**, 1444–1450 (1978).
[10] C. Lanczos, *Applied Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1956).
[11] T. J. Loredo, ''From Laplace to supernova SN 1987A: Bayesian inference in astrophysics,'' in *Maximum Entropy and Bayesian Methods*, edited by P. F. Fougere (Kluwer Academic, Dordrecht, The Netherlands, 1990), pp. 81–142.
[12] E. T. Jaynes, ''Highly informative priors,'' in *Bayesian Statistics 2*, edited by J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith (Elsevier Science, Amsterdam, The Netherlands, 1985), p. 329.
[13] P. M. Woodward, *Probability and Information Theory* (Pergamon, Oxford, UK, 1964).
[14] H. Jeffreys, *Theory of Probability* (Oxford University Press, London, UK, 1939) (3rd revised ed. 1961).
[15] L. Bretthorst, *Bayesian Spectrum Analysis and Parameter Estimation* (Springer-Verlag, New York, 1988).
[16] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, ''Equation of state calculations by fast computing machines,'' J. Chem. Phys. **21**, 1087–1092 (1953).
[17] S. Geman and D. Geman, ''Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,'' IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-6**, 721–741 (1984).
[18] A. E. Gelfand and A. F. M. Smith, ''Sampling based approaches to calculating marginal densities,'' J. Am. Stat. Assoc. **85**, 398–409 (1990).
[19] J. J. K. O Ruanaidh and W. J. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing* (Springer-Verlag, New York, 1996).

Linear prediction coding analysis and self-organizing feature map as tools to classify stress calls of domestic pigs (*Sus scrofa*)

Peter-Christian Schön,[a] Birger Puppe, and Gerhard Manteuffel
Forschungsinstitut für die Biologie landwirtschaftlicher Nutztiere, Forschungsbereich Verhaltensphysiologie, Wilhelm-Stahl-Allee 2, D-18196 Dummerstorf, Germany

(Received 15 August 2000; revised 23 April 2001; accepted 1 June 2001)

It is assumed that calls may give information about the inner (emotional) state of an animal. Hence, in the last years sound analysis has become an increasingly important tool for the interpretation of the behavior, the health condition, and the well-being of animals. A procedure was developed that allows the characterization, classification, and visualization of the cluster structures of stress calls of domestic pigs (*Sus scrofa*). Based on the acoustic model of the sound production the extraction of features from calls was performed with linear prediction coding (LPC). A vector-based self-organizing neuronal network was trained with the determined LPC coefficients, resulting in a feature map. The cluster structure of the calls was then visualized with a unified matrix and the neurons were labeled for their input origin. The basic applicability of the procedure was tested by using two examples which were of special interest for a possible evaluation of the normal farming practice. The procedure worked well both in discriminating individual piglets by their scream characteristics and in classifying pig stress calls vs other calls and noise occurring under normal farming conditions. [DOI: 10.1121/1.1388003]

PACS numbers: 43.60.Lq, 43.80.Ev, 43.60.Qv [WA]

I. INTRODUCTION

Vocalization may provide a useful tool for evaluating the emotional state of animals under captive and natural conditions (Jürgens, 1979; Crowell Comuzzi, 1993; Mulligan *et al.*, 1994; Weary and Fraser, 1995a, b; Schrader and Todt, 1998). The important advantage of this approach is a relatively objective, noninvasive, and real-time monitoring of emotions related to environmental changes.

Husbandry may cause stress in pigs by various reasons. The stressors activate the hypothalamo–pitutary–adrenal (HPA) and the sympathico–adrenomedullary (SAM) axes via the brain's sensory and limbic pathways. These pathways also reach central motor centers which eventually trigger behavioral stress responses. One such response is vocalization being performed by sets of muscles which are located around the pulmonary–pharyngeal tract.

In pigs, the outcome is a rather sustained cry with high-frequency bands that may be highly dynamical. It is well-known from farming practice that some handling procedures, especially the restraint of animals, can induce a number of vocalizations that may reflect discomfort or distress. There is evidence that peripheral endocrine stress responses are accompanied by changing rates of specific types of vocalizations (Schrader and Rohn, 1997; Schrader and Todt, 1998). Weary *et al.* (1998) have shown that an increased rate of high-frequency calls (>1 kHz) in young piglets is a useful indicator of the pain due to castration. The analysis and classification of pigs' screams may deliver the species' and the individual's phonetic characteristics that can be attributed to a particular stressor. If information on this interdependence is given it will be possible to judge the individual stress perception of an animal and, thus, its state of welfare or suffering. Given suitable analytic and diagnostic tools it should be possible to recognize stress quantitatively, immediately, and noninvasively. However, such tools are not available at present. In all mammals, the problem of which features are best suited for the analysis and subsequent classification has not yet been solved.

Techniques that produce good models of an arbitrary vocalization are still missing to date because it is not easy to decide which features are relevant for the exact characterization of the call. Too few features lead to an inadequate model, too many features easily overburden the computer performance for a statistical evaluation of the data (Hammerschmidt and Todt, 1995; Schrader and Hammerschmidt, 1997; Schön *et al.*, 1999). Hence, different and adapted procedures have to be applied. If such procedures are well developed they are distinguished by a good borderline of particular sounds towards other sounds, i.e., a clear, nonambiguous classification is performed.

Because of the apparent lack of effective methods, this paper presents a procedure to discriminate, classify, and visualize vocalizations of domestic pigs. The advantage of the described approach is its ability to include the dynamic structures of the calls as well as nonlinear effects like frequency steps or bifurcations. Further, decisions on the suitability for classification of the chosen features from the calls are possible. The whole procedure results in a system that will allow an assessment (e.g., classification of stressed vs not stressed). Finally, the performance of the system is demonstrated using two examples in pigs. Whereas in the first example individual piglets were discriminated by their scream character-

[a]Author to whom correspondence should be addressed; electronic mail: schoen@fbn-dummerstorf.de

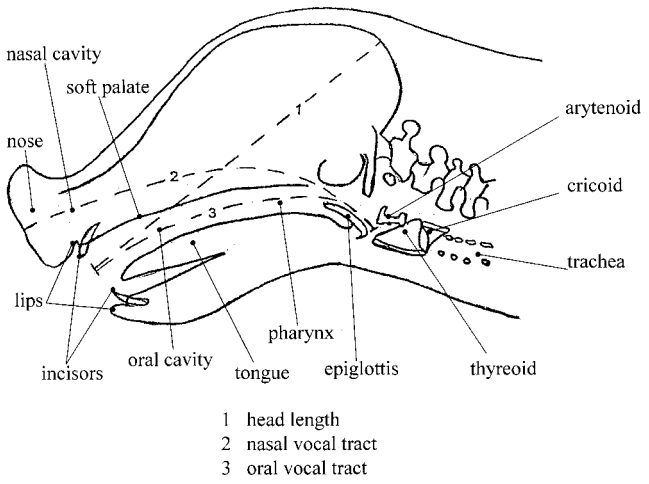


FIG. 1. Schematic drawing of the vocal tract of a piglet (redrawn from an x-ray picture).

istics, the second example shows the ability of the system to classify unknown calls as stress vocalizations when compared with calls from nonstressed animals or noises which occurred under normal farming conditions.

II. DESCRIPTION OF THE PROCEDURE

A. The acoustic model of sound production

According to Fant's (1970) acoustic model of sound production, vocalizations are produced by an apparatus consisting of a power supply (the lung, the thorax, and the diaphragm), the glottis as the sound source, and the nozzle formed by the vocal-tract cavern which serves as an acoustic resonator. The articulators (nose, tongue, and the soft palate; see Fig. 1) vary the size and the diameter of the nozzle. After the source-filter model, the signal that originates from the glottis is modulated by properties of the vocal tract.

The continuous air stream from the power supply is chopped by the glottis, resulting in pressure impulses (glottal source) with a certain fundamental frequency f_0 and its harmonics, which can be seen in the glottis source spectrum. The glottal source signal is then modified by the vocal tract, which is the portion of the system lying above the larynx. It includes the pharynx, the oral, and the nasal cavity. The vocal-tract shape can be varied by the specific placements of the tongue, lips, and jaw. Hence, the vocal tract operates as an all-pole linear filter that introduces resonance frequencies, the so-called formants in human speech. This procedure can be formulated mathematically as

$$X(z)=H(z)*I(z), \tag{1}$$

where $X(z)$ is the Z transformation of the generated sound signal, $I(z)$ the source signal of the glottis, and $H(z)$ the digital filter of the vocal tract. Then, the Z transformation of the all-pole linear filter is

$$H(z)=\frac{1}{1-\sum_{k=1}^{q} c_k z^{-k}}, \tag{2}$$

and the resulting transfer function

$$X(z)=\frac{1}{1-\sum_{k=1}^{q} c_k z^{-k}}*I(z), \tag{3}$$

can be regarded as a mathematical model of the vocal tract (Rabiner and Gold, 1975; O'Shaugnessy, 1987).

B. Linear predictive coding (LPC) as a method for feature formation and data reduction

The LPC is formed in the temporal domain from a time-sequenced data series derived from a continuous signal. It is used, for example, to extract features in the frequency domain (e.g., the frequency, amplitude, and bandwidth of resonance frequencies). In the LPC, changes in the signal are used instead of the signal itself. Thus, a sound sample $x(n)$ out of a series is taken together with a previous sample $x(n-1)$. A linear prediction of the actual sample is formed as a weighted sum of the past sample. The difference (prediction error) $e(n)$ between the two samples can be minimized by introducing a_1 as a coefficient, such that

$$e(n)=x(n)-a_1x(n-1). \tag{4}$$

The minimization of the error $e(n)$ can best be achieved if the number of previous samples is increased, which introduces also a number of additional coefficients $a_1,\ldots,a_p$.

$$e(n)=x(n)-a_1x(n-1)-a_2x(n-2)-\cdots-a_px(n-p). \tag{5}$$

Applying the Z transformation, this results in

$$E(z)=X(z)-\sum_{k=1}^{p} a_k z^{-k}*X(z), \tag{6}$$

or

$$X(z)=\frac{1}{1-\sum_{k=1}^{p} a_k z^{-k}}*E(z). \tag{7}$$

This is in formal equivalence to (3), taken $E(z)$ as the source signal from the glottis. Thus, LPC is equivalent to the source-filter model of Fant (1970) with the predictor coefficients a_k of LPC representing the vocal-tract filter coefficients c_k. For the calculation of the p coefficients $a_1,\ldots,a_p$ we used the autocorrelation method and the Levinson–Durbin recursion.

C. Self-organizing feature maps (SOFM)

Self-organizing feature maps (Kohonen, 1982, 1997) consist of a multiplicity of homogeneous processing units (neurons) with spatial relationships to their neighbors. In the basic setup input vectors are directly fed forward to the two-dimensional output layer (map). Using the neighborhood relationships between the neurons, the map organizes the input vectors according to their structure so that similar vectors are stored in neighboring areas of the map and different vectors in distant areas. According to the Kohonen algorithm we trained the feature map with the determined LPC coefficients (12-dimensional input vectors). The mapping of the input space can then be visualized with various methods. In the

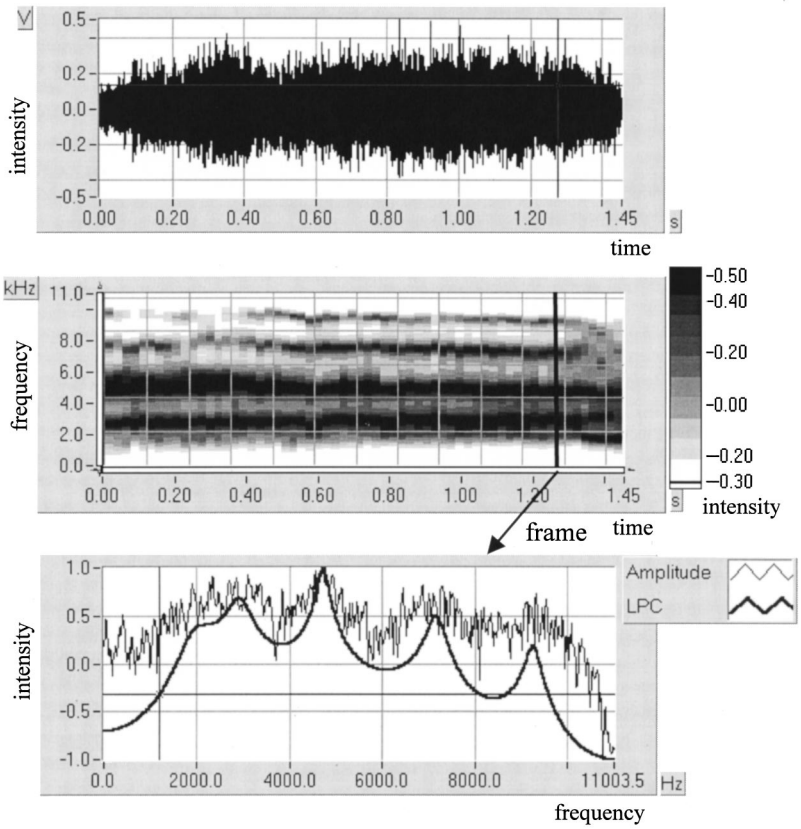


FIG. 2. Example of a stress call displayed by a piglet on the 5th day after birth. Top: Voltage signal in the time domain. Middle: Normalized logarithmic LP spectrum of the signal depicted at the top. The gray-scale density represents the intensity of the corresponding frequencies. Two resonance frequency bands at about 3 and 5 kHz are clearly visible. A narrower frequency band lies at about 8 kHz. A further, incomplete band is seen between 8 and 10 kHz. Bottom: Normalized logarithmic amplitude spectrum extracted from the FFT transformation and the LP spectrum of a single frame (46.44 ms). The local maxima of the LP spectrum coincide with the resonance spectra of the amplitude spectrum.

second step the mapping of the Kohonen network was tested with test data sets which were not included in the training data set.

A methodical problem before the self-organization can start is the choice of suitable network parameters (neurons *X*, neurons *Y*, learning radius, learning radius factor, learning rate, learning rate factor, learning steps). Still, a suitable map structure can be reached only in trying. In our case, we started with a small number of neurons (60×60) that was increased until no further improvement of the classification result and the classification parameters (maximum and mean distance—learning curve) occurred. This methodical problem could also be solved by the application of self-optimizing growing cell structures. A description of such structures is given by Fritzke (1992, 1998).

D. Extraction and graphical representation of the LPC coefficients

The number of the LPC coefficients $a_1,\ldots,a_p$ determines the number of the considered resonance frequencies of the vocal tract, with $p/2$ being the number of the resonance frequencies. Using up to 18 LPC coefficients we never found more than 6 resonance frequencies in the piglets' screams. Hence, for further analyses, 12 LPC coefficients were used. In order to take into account the dynamic structure of the calls, the LPC coefficients $a_1,\ldots,a_p$ were extracted from time windows (frames) of 46.44 ms length in each case. The LP spectrum was determined from the LPC coefficients by means of a polynomial development.

Figure 2 shows the results of the spectral analyses of a stress call as amplitude and LP spectra. The dynamic structure of the call is clearly visible by the bands of the LP spectrum in the frequency domain. Some bands alter their frequency and amplitude, and the low-frequency band bifurcates at the end of the recording. In the single-amplitude spectrum (bottom), two effects of the sound production mechanism are visible. The influence of the stimulation frequency (glottal source) results in a fundamental frequency f_0 at about 250 Hz and its harmonics. The harmonics are the quickly (200-Hz range) changing parts. The influence of the resonance qualities of the vocal tract causes slow changes. The LP spectrum represents a model of the vocal tract and may substitute the amplitude spectrum with respect to the resonance frequencies. Hence, it is possible to determine the resonance frequencies of the vocal tract from the local maxima of the LP spectrum.

III. EXAMPLES FOR THE APPLICABILITY OF THE PROCEDURE

A. Example I: Classification of individual piglets by their scream characteristics

1. Animals and recordings

The calls to be individually classified were recorded from three randomly selected piglets (*Sus scrofa*) of the ''German Landrace''. They were housed together with their mother and kept in a standard farrowing crate. ''Stress screams'' were recorded on the fifth day after birth within an interval of 2 min per animal. In this time we got between 26 and 62 screams per animal for classification. For the acoustical recordings the piglets were removed rapidly from their home pen and taken to the adjacent acoustic chamber with a minimum of disturbance. The chamber with low acoustic

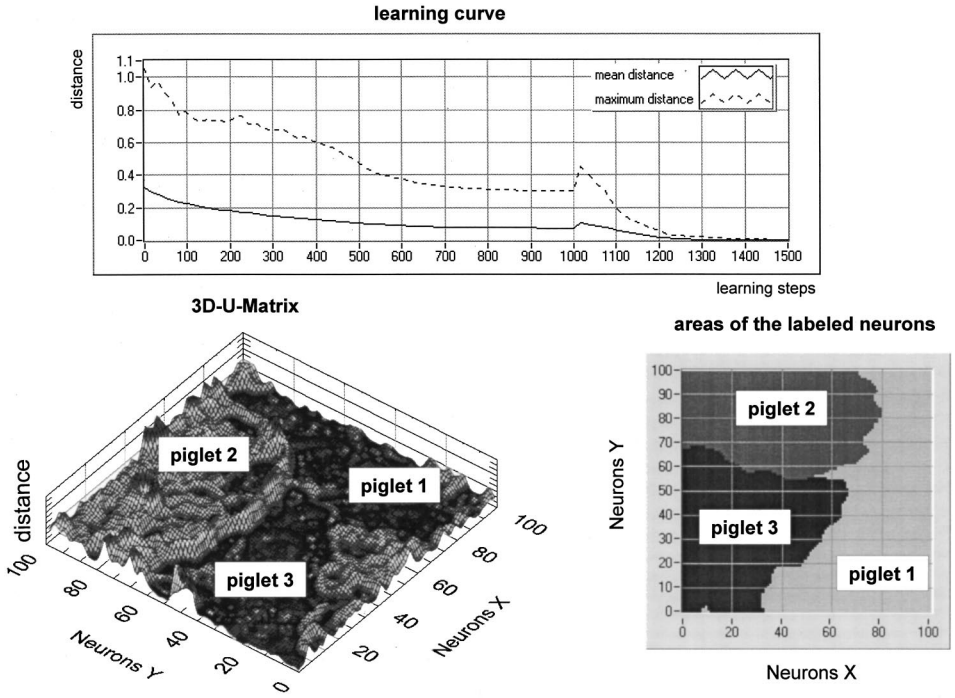


FIG. 3. Learning curve, 3D-U Matrix, and areas of the labeled neurons for the model on the 5th day after birth. The learning curve shows the quality of the training process of the SOFM with the help of the gradual descent of the mean and maximum distance of the winner neurons to the respective input vectors; the U matrix visualizes the separation of clusters by the SOFM; the areas of labeled neurons decides further about the assignment of the test vectors to the individual clusters.

reflection by the walls and the ceiling was described earlier by Schön *et al.* (1998). Screams were experimentally induced by keeping up the piglets at the thorax by a person for 1 to 2 min—a stressful situation which is quite similar to the handling immediately prior to the castration of male piglets in the normal farm animal practice, or, more common, to a situation where the sow lays down and accidentally squeezes a piglet to death. In the latter situation the screaming triggers the sow to stand up immediately (Hutson *et al.*, 1993; Weary *et al.*, 1996). All calls were recorded with a DAT recorder (Sony DCT-790) and a separate microphone (Sennheiser MKE 46). The procedures for data processing were developed and programed using the graphical programming language LABVIEW® with the additional tool DATAENGINE V.I.© (LABVIEW®, 1998; DATAENGINE V.I.©, 1999).

2. Training and test phase of the system

Fifteen screams of each individual were used to train the map and further ten screams of each piglet were used to test the map. The test screams were not included in the training data set.

a. Training phase. The SOFM was trained with 12-dimensional input vectors derived from the 12 LPC coefficients per time window (46.44 ms). Hence, 9 to 45 LPC vectors per scream were obtained. The maps used in this experiment consisted of 100×100 neurons and were trained in 3×500 learning steps. The initialization of the first 1000 steps was performed with a learning radius of 60, a learning radius factor of 0.995, a learning rate of 0.999, and a learning rate factor of 0.99. The continuous training during the following 500 learning steps was performed with a learning radius of 4; all other parameters were left unchanged. Figure 3 shows the achieved learning curves that represent the gradual descent of the mean and maximum distance of the winner neurons to the respective input vectors.

In order to assess the result of the training process, a graphic representation has proven to be helpful. This can be obtained by a U-matrix (unified matrix) representation (Ultsch and Simon, 1990), where the distance between neighboring weight vectors of the map is determined and applied to the matrix. The result can be plotted three-dimensionally where the height of the hills is proportional to the distance of the neurons. In this representation information exists over the entire input space. For each neuron, the Euclidian distance to its neighbors is determined. The obtained result is depicted in Fig. 3.

The representation with the U matrix is not dependent on the dimension of the input space. Based on the topological representation of the map, the input vectors that belong to the same cluster are found in the same areas (valleys) on the map. The valleys are separated by the hills. The scatter of the single-resonance frequencies of the vocal tract over the call and the differences between the single calls of an animal are reflected by the neighborhood relationships of the neurons. The clusters 1–3 were assigned to the three examined piglets. While piglets 1 and 3 displayed a very homogeneous structure in the screams, indicating a low variability of the calls, piglet 2 delivered a strong variability within and between the screams. Nevertheless, piglet 2 was clearly separated from the other two piglets, while the separation between piglet 1 and 3 was not as strong. The U matrix makes it possible to recognize the separation of clusters by the SOFM without having exact knowledge of the input data set. This is an important aspect in sound analysis. Thus, it is possible to decide whether or not a classification result can be reached with the chosen feature vectors from the calls.

After training of the SOFM the resulting clusters were labeled using the knowledge of the individual origins of the training data set. The result is shown in Fig. 3, where the neuronal areas were labeled with respect to the three piglets.

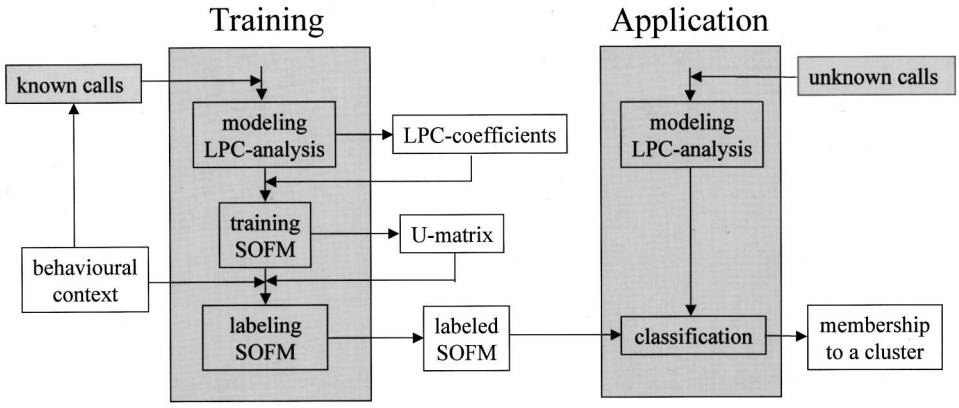


FIG. 4. Schematic representation of the process of the training and the application (testing) phase by using the SOFM.

With the labeling the training phase was finished.

b. Testing phase. The trained classification model was now ready for testing with unknown test vectors from screams that were not used for training (Fig. 4). The results of the assignment of the test LPC vectors are given in Table I. The misclassification rate was very low (<3%).

B. Example II: Classification of piglet screams versus other calls and noise

1. Animals, calls, and recordings

The procedure is not only able to deliver an individual assignment of calls, but also a classification regarding different call types. Hence, different vocal responses of pigs regarding their environment and treatment were exemplarily recorded to be analyzed later.

First, stress screams of 19 piglets from 4 different litters (litter 1–litter 3: each with three 2-week-old piglets [training], litter 4: ten 2-week-old piglets [testing]) were recorded in the same way as described in example I. We further used the screams of 16 growing pigs (5-week-old pigs) from two different litters (litter 5: nine 5-week-old piglets [training], litter 6: seven 5-week-old piglets [testing]).

Second, grunts vocalized in various social situations were used (calls displayed in a ''nonstress'' context). Both the grunts of six 2-week-old piglets (three for training and three for testing) and the grunts of six 5-week-old growing pigs (three for training and three for testing) were induced by a short social isolation of the animals with vocal contact to a companion. Further, the nursing grunts (see Schön *et al.*, 1999) of ten first-lactating sows (five for training and five for testing) were involved.

Third, three examples of the background noise occurring in the housing environment of the pigs were used. They consist of arbitrary sounds (e.g., the talking of humans, air ventilation, and rattle of the equipment) without any animal calls (noise I for training, noise II for testing) or with a small rate of animal vocalizations included in the background noise (noise III for testing). All recordings were made randomly under normal keeping conditions in the stable. The recording equipment was the same as described in example I.

TABLE I. Example I: Results for testing the trained SOFM with unknown stress screams of three piglets on their 5th day after birth.

Piglet	No. of analyzed LPC vectors	Misclassification rate (%)
1	380	0.00
2	281	1.78
3	272	2.94

2. Training and test phase of the system

a. Training phase. The map was trained with LPC vectors that were determined from screams of piglets and growing pigs. We used 7673 LPC vectors to characterize the stress (scream) area of the Kohonen map (screams from litter 1–3 and litter 5). On the other side we used LPC vectors from grunts of three piglets (724 LPC vectors), three growing pigs (2716 LPC vectors), and five first-lactating sows (178 LPC vectors), and noise I (1263 LPC vectors). Hence, the non-stress area of the Kohonen map was characterized by 4881 LPC vectors and the map was trained with a total of 12 554 LPC vectors. The maps used in this experiment consisted of 150×150 neurons and were trained in 3×500 learning steps. The initialization of the first 1000 steps was performed with a learning radius of 80, a learning radius factor of 0.995, a learning rate of 0.999, and a learning rate factor of 0.99. The continuous training during the following 500 learning steps was performed with a learning radius of 4; all other parameters were left unchanged.

b. Testing phase. The labeled SOFM was then tested with unknown LPC vectors from screams (litter 4, litter 6), grunts (three 2-week-old piglets, three 5-week-old piglets, and five first-lactating sows), and noise (noise II, noise III). The findings are shown in Table II.

The results show that the classification of screams to the ''stress area'' was possible with a misclassification rate lower than 1%. The grunts were to >97.5 correctly attributed to the ''non-stress area.'' A similar classification result was achieved with noise II. In addition to noise II, noise III consisted also of a small rate of animal communication. This was a mixture of grunts, squeals, and screams (e.g. fighting for rank order, waiting and fighting in front of the trough before feeding, attempts of escaping after isolation from the social group) which might be regarded as typical examples of vocalizations in a housing environment. When a ''non-stress'' context is assumed (Table II) the classification shows an error of 7%.

TABLE II. Example II: Results for testing the trained SOFM with unknown calls or noise classified as ''stress'' or ''nonstress'' calls.

Animals (age)	*n*	Calls/noise	No. of analyzed LPC vectors	Type	Misclassification rate means±s.d. (%)
Piglets (2 weeks)	10	screams	1904	stress	0.58± 1.11
Growing pigs (5 weeks)	7	screams	2476	stress	0.85± 1.27
Piglets (2 weeks)	3	grunts	171	nonstress	2.34± 2.55
Growing pigs (5 weeks)	3	grunts	245	nonstress	2.04± 1.70
Sows (1st lactation)	5	nursing grunts	60	nonstress	1.67± 2.89
Noise II	···	without animal calls	1706	nonstress	1.23± 1.15
Noise III	···	with animal calls	1072	nonstress	7.00± 2.79

IV. DISCUSSION

The aim of the present paper is to introduce a procedure for the classification and identification of stress screams of domestic pigs that allows the classification of animal calls using the LPC analysis and a subsequent SOFM of the Kohonen type (Kohonen, 1992).

Previous attempts of computer-aided analysis of animal vocalization were mainly based on conventional statistical methods (e.g., a discriminant analysis) with a multiparametric approach (Schrader and Hammerschmidt, 1997; Schön *et al.*, 1999). Where features like FFT coefficients or cepstrum analysis were used in combination with neural networks, the biological meaning of the respective sounds were not in the center of interest (cf. Datum *et al.*, 1996), or the dynamic structure and/or the large amount of parameter coefficients of the sounds to be recognized demanded large and complicated systems (cf. Kohonen, 1992; Schuchardt, 1992; Reby *et al.*, 1996, 1997).

Calls of mammals often have a complex structure. In past years it was shown that the sound production is not only a deterministic process but rather the result of a highly nonlinear dynamic system (Herzel *et al.*, 1995). The model of Fant (1970) represents only an idealized view on the underlying processes. In nonhuman mammal communication the appearance of nonlinear phenomena seems to be normal and it seems to occur more often in disordered animals (Riede *et al.*, 1997, 2000; Wilden *et al.*, 1998). Thus, every kind of modeling has to take these aspects into account.

Linear predictive coding (LPC), based on a source-filter model, is compatible with the mechanism of sound generation in the vocal tract and, most importantly in our application, reduces the number of coefficients necessary for a successful classification of stress calls to 12. This is a considerable reduction compared to 512 Fourier coefficients, and still more than the 50 cepstral coefficients used in an earlier study of pig vocalization (Schön *et al.*, 1999).

The coefficients of the LPC analysis were topologically represented on the feature map. Subsequently, the U matrix delivered the number of classes on the map. The result was identical with the number of animals from which the labeled data set originated. After the training phase we tested the structure of the network with unknown test screams. The classification results clearly showed that an individual assignment of the three exemplary selected piglets was possible on the basis of their particular scream characteristics.

A further task was testing a trained network regarding the classification performance with respect to the classification of piglet screams vs other calls and noise. Here, after the training was completed, unknown stress calls from pigs had to be assigned to the trained clusters. Most of the unknown calls were correctly assigned to the area of stress screams.

In order to obtain reasonable training times and an online capability of the trained system, as well as a good classification, it is necessary to estimate the optimum number of LPC coefficients that are inputs to the SOFM. This, like the size of the Kohonen network, is still not solved in general. Clearly, this optimum number depends on the sounds to be discriminated and, hence, has to be evaluated with each new task. However, this method also bears some advantages. The flexible arrangement of the feature extraction as well as the arrangement of the topology of the neuronal network allows extensive simulations of the training process and makes it possible to solve a multiplicity of classification tasks and have visual access to them.

The functionality of the described approach clearly depends on the fact that pig vocalizations are more or less sustained. That means that only weak frequency modulations occur during a call. Hence, the temporal order of the LPC coefficients was irrelevant to the SOFM. However, the comparatively simple design of the LPC-SOFM procedure for stress-call classification and identification makes it possible to create an on-line monitoring system which can detect stress calls immediately, since the necessary number of LPC coefficients and the classification by the trained neural network can be calculated in real time.

Hopefully, further development of the method extending it to other types of vocalizations will enable us to create systems that are able to recognize, discriminate, and output the semantic meaning of various animal calls. Similar to the proposal of Ritter and Kohonen (1989) for the semantics of words, this could be reached by grouping the calls according to their behavioral context.

Thus, the present procedure may be used as a methodological approach to solve different analysis and classification tasks in animal vocalization. Besides the clarification of communicative aspects, it would allow us to automatically monitor behavioral responses of farm animals in a housing situation with respect to their well-being (e.g., stressed vs nonstressed) or suffering (e.g., sick vs healthy). This will be

of increasing significance in all fields of captive and restricted animals in human charge.

ACKNOWLEDGMENTS

The authors thank the staff of the Department of Behavioral Physiology for expert technical assistance. Special thanks go to Tobias Riede for fruitful discussions and helpful comments during the work and for kindly allowing us to use one of his drawings from x rays (Fig. 1).

Crowell Comuzzie, D. K. (**1993**). ''Baboon vocalizations as measures of psychological well-being,'' Lab. Prim. Newsletter **32**, 5–6.

DATAENGINE V.I.© (**1999**). *User Manual, Function Reference, Tutorials, Basics* (MIT-Management Intelligenter Technologien GmbH, Aachen, Germany).

Datum, M. S., Palmieri, F., and Moiseff, A. (**1996**). ''An artificial neural network for sound localization using biaural cues,'' J. Acoust. Soc. Am. **100**, 372–383.

Fant, G. (**1970**). *Acoustic Theory of Speech Production* (2nd printing, Mouton, The Hague).

Fritzke, B. (**1992**). ''Wachsende Zellstrukturen—ein selbstorganisierendes neuronales Netzwerk,'' Ph.D. thesis (unpublished), Universität Erlangen-Nürnberg.

Fritzke, B. (**1998**). *Vektorbasierte Neuronale Netze* (Shaker GmbH).

Hammerschmidt, K., and Todt, D. (**1995**). ''Individual differences in vocalisations of young Barbary macaques (*Macaca sylvanus*): A multiparametric analysis to identify critical cues in acoustic signaling,'' Behaviour **132**, 381–399.

Hutson, G. D., Price, E. O., and Dickenson, L. G. (**1993**). ''The effect of playback volume and duration on the response of sows to piglet distress calls,'' Appl. Anim. Behav. Sci. **37**, 31–37.

Herzel, H., Berry, B., Titze, I., and Steinecke, I. (**1995**). ''Nonlinear dynamics of the voice: Signal analysis and biomechanical modeling,'' Chaos **5**, 1–5.

Jürgens, U. (**1979**). ''Vocalization as an emotional indicator. A neuroethological study in the squirrel monkey,'' Behaviour **69**, 88–117.

Kohonen, T. (**1982**). ''Self-organized formation of topologically correct feature maps,'' Biol. Cybern. **43**, 59–69.

Kohonen, T. (**1992**). ''How to make a machine transcribe speech,'' in *Application of Neural Networks*, edited by H. G. Schuster (Chemie, Weinheim), pp. 25–34.

Kohonen, T. (**1997**). *Self-organizing Maps*. Springer Series in Information Sciences, Vol. 30, 2nd ed. (Springer, Berlin, Heidelberg, New York).

LABVIEW® (**1998**). *Complete Software Documentation* (National Instruments Corporation, Austin, TX).

Mulligan, B. E., Baker, S. C., and Murphy, M. R. (**1994**). ''Vocalizations as indicators of emotional stress and psychological well being in animals,'' Anim. Welfare Inform. Center Newsletter **5**, 3–4.

O'Shaughnessy, D. (**1987**). *Speech Communication* (Addison-Wesley, Reading, MA).

Rabiner, L. R., and Gold, B. (**1975**). *Theory and Application of Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ).

Reby, D., Joachim, J., Lauga, J., Cargnelluti, B., and Gonzalez, G. (**1996**). ''Using voice recognition as a tool in population biology and management,'' 1st International Conference on Methods and Techniques in Behavioural Research, Utrecht, 16–18 October 1996.

Reby, D., Lek, S., Dimopoulos, I., Joachim, J., Lauga, J., and Aulagnier, S. (**1997**). ''Artificial neural networks as classification method in the behavioural sciences,'' Behav. Processes **40**, 35–43.

Riede, T., Tembrock, G., Herzel, H., and Brunnberg, L. (**1997**). ''Vocalization as indicator for disorders in mammals,''137th meeting of the Acoustical Society of America, San Diego, 1–5 December 1997.

Riede, T. (**2000**). ''Vocal changes in animals during disorders,'' Ph.D. thesis (unpublished), Humboldt University, Berlin.

Ritter, H., and Kohonen, T. (**1989**). ''Self-organizing semantic maps,'' Biol. Cybern. **61**, 241–254.

Schrader, L., and Hammerschmidt, K. (**1997**). ''Computer-aided analysis of acoustic parameters in animal vocalization: A multi-parametric approach,'' Bioacoustics **7**, 247–265.

Schrader, L., and Rohn, C. (**1997**). ''Lautäußerungen von Hausschweinen als Indikator für Stressreaktionen,'' Landbauforsch. Völkenrode **47**, 89–95.

Schrader, L., and Todt, D. (**1998**). ''Vocal quality is correlated with levels of stress hormones in domestic pigs,'' Ethology **104**, 859–876.

Schön, P. C., Puppe, B., and Manteuffel, G. (**1998**). ''A sound analysis system based on LABVIEW® applied to the analysis of suckling grunts of domestic pigs (*Sus scrofa*),'' Bioacoustics **9**, 119–133.

Schön, P. C., Puppe, B., Gromyko, T., and Manteuffel, G. (**1999**). ''Common features and individual differences in nurse grunting of domestic pigs (*Sus scrofa*): A multi-parametric analysis,'' Behaviour **136**, 49–66.

Schuchardt, J., Gruel, J. C., Lüthje, N., Molgedy, L., Radons, G., and Schuster, H. G. (**1992**). ''Neural networks for the classification of sound patterns,'' in *Application of Neural Networks*, edited by H. G. Schuster (Chemie, Weinheim), pp. 239–249.

Ultsch, A., and Simon, H. P. (**1990**). ''Kohonen's self-organizing feature map for exploratory data analysis,'' in *Proceedings of International Neural Networks* (Kluwer Academic, Paris), pp. 305–308.

Weary, D. M., and Fraser, D. (**1995a**). ''Calling by domestic piglets: Reliable signals of need?'' Anim. Behav. **50**, 1047–1055.

Weary, D. M., and Fraser, D. (**1995b**). ''Signaling need: Costly signals and animals welfare assessment,'' Appl. Anim. Behav. Sci. **44**, 159–169.

Weary, D. M., Pajor, E. A., Thompson, B. K., and Fraser, D. (**1996**). ''Risky behavior by piglets: A trade off between feeding and risk of mortality by maternal crushing?'' Anim. Behav. **51**, 619–624.

Weary, D. M., Braithwaite, L. A., and Fraser, D. (**1998**). ''Vocal response to pain in piglets,'' Appl. Anim. Behav. Sci **56**, 161–172.

Wilden, I., Herzel, H., Peters, G., and Tembrock, G. (**1998**). ''Subharmonics, biophonation, and deterministic chaos in mammal vocalization,'' Bioacoustics **9**, 171–196.

Middle-ear function with tympanic-membrane perforations. I. Measurements and mechanisms

Susan E. Voss[a)]
Picker Engineering Program, Smith College, 51 College Lane, Northampton, Massachusetts 01063; Eaton-Peabody Laboratory of Auditory Physiology and Department of Otolaryngology, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114; Speech and Hearing Sciences Program, Harvard–M.I.T. Division of Health Sciences and Technology, Cambridge, Massachusetts 02139; Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts; and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

John J. Rosowski
Eaton-Peabody Laboratory of Auditory Physiology and Department of Otolaryngology, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114; Speech and Hearing Sciences Program, Harvard–M.I.T. Division of Health Sciences and Technology, Cambridge, Massachusetts 02139; Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts; and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Saumil N. Merchant
Eaton-Peabody Laboratory of Auditory Physiology and Department of Otolaryngology, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114; Speech and Hearing Sciences Program, Harvard–M.I.T. Division of Health Sciences and Technology, Cambridge, Massachusetts 02139; and Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts 02114

William T. Peake
Eaton-Peabody Laboratory of Auditory Physiology and Department of Otolaryngology, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114; Speech and Hearing Sciences Program, Harvard–M.I.T. Division of Health Sciences and Technology, Cambridge, Massachusetts 02139; and Department of Electrical Engineering and Computer Science and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 27 December 2000; revised 1 March 2001; accepted 24 May 2001)

Sound transmission through ears with tympanic-membrane (TM) perforations is not well understood. Here, measurements on human-cadaver ears are reported that describe sound transmission through the middle ear with experimentally produced perforations, which range from 0.5 to 5.0 mm in diameter. Three response variables were measured with acoustic stimulation at the TM: stapes velocity, middle-ear cavity sound pressure, and acoustic impedance at the TM. The stapes-velocity measurements show that perforations cause frequency-dependent losses; at low frequencies losses are largest and increase as perforation size increases. Measurements of middle-ear cavity pressure coupled with the stapes-velocity measurements indicate that the dominant mechanism for loss with TM perforations is reduction in pressure difference across the TM; changes in TM-to-ossicular coupling generally contribute less than 5 dB to the loss. Measurements of middle-ear input impedance indicate that for low frequencies, the input impedance with a perforation approximates the impedance of the middle-ear cavity; as the perforation size increases, the similarity to the cavity's impedance extends to higher frequencies. The collection of results suggests that the effects of perforations can be represented by the path for air-volume flow from the ear canal to the middle-ear cavity. The quantitative description of perforation-induced losses may help clinicians determine, in an ear with a perforation, whether poor hearing results only from the perforation or whether other pathology should be expected. [DOI: 10.1121/1.1394195]

PACS numbers: 43.64.Ha [BLM]

I. INTRODUCTION

Perforations of the tympanic membrane (TM) can result from trauma, middle-ear disease, or the treatment of middle-ear disease. Perforations occur as a result of the disease process in chronic otitis media, which affects at least 0.5% of the population (Sadé, 1982). Additionally, it is estimated that 1.3% of American children have tympanostomy tubes (Bright *et al.*, 1993), which are tubes placed in the TM that, like a perforation, connect the ear canal to the middle-ear cavity. Although TM perforations occur frequently, their effects on hearing are uncertain: ''There is no general agreement among clinicians about the magnitude and the configu-

[a)]Author to whom correspondence should be addressed. Address for correspondence: Smith College, Picker Engineering Program, 51 College Lane, Northampton, MA 01063; electronic mail: svoss@email.smith.edu

ration of the hearing loss that is caused by various types of tympanic-membrane perforations'' (Terkildsen, 1976). This paper, along with the theoretical companion paper (Voss *et al.*, 2001d), provides controlled measurements together with a theory for middle-ear function with perforations. Results from both papers will be important in the future design of controlled clinical experimental studies.

To determine the effects of perforations, we have made measurements on human cadaveric temporal-bone preparations with controlled perforations in otherwise normal ears. The work reported here is organized both (1) to describe how perforations alter middle-ear function from normal and (2) to determine the relative importance of particular mechanisms that contribute to hearing loss with perforations. The companion paper (Voss *et al.*, 2001d) uses these experimental results to develop a mathematical model of sound transmission with perforations.

We present measurements of several acoustic quantities with perforations. The ratio between the stapes velocity (V_S) and the sound pressure in the ear canal at the TM (P_{TM}) is our measure of middle-ear sound transmission to the cochlea, i.e.,

$$T \equiv \frac{V_S}{P_{\mathrm{TM}}} \equiv \text{middle-ear transmission.} \tag{1}$$

Perforations of the TM may change middle-ear sound transmission through at least three mechanisms.

(1) Perforations may alter the pressure difference across the TM, which drives the motion of the TM and ossicular chain, and thereby change ossicular motion (e.g., Mehmke, 1962; McArdle and Tonndorf, 1968; and Kruger and Tonndorf, 1977, 1978).

(2) Perforations may alter the coupling between the pressure difference across the TM and the malleus motion, thereby changing ossicular motion. Specific suggestions that fall in this category include (a) decrease in the effective TM area (e.g., Austin, 1978; Shambaugh, 1967), (b) change in the coupling between TM motion and malleus motion, and (c) change in tension of the TM that results from disruption of its fibrous structure (Lim, 1970).

(3) Perforations may alter the sound pressures that act directly on the oval and round windows (e.g., Shambaugh, 1967; Hughes and Nodar, 1985, p. 72; Schuknecht, 1993b, p. 196), thereby changing the pressure difference between the windows, which is a component of the stimulus to the cochlea (Voss *et al.*, 1996).

Measurements reported here determine the relative importance of mechanisms (1) and (2) above; we have shown elsewhere that mechanism (3) is not an important route of sound transmission with most perforations (Voss 1998, Chap. 3). To separate the effects of mechanisms (1) and (2), we use measurements of stapes velocity (V_S), pressure at the TM (P_{TM}), and middle-ear cavity pressure (P_{cav}), to define two ratios,

$$H_{\Delta \mathrm{TM}} \equiv \frac{P_{\mathrm{TM}} - P_{\mathrm{cav}}}{P_{\mathrm{TM}}} \equiv \text{pressure-difference ratio,} \tag{2}$$

and

$$H_{\mathrm{TOC}} \equiv \frac{V_S}{P_{\mathrm{TM}} - P_{\mathrm{cav}}} \equiv \text{TM coupling ratio.} \tag{3}$$

[The subscript TOC stands for TM–ossicles–cochlea.] H_{TOC} is a measure of sound transmission through the TM and ossicular chain that eliminates the effects of changes in pressure difference across the TM.

The product of the two ratios is middle-ear transmission T, i.e.,

$$T \equiv \frac{V_S}{P_{\mathrm{TM}}} = \underbrace{\frac{P_{\mathrm{TM}} - P_{\mathrm{cav}}}{P_{\mathrm{TM}}}}_{H_{\Delta \mathrm{TM}}} \underbrace{\frac{V_S}{P_{\mathrm{TM}} - P_{\mathrm{cav}}}}_{H_{\mathrm{TOC}}}. \tag{4}$$

Changes in the factors, $H_{\Delta \mathrm{TM}}$ and H_{TOC}, provide measures of the importance of mechanisms (1) and (2) in determining transmission loss with perforations. As we will see, perforation-induced changes in transmission (i.e., changes in V_S/P_{TM}) result primarily from one of these mechanisms.

In addition to measurements of the quantities in Eq. (4), we also present measurements of the impedance at the TM (Z_{TM}). These impedance measurements are critical in defining a model of the middle ear with a perforation (Voss *et al.*, 2001d), and they also allow estimation of the effects of perforations on the sound pressure generated at the TM by audiologic earphones (Voss *et al.*, 2000a, e).

II. METHODS

A. Temporal bones and their preparation

Acoustic measurements of stapes velocity, middle-ear cavity pressure, and impedance in the ear canal near the tympanic membrane (TM) were made in cadaveric temporal bones with both normal and perforated TMs. The subjects were 11 of the ears for which normal results are presented in Voss *et al.* (2000b), where measurement techniques are described in detail. Measurements are reported from 100 to 4000 Hz, as some measurements above 4000 Hz were in the noise floor.

Measurements were made on fresh temporal bones for which no evidence of otologic disease was found either in medical records or in oto-microscopic examination. In each temporal bone, the ear canal was drilled away to expose the TM, and a brass ring was cemented to the bony rim around the TM to allow repeatable coupling to the sound source. Access to the stapes footplate was obtained by opening the ''facial recess'' from a posterior-tympanotomy approach (e.g., Shambaugh and Glasscock, 1980, pp. 704–705). To increase visibility of the stapes, (1) the stapedius tendon was cut with alligator-type surgical scissors, (2) the pyramidal process was curetted away, and (3) the mastoid segment of the facial nerve was removed. A 0.25 mm^2 piece of reflective tape, coated with 50 μm polystyrene spheres packed side by side and weighing 0.05 mg, was placed on either the stapes footplate or the posterior crust of the stapes. Voss *et al.* (2000b) and Voss (1998) illustrate the (minimal) effects of these manipulations on the measured stapes velocity. The middle-ear cavity pressure was measured adjacent to the stapes via a probe tube that was cemented to the temporal

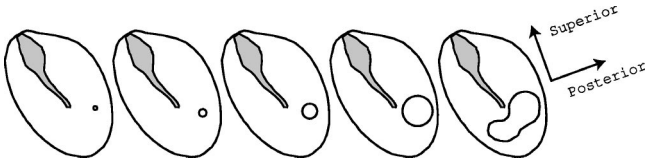


FIG. 1. Schematic drawing of the TM and the manubrium of the malleus that illustrates how perforations of increasing size were made. The diameters for the illustrated circular perforations are, from left to right, 0.5, 1, 2, and 4 mm.

bone and terminated by a calibrated microphone. Measurements were made with the middle-ear cavity closed. Periodically during the measurement sessions, the bone was submerged in normal saline for several seconds, and the excess fluid removed with gentle suction.

B. Perforation of the tympanic membrane

In each ear, perforations of different sizes and locations were made. Typically, increasing-sized circular perforations were made in the pars tensa in either the posterior-inferior (PI) or the anterior-inferior (AI) quadrant. In most cases, once an entire quadrant was perforated, a kidney-shaped perforation that included both inferior quadrants was made so that most of the inferior half of the TM was removed. A typical sequence of perforations is illustrated in Fig. 1.

The controlled perforations were made with an otosurgical Argon laser (i.e., an HGM, Inc. laser used at a power level of 1–2 mW and pulse duration of 100 ms). In order to measure the perforation sizes, an image of the TM was video taped as each perforation was made, and a scale was placed next to the perforation. Later, the video tape was viewed with a video monitor and the perforation was traced and measured to obtain its area. We report a diameter calculated from the area measurement for each perforation. For kidney-shaped perforations that involve more than one quadrant, we report the diameter of a circular perforation of the same area.

C. Calculation of pressure at the tympanic membrane

We measured the ear-canal sound pressure about 3 mm lateral to the TM, as described in detail in Voss *et al.* (2000b) and Voss (1998). To correct for differences in pressure between this location and the TM we use a lossless cylindrical-tube model to represent the residual ear-canal air space (e.g., Møller, 1965; Rabinowitz, 1981; Lynch *et al.*, 1994; Huang *et al.*, 1997), and we estimate the pressure P_{TM} at the TM as

$$P_{\mathrm{TM}}=\frac{P_{\mathrm{EC}}Z_{\mathrm{TM}}}{Z_{\mathrm{TM}}\cos(kl)+jZ_0\sin(kl)}, \tag{5}$$

where P_{EC} is the measured pressure at the microphone probe-tube orifice; Z_{TM} is the measured impedance at the TM (see Sec. II G); $Z_0=\rho c/A$ is the characteristic impedance of the tube, and $A=70\,\mathrm{mm}^2$ is the area of the tube which is defined by the average area of the TM (Wever and Lawrence, 1954, p. 416); l is the length of the tube and is defined as $l=V_{\mathrm{EC}}/A$, where V_{EC} is the ear-canal air volume between the microphone probe-tube orifice and the TM, which is measured by filling the residual ear-canal air space with saline using a calibrated syringe (range 0.03–0.1 cm^3); $k=2\pi f/c$, ρ is the density of air, c is the velocity of sound in air, and f is frequency.

The ratio $P_{\mathrm{TM}}/P_{\mathrm{EC}}$ is less than 1 dB in magnitude and only a few hundredths of a cycle in angle for perforations that are less than 1 mm in diameter. At the other extreme, when the TM is entirely removed, the ratio $P_{\mathrm{TM}}/P_{\mathrm{EC}}$ can approach ± 2 dB in magnitude and 0.05 cycles in angle for frequencies above 1000 Hz. Below 1000 Hz, the differences between P_{EC} and P_{TM} are smaller than 1 dB for all TM conditions.

In this paper, we apply the sound-pressure correction of Eq. (5) to convert measurements of P_{EC} to P_{TM}. However, as measurements of Z_{TM} were not obtained on one ear (bone 18), we use P_{EC} for all results on bone 18.

D. Measurement of the stapes velocity

1. Scope

We use our measurements of T (i.e., V_S/P_{TM}) as a description of the sound transmission through the middle ear. Our techniques are described in Voss *et al.* (2000b) and Voss (1998). Specifics relevant to the measurements with perforations are presented here.

2. Acoustic stimuli

Unless otherwise noted, the measurements of stapes velocity were made in response to tonal stimuli that were typically between 90 and 120 dB SPL. [The system behaved linearly for these levels, as described by Voss *et al.* (2000b).] For each tone, the two responses, ear-canal pressure and stapes velocity, were typically the average of 1000 or 2000 41-ms-long responses for a total of 41 or 82 s of averaging. The measured frequency range was 25 to 10 000 Hz; the frequency resolution was not identical for all experiments. Because the experiments on the first three ears had a poor frequency resolution (only 16 or 25 logarithmically spaced points), the reported means include only the experiments on the final eight ears, which had 68 logarithmically spaced points. [In all cases, data from the first three ears are consistent with those from the final eight ears.]

3. Removal of data affected by the mechanical artifact

Our stapes-velocity measurements are affected by a mechanical artifact, which appears to result from vibration of the sound source. This artifact confounds the measurement of stapes velocity. We measured the artifact as the velocity of the temporal bone (at the round-window niche) in response to sound in the ear canal. Specifically, we placed a piece of reflective tape on the bony round-window niche and measured the ratio between this temporal-bone velocity and the ear-canal pressure, $V_{\mathrm{bone}}/P_{\mathrm{TM}}$. With a normal TM, $|V_{\mathrm{bone}}/P_{\mathrm{TM}}|$ was usually more than 20 dB smaller than $|V_S/P_{\mathrm{TM}}|$, and therefore the artifact was negligible. However, as perforations were made, the stapes velocity decreased for a constant pressure at the TM, and the mechanical artifact could interfere with measurement of the stapes velocity. Unless noted, data corresponding to stapes-velocity

magnitudes within 20 dB of the artifact magnitude $|V_{\text{bone}}/P_{\text{TM}}|$ are not included. Further details are in Voss (1998, pp. 34–35, Fig. 1–5).

The requirement that the stapes-velocity magnitude be at least 20 dB greater than the artifact's magnitude has the effect of eliminating data in which the stapes-velocity magnitude is smallest, i.e., at the lowest frequencies with the largest perforations.

4. Definition of transmission loss

To describe sound transmission with perforations, we compare the transmission measured with an intact TM, T^{norm} [i.e., $(V_S/P_{\text{TM}})^{\text{norm}}$], to the same ratio measured with a perforated TM, T^{perf}. The ratio between the normal and the perforated condition serves as our measure of transmission loss

$$\Delta T \equiv \frac{T^{\text{norm}}}{T^{\text{perf}}} \equiv \frac{(V_S/P_{\text{TM}})^{\text{norm}}}{(V_S/P_{\text{TM}})^{\text{perf}}} \equiv \text{transmission loss}. \quad (6)$$

E. Pressure-difference ratio, $H_{\Delta\text{TM}}$

1. Calculation of the pressure-difference ratio

We calculate the pressure difference across the TM from measurements of ear-canal pressure and middle-ear cavity pressure. The ear-canal pressure, P_{EC}, is generated and measured with the acoustic assembly described by Voss *et al.* (2000b), and P_{TM} is computed from P_{EC} via Eq. (5). The middle-ear cavity pressure, P_{cav}, is measured with a probe-tube microphone that is placed near the stapes footplate within the middle-ear cavity. Details of this probe-tube microphone and its calibration are in Voss *et al.* (2000b). We compute $H_{\Delta\text{TM}}$, the pressure-difference ratio, as

$$H_{\Delta\text{TM}} \equiv \frac{\Delta P_{\text{TM}}}{P_{\text{TM}}} \equiv \frac{P_{\text{TM}} - P_{\text{cav}}}{P_{\text{TM}}}. \quad (7)$$

To describe how perforations affect $H_{\Delta\text{TM}}$, we compute changes in $H_{\Delta\text{TM}}$ for the intact TM relative to the perforated TM as

$$\Delta H_{\Delta\text{TM}} \equiv \frac{H_{\Delta\text{TM}}^{\text{norm}}}{H_{\Delta\text{TM}}^{\text{perf}}} \equiv \text{change in } H_{\Delta\text{TM}}. \quad (8)$$

2. Limits on the accuracy of pressure-difference calculations

With the TM perforated, the middle-ear cavity pressure and the pressure at the TM are nearly equal at low frequencies. Thus, small errors in the relative calibration of the two microphones that measure these two pressures may introduce large errors in the computed pressure difference. The calibration procedure for the microphone that measures P_{cav} is described by Voss *et al.* (2000b). This microphone is calibrated relative to the microphone that measures P_{EC} by acoustically coupling the two microphones together and comparing their responses to the common stimulus. To estimate the variability in the relative calibrations between the two microphones, we examined repeated relative calibrations over the course of a single experiment. In general, the ratio between repeated calibrations during an experiment varied as much as a factor of 1.1 in magnitude and 0.01 cycles in angle. With such an error and with P_{cav} equal to P_{TM}, a measurement would yield

$$|H_{\Delta\text{TM}}| = \left| \frac{P_{\text{TM}} - 1.1 P_{\text{TM}} e^{j2\pi 0.01}}{P_{\text{TM}}} \right| = 0.12, \quad (9)$$

although $H_{\Delta\text{TM}}$ would really be zero. Thus, measured $|H_{\Delta\text{TM}}|$ values of less than 0.12 may be highly affected by calibration errors. Thus, we impose the lower limit of 0.12 on the magnitude of the pressure-difference ratio ($|H_{\Delta\text{TM}}|$) computed from Eq. (7), and any calculated $|H_{\Delta\text{TM}}|$ that is below 0.12 is eliminated from the results.

3. Acoustic stimuli for cavity-pressure measurements

The voltages that correspond to the pressures P_{EC} and P_{cav} are measured simultaneously on two channels in response to a chirp stimulus. The chirp contains 1024 linearly spaced frequencies from 24 to 25 000 Hz. The reported response is the average of 200 responses (8.2 s of averaging).

F. TM coupling ratio, H_{TOC}

Perforation-induced changes in H_{TOC} are a measure of sound transmission through the TM and ossicular chain that eliminates the effects of changes in pressure differences across the TM with different perforations. H_{TOC} is calculated from measurements as

$$H_{\text{TOC}} \equiv \frac{V_S/P_{\text{TM}}}{(P_{\text{TM}} - P_{\text{cav}})/P_{\text{TM}}} \equiv \frac{V_S/P_{\text{TM}}}{H_{\Delta\text{TM}}}. \quad (10)$$

Changes in H_{TOC} are calculated as the ratio of the normal $H_{\text{TOC}}^{\text{norm}}$ to the perforated $H_{\text{TOC}}^{\text{perf}}$.

$$\Delta H_{\text{TOC}} \equiv \frac{H_{\text{TOC}}^{\text{norm}}}{H_{\text{TOC}}^{\text{perf}}} \equiv \text{change in } H_{\text{TOC}}. \quad (11)$$

To calculate H_{TOC} [Eq. (10)], some data manipulation was required. The measurements of V_S/P_{TM} were made with a frequency resolution of 68 points from 24 to 10 000 Hz, whereas the measurements of $H_{\Delta\text{TM}} = \Delta P_{\text{TM}}/P_{\text{TM}}$ [Eq. (7)] were made with a frequency resolution of 1024 points from 24 to 25 000 Hz. Thus, the two kinds of measurements are not at exactly the same frequencies. We resolved this problem through interpolation of the $H_{\Delta\text{TM}}$ data to the frequencies of the V_S/P_{TM} data, where the $H_{\Delta\text{TM}}$ data have the larger frequency resolution of the two measurements. Interpolation is done using cubic spline interpolation performed using the software package MATLAB (The Mathworks, Inc.). Note that the displayed H_{TOC}'s do not include points where either (1) $|V_S/P_{\text{TM}}|$ is within 20 dB of the mechanical artifact or (2) $|H_{\Delta\text{TM}}| < 0.12$.

G. Measurement of the middle-ear input impedance

Acoustic impedance measurements were made with a method similar to that used by others (e.g., Rabinowitz, 1981; Allen, 1986; Lynch *et al.*, 1994), which is thoroughly discussed elsewhere (Voss, 1998; Voss *et al.*, 2000b). The Thévenin equivalent of the sound-delivery system was determined by pressure measurements in two ''reference loads'' of known theoretical impedance. The two theoretical imped-

ances were calculated from the equations of Egolf (1977) and combined with the pressure measurements made in the two reference loads to calculate the source's Thévenin pressure and impedance equivalents P_{TH} and Z_{TH}. The ear's impedance was then calculated from a pressure measurement in response to a chirp stimulus in the ear canal. The impedance measurements were estimated to be generally accurate to within about 10% in magnitude (about 1 dB) and 10° in angle (0.025 cycles).

Impedances were measured with a different acoustic assembly from that used for the stapes-velocity and the middle-ear cavity pressure measurements. (The source used for the stapes-velocity and middle-ear cavity pressure measurements had a source impedance magnitude that was small compared to the magnitude of the input impedance of the middle ear; thus, it could not be used to make accurate impedance measurements.) The impedance-measurement assembly consisted of a Knowles ED-1913 hearing aid receiver as a sound source (Knowles Electronics, Elk Grove, IL) and a Knowles EK-3027 microphone (Ravicz *et al.*, 1992; Voss, 1998). (The source used to measure impedance in the ear canal was not useful for the stapes velocity measurements because it could not generate large enough sound-pressure levels in all situations.)

The ear's impedance was measured in the ear canal (1) with an intact TM, (2) after each perforation was made (Z_{TM}^{perf}), and (3) with the TM removed. The impedance measurement was made at the same ear-canal location as the ear-canal pressure (Sec. II C), and we approximate the effect of the ear-canal air volume on the measured impedances by a lossless cylindrical-tube model of the residual ear-canal air space (e.g., Møller, 1965; Rabinowitz, 1981; Lynch *et al.*, 1994; Huang *et al.*, 1997), so that the impedance at the TM (Z_{TM}) is determined from the impedance measured in the ear canal Z_{EC} as

$$Z_{TM} \equiv Z_0 \frac{Z_{EC} - jZ_0 \tan(kl)}{Z_0 - jZ_{EC} \tan(kl)}, \tag{12}$$

where the variables in Eq. (12) are defined in conjunction with Eq. (5).

With an intact TM, the ratio $|Z_{EC}/Z_{TM}|$ approaches -3 dB with the largest ear-canal volume of 0.1 cm^3; in cases with smaller ear-canal air volumes, $|Z_{EC}/Z_{TM}|$ is between 0 and -1 dB. With perforations, the ratio $|Z_{EC}/Z_{TM}|$ can approach 3 dB for frequencies near 2000–4000 Hz.

III. RESULTS

A. Organization

The results are organized into five sections. Each of the first four sections focuses on one of the four measured ratios for a series of perforations: (1) middle-ear transmission $T \equiv V_S/P_{TM}$; (2) the pressure-difference ratio $H_{\Delta TM}$; (3) the TM coupling ratio H_{TOC}; and (4) the middle-ear input impedance Z_{TM}. In each section, we show data from a typical "example" ear (bone 24L). Additionally, we show the means from eight ears for the quantities ΔT, $\Delta H_{\Delta TM}$, and ΔH_{TOC}. In all cases, results from the example ear are similar to those from the other ears, which are plotted in the appendices of Voss (1998). Measurements were made on eleven ears. However, measurements on the first three ears had poor frequency resolution (only 16 or 25 logarithmically spaced points) whereas the experiments on the final eight ears had a frequency resolution of 68 logarithmically spaced points. For this reason, the reported means include only the final eight experiments. In all cases, data from the first three ears are consistent with those from the final eight ears.

A basic conclusion supported by the results is that the effects of perforations on sound transmission result primarily from perforation-induced changes in the pressure difference across the TM, with alterations in TM-to-ossicular-coupling (H_{TOC}) being relatively small. One experiment, reported in Sec. III F, was designed to determine what modifications in the TM are required to make major changes in this coupling; this experiment determined the effects of extensive slits in the TM on H_{TOC}.

B. Middle-ear transmission *T*

1. Components of stapes motion

Voss *et al.* (2000b) show that the stapes translates in and out of the oval window with a piston-like motion for frequencies up to at least 2000 Hz when the TM is normal. Voss *et al.* (2000b) argue that to be consistent with translational motion, the ratio between the velocities measured at two locations on the stapes has to (1) have a magnitude that is constant versus frequency, and (2) have an angle that is zero. To determine whether the stapes motion is piston-like when the TM is perforated, stapes-velocity measurements at two stapes locations were made on two ears with perforations [two of the five ears in Fig. 6 of Voss *et al.* (2000b)]. For each perforation condition, the ratio of the complex velocities measured at two locations was computed. In Fig. 2, the magnitudes and angles of the velocity ratios associated with these two ears are plotted for four perforation sizes. Below 2000 Hz, the magnitudes and angles with the perforations appear similar to those shown for the normal TM: the magnitudes are nearly constant and the angles are near zero, consistent with a translational motion. Above 2000 Hz, there is more variability in both the magnitudes and angles, which might result from a more complicated motion. However, moderate changes in the preparation between measurements could produce the differences seen in Bone 29. In neither ear do the changes have a systematic dependence on perforation size, which is consistent with another source for the change. In summary, up to at least 2000 Hz, the stapes appears to move with a one-dimensional translational motion with both a normal and a perforated TM.

2. Effects of perforations on transmission

a. Example ear. Measurements of middle-ear transmission T (i.e., V_S/P_{TM}) from the example ear are shown as a function of frequency with perforation diameter as a parameter in Fig. 3 (left). The magnitude and angle of T have several features that are consistent across all preparations. First consider the magnitude $|T|$: (1) As perforation diameter increases, $|T^{perf}|$ decreases systematically at frequencies below 1000–2000 Hz; (2) at frequencies less than 1000 Hz, for all perforation diameters, $|T^{perf}|$ increases with increasing

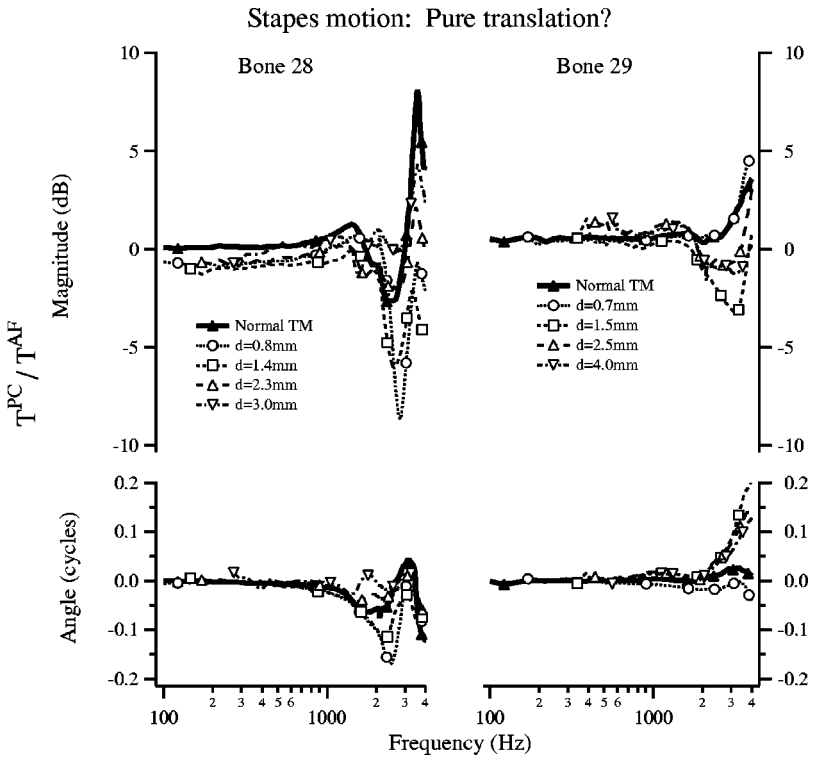


FIG. 2. Magnitude and angle of ratio of stapes velocities measured at two locations on the stapes. $T \equiv V_S / P_{TM}$ was measured at two locations on the stapes of bones 28 and 29 with the TM both normal and perforated. For both bones, the measurement locations were the anterior end of the stapes footplate (AF) and the posterior crus (PC). Thus, for each TM condition noted in the legends, "Magnitude (dB)" refers to $20 \log_{10}|T^{PC}/T^{AF}|$, and "Angle" refers to the difference $\angle(T^{PC}) - \angle(T^{AF})$. The middle-ear cavities were open for the measurements on bone 28 and sealed for the measurements on bone 29. Measurement stimuli were chirps, and symbols that distinguish between measurements are plotted at every 30th data point.

frequency such that in the 1000 Hz range $|T^{perf}|$ approaches the normal value, and in many cases $|T^{perf}|$ exceeds $|T^{norm}|$ slightly; and (3) above 1000 Hz, the perforations' effects are generally smaller than at frequencies below 1000 Hz.

Next, consider the angle $\angle T$. With the TM intact, $\angle T^{norm}$ is constant at about 0.25 cycles at frequencies up to at least 500 Hz, and above about 500 Hz, $\angle T^{norm}$ decreases gradually with increasing frequency. When the TM is perforated: (1) At low frequencies, $\angle T^{perf}$ is roughly constant with frequency but its low-frequency asymptote appears to be between 0.25 and 0.75 cycles. Larger perforations result in larger low-frequency angles; and (2) as frequency increases,

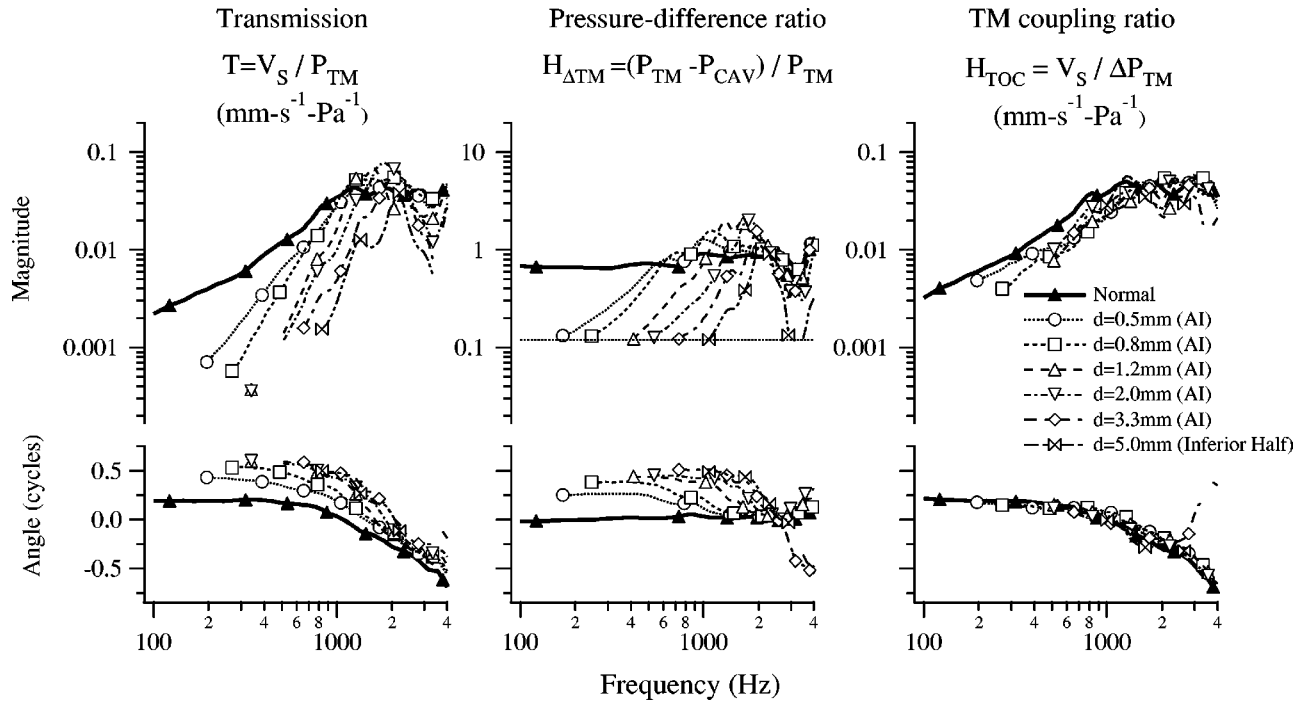


FIG. 3. Measurements of T (left), $H_{\Delta TM}$ (center), and H_{TOC} (right) from the example ear (Bone 24L). For each quantity, the magnitude (upper) and angle (lower) of each measurement are plotted. The legend in the rightmost panel specifies the perforation diameter. The smaller perforations were in the anterior-inferior (AI) quadrant, and the largest perforation contained most of the inferior half of the TM. Symbols are plotted at every eighth data point for T and H_{TOC} and every 25th data point for $H_{\Delta TM}$. "Gaps" between T data points result when measurements at some frequencies are excluded by the mechanical artifact criterion. The horizontal dashed line in $|H_{\Delta TM}|$ indicates a pressure difference of 0.12, which is the lower limit for the calculated $|H_{\Delta TM}|$.

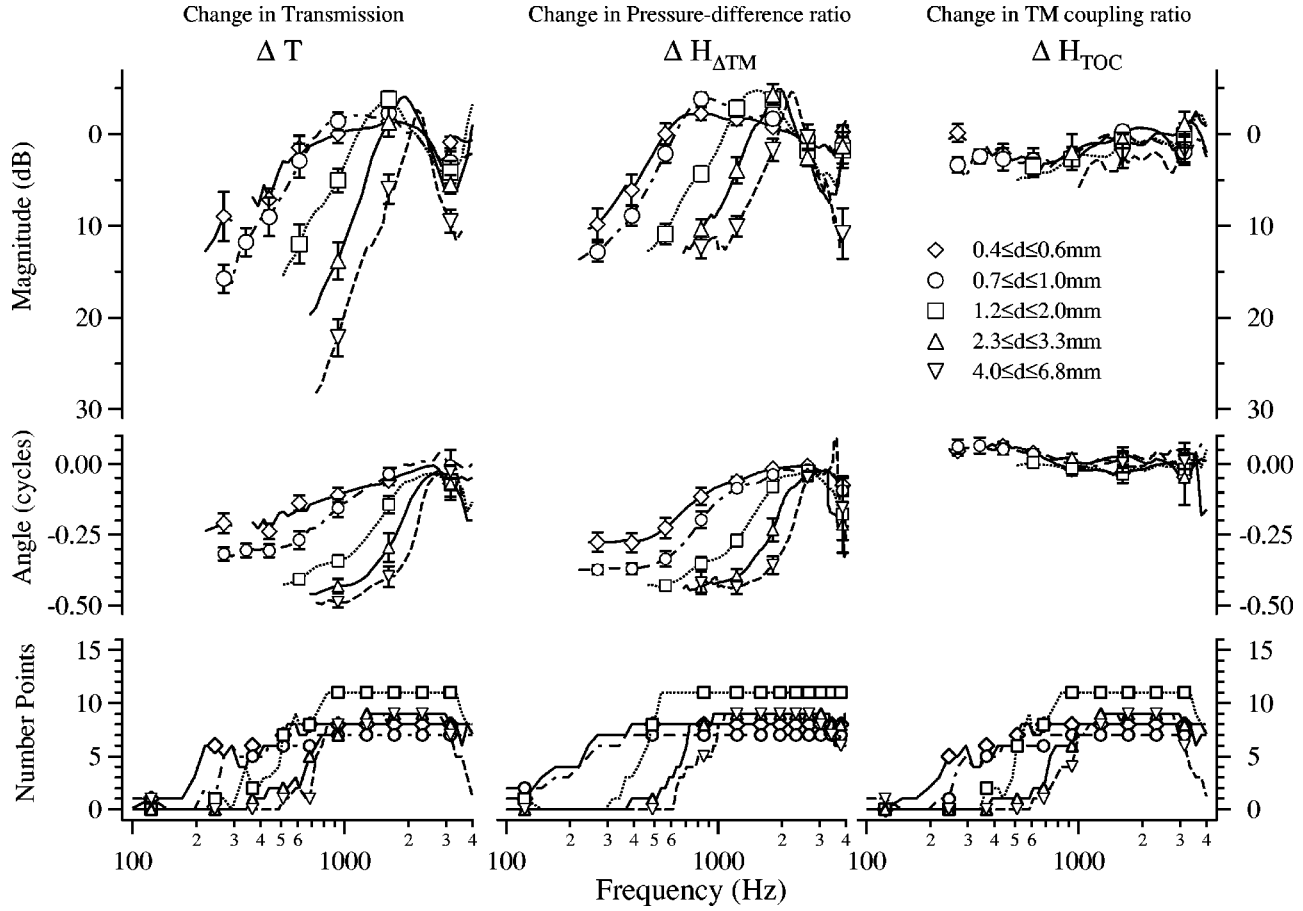


FIG. 4. Mean change (±standard error) in T (left), $H_{\Delta TM}$ (center), and H_{TOC} with perforation diameter as a parameter; the changes from normal are defined as the ratio between the normal measurement and the perforated measurement (i.e., norm/perf), with |norm/perf| plotted in dB with positive down so that these plots have the format of standard audiograms. Mean magnitudes (upper) were calculated in the logarithmic domain, and mean angles (middle) in the linear domain. Changes from all perforations are included from the final eight experiments. Each change curve, with diameter as a parameter, includes all perforations made on the eight ears that fit the diameter category. The lower plots indicate the total number of measurements in each diameter category that are available at each frequency: N is not constant across frequency because points where either V_S/P_{TM} is within 20 dB of the artifact or where $|H_{\Delta TM} < 0.12|$ are not included in calculation of the means. Means are only calculated at frequencies where data from more than 50% of the cases in the range category are above our noise limits. To increase visibility, symbols and standard error bars are indicated at a subset of data points.

$\angle T^{perf}$ decreases and approaches the value for the intact TM $\angle T^{norm}$.

b. Means from eight ears. In this section, we compare the effects of perforations on T for measurements made on eight ears. In order to compare the effect of the perforation across different ears that vary in their baseline (i.e., intact TM) responses, we compare changes from normal in stapes velocity [i.e., ΔT of Eq. (6)].

A total of 44 perforations were made on the eight ears. To compare the effects of all 44 perforations, we grouped the perforations by size. Diameter categories were selected by grouping the 44 perforations into 5 categories so that (1) there were approximately the same number in each group (range 7–11) (Fig. 4 lower-left), (2) the ratio of the maximum-to-minimum perforation diameter in each group is similar for all groups (range 1.4–1.7), and (3) gaps between groups are similar across all five groups (15%–20%).

The mean perforation-induced changes from the intact condition (i.e., transmission loss) are plotted in Fig. 4 (left-hand side). The mean changes in transmission, ΔT, are consistent with the features described above for the measurements from the example ear: (1) ΔT is frequency dependent with the largest reductions in magnitude occurring at low frequencies; (2) at low frequencies, ΔT increases with perforation size; (3) between 1000 and 2000 Hz, there can be small increases (i.e., 3 dB) in ΔT with perforations; both the frequency and the sharpness of this peak increase with perforation-diameter increases; (4) from about 2000 to 4000 Hz, changes are less than 12 dB and increase with perforation size; and (5) for the lowest frequencies the mean angle change approaches -0.25 cycles for the smallest perforations and -0.5 cycles for the larger perforations; for the middle frequencies the angle changes are near zero, and for the higher frequencies the mean angle changes are between zero and -0.25 cycles.

C. Pressure-difference ratio: $H_{\Delta TM}$

Measurements of the pressure-difference ratio $H_{\Delta TM}$ from the example ear are shown as a function of frequency with perforation diameter as a parameter in Fig. 3 (center). With the TM intact, the magnitude ranges from about 0.7 to 1 and is nearly independent of frequency; the angle is about zero. Perforations have their biggest effects on $H_{\Delta TM}$ at fre-

quencies below about 1000 Hz, where many of the effects are similar to those described previously for T. For example: (1) the low-frequency magnitude $|H_{\Delta TM}|$ decreases systematically as perforation diameter increases and (2) $|H_{\Delta TM}|$ has a maximum where the ''perforated'' measurement exceeds the normal measurement, and the frequency of this maximum increases with perforation diameter. At the frequency of the maximum, the angle $\angle H_{\Delta TM}$ begins to decrease toward 0 cycles: For the smallest perforation this angle decrease is about 0.25 cycles and for the largest perforations the angle decrease approaches 0.5 cycles. For the larger perforations, the combination of the magnitude maximum and the half-cycle angle change suggests that the perforation introduces a resonance—which might involve the effective mass of the perforation and the compliance of the middle-ear cavity (Voss *et al.*, 2001d).

Figure 4 (center) plots the mean changes from normal in $H_{\Delta TM}$ [i.e., $\Delta H_{\Delta TM}$ of Eq. (8)]. Both the magnitude and the angle of the changes from normal are similar to the measurements of $H_{\Delta TM}$ (Fig. 3 center) because with a normal TM $H_{\Delta TM}$ is nearly one in magnitude and zero in angle.

D. TM coupling ratio: H_{TOC}

The TM coupling ratio is a measure of signal transmission coupled by the TM and ossicular chain with the effects of perforations on the pressure difference across the TM removed. H_{TOC} with an intact TM is roughly the same as the transmission T with an intact TM (Fig. 3) because $|P_{cav}| < 0.3|P_{TM}|$. Perforations have only moderate effects on H_{TOC} (Fig. 3 right and Fig. 4 right). With all perforations, the changes from normal in H_{TOC} are generally less than 5 dB in magnitude and less than 0.1 cycles in angle for all frequencies. These changes are small compared to the changes in T and $H_{\Delta TM}$. Thus, changes in transmission T appear to result primarily from changes in the pressure difference across the TM, with smaller changes in the way the TM couples to the cochlea. We address this finding further in Sec. III F. [One limitation on this description results from the lack of measurements of either stapes velocity or pressure difference across the tympanic membrane at the lowest frequencies for the larger perforations, as measurements of V_S were limited at the lowest frequencies by the mechanical artifact and measurements of ΔP_{TM} with perforations yielded low-frequency pressure differences that were too small to measure accurately.]

E. Impedance at the tympanic membrane: Z_{TM}

In Fig. 5, measurements of the acoustic impedance at the TM [Eq. (12)] are plotted for our example ear with a normal (Z_{TM}^{norm}), perforated (Z_{TM}^{perf}), and removed TM (Z_{cav}). Below 500 Hz, Z_{TM} is compliance-like for all conditions, i.e., the slope of the magnitude $|Z_{TM}|$ approximates -20 dB/decade, and the low-frequency angle of $\angle Z_{TM}$ is near -0.20 cycles.

First, consider the perforation's effect on the impedance magnitude. At the lower frequencies, a perforation reduces $|Z_{TM}^{norm}|$ by a constant, frequency-independent factor of about 0.3. This reduction is independent of perforation diameter. Above about 500 Hz, $|Z_{TM}^{perf}|$ shows perforation-diameter dependent variations. First, $|Z_{TM}^{perf}|$ has a well-defined local minimum that depends on diameter in a systematic way; as perforation diameter increases, the frequency of the minimum increases, the magnitude at the minimum frequency decreases, and the sharpness of the minimum increases. Second, $|Z_{TM}^{perf}|$ has a well-defined local maximum around 3000 Hz; the frequency of this maximum is not affected by the perforation diameter and the magnitude of the maximum decreases as perforation diameter increases.

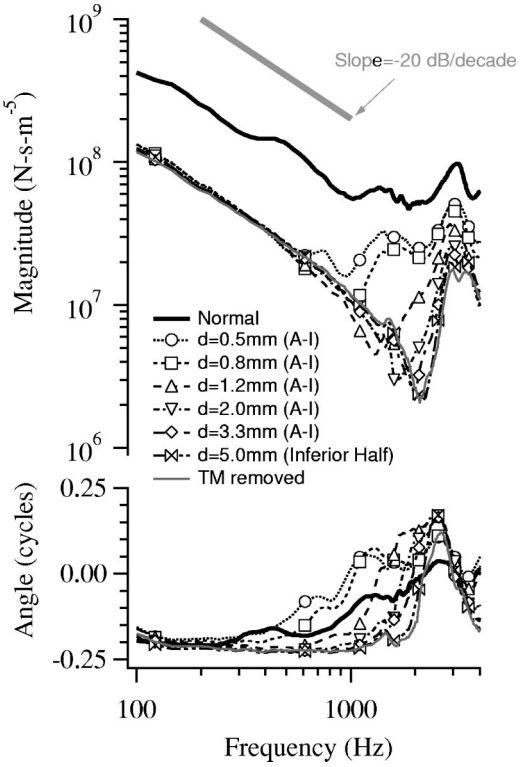


FIG. 5. Impedances at the TM measured on the example ear (bone 24 Left) for the normal TM (Z_{TM}^{norm}), perforated TM (Z_{TM}^{perf}), and TM removed (Z_{cav}) conditions. Symbols indicate every 20th data point. Upper: magnitude, lower: angle.

Perforations also affect the angle of the impedance. At low frequencies, the compliant-like angle approximates -0.20 cycles for both the normal and the perforated conditions. As frequency increases, the angle increases to a positive value between 0 and 0.25 cycles. With the moderate- and larger-sized perforations, the increase in angle occurs across the same frequency range as the first local minimum in magnitude described previously. As perforation diameter increases, the transition frequency from compliant-like to resistive and mass-like increases. Around 3000 Hz, corresponding to the local magnitude maximum described previously, $\angle Z_{TM}^{perf}$ decreases.

The measurements of the impedance with the TM removed (designated as Z_{cav}) are included in Fig. 5. Comparison of the measurements of Z_{TM}^{perf} to Z_{cav} indicate that (1) at the lowest frequencies $Z_{TM}^{perf} \approx Z_{cav}$ for all perforation sizes, and (2) at higher frequencies, as the perforation diameter increases, Z_{TM}^{perf} approaches Z_{cav}.

The regular occurrence of a minimum in $|Z_{TM}|$ (Fig. 5) at a frequency where the angle changes rapidly, suggests a

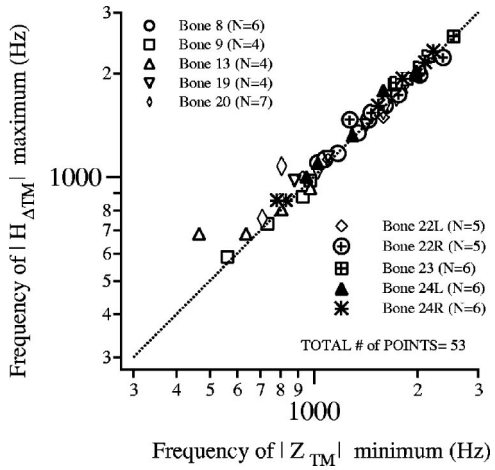


FIG. 6. Comparison of frequencies of the first maximum in $|H_{\Delta TM}|$ to frequencies of the first minimum in $|Z_{TM}|$ for 53 pairs of measurements with perforations (in 10 ears). A symbol is plotted for each perforation including the condition of TM removed. The dotted line is $y=x$.

perforation-dependent resonance involving the acoustic mass of the perforation and the acoustic compliance of the middle-ear cavity. In this case, at the resonant frequency, the pressure difference across the TM should be a maximum [Voss *et al.*, 2001d, Eq. (3)]. This prediction can be tested by comparing the frequencies of the minima of $|Z_{TM}|$ to those of the maxima of $|H_{\Delta TM}|$. Figure 6 shows that these two independent sets of measurements fit this prediction.

F. Effects of slit-like perforations

The results presented previously indicate that even large perforations produce relatively small (i.e., less than 10 dB) changes in the coupling of pressure difference across the TM to the ossicles and cochlea (i.e., ΔH_{TOC} as in Fig. 4, right). This result, which simplifies understanding of the effect of perforations, seems to contradict some conceptions of the TM-to-ossicular chain coupling mechanisms. For example, this result indicates that interruption of a sizable fraction of TM radial fibers (e.g., as in Fig. 1, right) may have little effect on TM-coupling function. To demonstrate larger effects of changes in TM integrity, a structural modification was designed to disconnect more extensively the TM from the manubrium (handle) of the malleus, while minimizing changes in the pressure difference across the TM. In other words, in contrast to the perforations, this modification might produce only small changes in the pressure difference across the TM ($H_{\Delta TM}$) while dramatically changing the coupling H_{TOC}.

In one ear, we slit the TM with a myringotomy knife along the manubrium of the malleus in four stages as schematized in Fig. 7. Each slit completely penetrated the TM, as it was possible to view the middle-ear cavity through the slit.

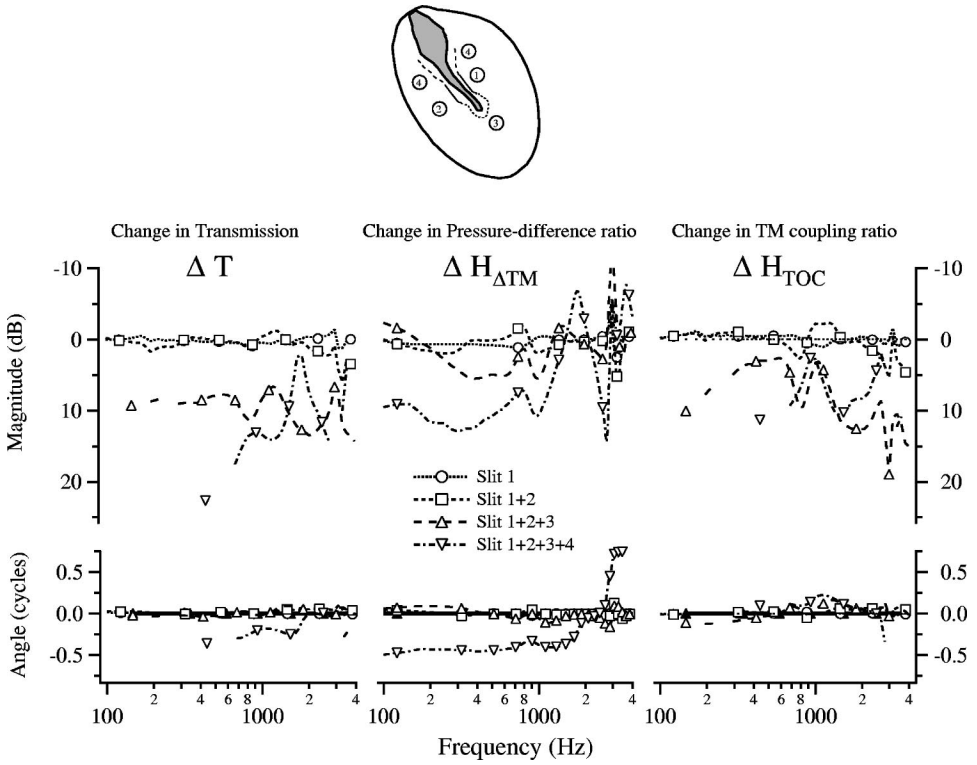


FIG. 7. Change in V_S/P_{TM} (left), $H_{\Delta TM}$ (center), and H_{TOC} with slit configuration the parameter; changes are defined as the ratio between the normal measurement and the perforated measurement (i.e., norm/perf), with |norm/perf| plotted in dB with positive down so that these plots have the format of standard audiograms. The slit configuration is schematized in the center of the figure. Slit 1, indicated by a solid line, was made first along the posterior edge of the manubrium of the malleus. Slit 1 was about 2 mm long. Next slit 2, also about 2 mm long and indicated by a solid line, was made along the anterior border of the manubrium. Next, slit 3, indicated by a dashed line, was made to connect slits 1 and 2 around the distal end of the manubrium (umbo). Finally, slit 4, with two components, was made so that the combined slit encompassed the perimeter of the manubrium and extended approximately to the lateral process of the malleus. Symbols are plotted at every eighth data point for V_S/P_{TM} and H_{TOC} and every 25th data point for $H_{\Delta TM}$. "Gaps" between V_S/P_{TM} data points result when measurements at some frequencies are excluded as a result of the mechanical artifact; these exclusions generally occur when the reduction in stapes velocity is large.

Figure 7 shows changes in measurements made with four slit conditions. First, consider the single 2 mm slit along the posterior part of the manubrium (labeled slit 1). The left column of Fig. 7 indicates that there was no change in T from normal with this first slit. This slit certainly penetrated the TM, but after the slit was made the margins of the slit were observed under a microscope to be bridged by moisture. Figure 7 (center column) shows that slit 1 did not affect the pressure-difference ratio ($H_{\Delta TM}$) either. Similar to slit 1, slit 2 was also self sealing with moisture and the responses with slit 2 added were similar to those with slit 1 only: neither T nor $H_{\Delta TM}$ changed much from normal (<2 dB in magnitude).

Larger changes in T and $H_{\Delta TM}$ occurred with the addition of slits 3 and 4, in which the slits were longer and remained open over their entire length. With the addition of slit 3, the change in $|T|$ is relatively flat below 1000 Hz at about 10 dB, and with the addition of slit 4 the change is between 10 and 25 dB and has some variation with frequency. Examination of the second and third columns of Fig. 7 shows that for the longer third and fourth configurations, changes in $|H_{\Delta TM}|$ increase to up to 10 dB for frequencies from 400 to 2000 Hz, and changes in $|H_{TOC}|$ can approach 10–20 dB at some frequencies. Measurements of T below 400 are not available because of the mechanical artifact, and it is possible that changes in $|H_{TOC}|$ could greatly exceed 20 dB at these lower frequencies. Even though changes in H_{TOC} are larger than with our circular perforations, the changes of only 10–20 dB at higher frequencies in configuration 4 seem inconsistent with the extensive disruption of the manubrium's attachment to the TM. In summary, this experiment suggests that major interruption of the TM structure near the manubrium has only moderate effects (i.e., <20 dB) on H_{TOC} for frequencies greater than 400 Hz.

IV. DISCUSSION

A. Comparison to other work

1. Frequency dependence of loss with perforations

Our measurements of perforation-induced changes in transmission show a clear frequency dependence; Fig. 4 shows that for all perforation sizes, the transmission changes are greatest at the lowest frequency and decrease toward zero as frequency increases toward 1000–2000 Hz. There can be increases (up to 20 dB) in transmission in the 1000–2000 Hz region. Above 2000 Hz, the losses are typically less than 10 dB. (We note that our largest perforations covered 50% of the TM; larger perforations could produce different loss characteristics.)

These results are consistent with the cat cochlear-potential measurements of McArdle and Tonndorf (1968) and Kruger and Tonndorf (1977,1978) (Fig. 8), and the umbo-velocity measurements of Bigelow *et al.* (1996) in rat, in all of which perforations in the TM produced their largest effects at the lowest frequencies. Similarly, the temporal-bone measurements of Nishihara *et al.* (1993) and the audiometric results of both Tavin *et al.* (1988) and Rosowski *et al.* (1996) show that tympanostomy tubes in the TM produced their largest losses at the lowest frequencies. The perforation-induced loss was most prominent at the lowest frequencies and the loss decreased with increasing frequency in the audiometric data of Ahmad and Ramani (1979) (Fig. 8). Our measurements show increases in sensitivity in the 2000 Hz region that are not found in the average of the measurements of Ahmad and Ramani (1979); this difference may result from variations in middle-ear cavity volume (Voss *et al.*, 2001d). In our population of ears, the volumes were very similar for all ears and thus the resonant frequency between the acoustic mass of the perforation and the acoustic compliance of the middle-ear cavity would be similar for perforations of the same size. In a clinical population, however, larger variations in middle-ear cavity volume are likely (Molvaer *et al.*, 1978), and thus this resonant frequency where sensitivity is increased, would not be the same across ears. As a result, averaging audiograms from a clinical population would tend to obscure the region of the increased sensitivities.

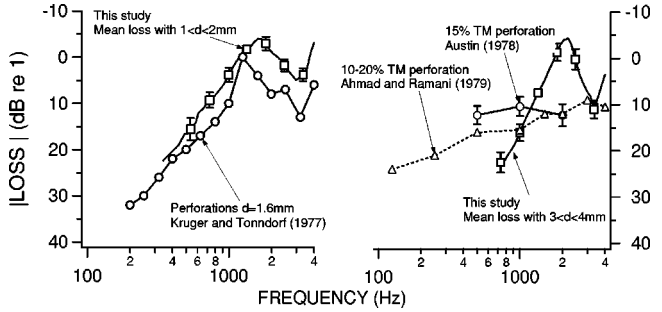


FIG. 8. Comparison of our mean change in transmission (i.e., ''loss'') to measurements of transmission loss from Kruger and Tonndorf (1977) and Austin (1978). Left: Comparison of results of Kruger and Tonndorf (1977) to our mean transmission change (ΔT) with perforations of diameter d such that $1<d\leq 2$ mm (taken from Fig. 4). The perforation diameter of the Kruger and Tonndorf (1977) measurements was $d=1.6$ mm, and their loss measurements were from cat cochlear potential data. Right: Comparison of audiologic data from Austin (1978) with perforations that covered about 15% of the TM (about 3.4 mm in diameter) and perforations from Ahmad and Ramani that covered 10%–20% of the TM. Here, we plot our mean ΔT with perforation diameters d such that $3<d\leq 4$ mm.

Our results disagree with results of Békésy (1936) in human, measurements of Payne and Githler (1951) in cat, and the audiologic data of Austin (1978) (Fig. 8). Using a temporal-bone preparation, Békésy (1936) found no differences in the motion of the malleus with a normal TM and a 0.6-mm-diam perforation for frequencies above 400 Hz, although below 100 Hz he did find a reduction in motion that increased inversely with frequency. However, if Békésy's perforations were effectively closed by moisture on the TM, as happened with our slit experiments (Fig. 7), then his measurements are consistent with our results. Békésy also investigated the effect of a ''lens-shaped tear'' of the TM [''extending from the end of the manubrium to one edge and having a width of 2 mm''] on a living human subject (Békésy, 1936). In this case, hearing in the contralateral ear was compared to hearing in the ear with the perforation, and Békésy concluded that the perforation ''had no noticeable effect on thresholds from 50 to 4000 cps [Hz] except to alter slightly the small deviation in the frequency function.'' As no further details are given, it is difficult to explain the clear

difference between this observation and our results.

Payne and Githler (1951) measured a reduction in cochlear potential that increased with perforation size, and they found these reductions were nearly frequency independent. McArdle and Tonndorf (1968) showed that the Payne and Githler (1951) results have a serious methodological problem. In the Payne and Githler study, the middle-ear cavity was opened to surrounding space. The open cavity caused the pressure on the middle-ear side of the TM to be smaller than it would have been with an intact middle-ear cavity, thus increasing the pressure difference across the TM and decreasing the effect of the perforation on the pressure difference across the TM.

Austin (1978) describes the frequency-dependence of loss with perforation size as: ''the presence of a perforation does not significantly affect the frequency response of the middle ear, since a flat hearing loss was observed for the three frequencies studied as well as for each size of perforation'' (Austin, 1978, p. 372). Indeed, the means of audiograms at 500, 1000, and 2000 Hz that Austin (1978) presents do not show appreciable frequency dependence. However, our measurements with perforation sizes similar to Austin's sizes show a clear frequency dependence (Fig. 8, right). One hypothesis for the differences between our measurements and Austin's audiograms is that errors occur with audiograms measured in ears with perforations (Voss *et al.*, 2000a, e). Because the perforated ear's impedance may be substantially lower than a normal ear, audiometric earphones can generate a lower-than-normal sound-pressure level when they are coupled to a perforated ear. This lower-than-expected sound-pressure level results in audiograms that make the hearing loss at the lower frequencies appear larger than they actually are. If such an error affects Austin's audiograms, the ''corrected'' audiograms would have smaller loss at low frequencies and the difference between Austin's audiometric results and our measurements would be even greater than that shown in Fig. 8 (right). Thus, the possibility that errors occurred in the ear-canal pressures generated during audiometry does not seem to account for the differences between Austin's data and ours. We have no explanation for the differences.

2. Effect of perforation size on transmission

Many studies of perforations, both animal and clinical studies, show that loss increases as perforation size increases (e.g., Anthony and Harrison, 1972; Austin, 1978; Ahmad and Ramani, 1979; Bigelow *et al.*, 1996). Our measurements show that perforation size has a big effect, and that for each perforation size, the perforation-induced transmission change can be described efficiently in three frequency regions. First, for the lowest frequencies, as perforation size increases, the loss (i.e., change from normal in transmission) increases monotonically and the slope of the loss magnitude versus frequency is about 40 dB per decade. Second, for ''middle-frequencies,'' the loss is near zero or slightly negative, and the frequency of this increased transmission increases with perforation size. The limits of the ''middle-frequencies'' range depend on and increase monotonically with perforation size. Third, for frequencies above 2000 Hz, the larger perforation sizes show the larger deviations from zero; both the gains and the losses increase with perforation size and can approach 10 dB. Thus, strictly speaking, loss increases monotonically with perforation size only in the low-frequency region.

3. Effect of perforation location on transmission

The results presented here are from perforations made in both the anterior-inferior and the posterior-inferior quadrants of the TM. We analyzed our results to determine if the perforation location plays an important role in the perforation-induced changes in stapes velocity. These comparisons are presented in detail elsewhere (Voss, 1998; Voss *et al.*, 2000c); the main conclusion is that changes in stapes velocity (i.e., transmission) do not appear to depend on perforation location in our temporal-bone preparation, with the kinds of perforation illustrated in Fig. 1.

This result contradicts the widely held clinical view that a posterior-inferior perforation results in larger hearing loss than an anterior-inferior perforation (e.g., Schuknecht, 1993b, p. 196; Glasscock and Shambaugh, 1990, p. 314; Pickles, 1987, pp. 60–61). The usual explanation for the location dependence is that a posterior perforation is closer to the round window, and as a result the pressure acting at the round window ''cancels'' the cochlear response more than the round-window pressure associated with perforations at other locations. Voss (1998, Chap. 3) used measurements of the oval and round window pressures with perforations at different locations to show that the perforation location has no effect on the pressure difference between the oval-window and the round-window pressures. Thus, our measurements reject both the view that loss depends on perforation location and the presumed theoretical basis.

Ahmad and Ramani (1979) investigate how the perforation locations of anterior-inferior versus posterior-inferior affect hearing levels. For their smaller perforations ($<$10% of the TM), they find hearing is independent of perforation location. With larger perforations, it appears that hearing levels are slightly more sensitive (i.e., lower) with anterior perforations; however, the difference between the locations is only for the lowest frequencies ($<$1000 Hz) and the differences appear small. In fact, it is not clear that there is a statistical difference between the locations, as no measures of intersubject variation are provided. Ahmad and Ramani (1979) conclude: ''It is seen that the difference in hearing losses...between antero- and postero-inferior perforations, is appreciable only at the lowest frequencies. At other frequencies it is minimal; indeed, it [effects of location on hearing losses] is almost negligible for clinical purposes.'' Thus, this study of audiograms, for a fairly large sample ($N=70$) of well-described perforations, provides, at best, weak support for the common clinical view.

B. Dominant loss mechanism: Change of pressure difference across the TM

Equation (4) expresses middle-ear sound transmission as the product of two ratios, representing the effects of (1) the pressure difference across the TM ($H_{\Delta TM}$) and (2) the cou-

pling of the pressure difference to the motion of the TM, ossicles, and cochlea (H_{TOC}). Our results show that perforation-induced changes in $H_{\Delta TM}$ are similar to perforation-induced changes in stapes velocity (Fig. 4), while perforation-induced changes in H_{TOC} are generally smaller than changes in either stapes velocity or $H_{\Delta TM}$ (i.e., Fig. 4). Thus, we conclude that (1) the dominant mechanism for changes in sound transmission is a perforation-induced change in the pressure difference across the TM, and (2) changes in the mechanical linkage of the TM and ossicles to the cochlea make secondary contributions to the total transmission changes.

C. Dependence of TM function on its structural integrity

A main result of this paper is that perforations do not alter (much) the coupling of the pressure difference across the TM to ossicular motion (i.e., H_{TOC}). Some basic questions arise concerning the processes involved. For example, if the TM's coupling of force to the manubrium occurs at the umbo and/or along the length of the manubrium (e.g., Dallos, 1973, Fig. 3.4; Wever and Lawrence, 1954, pp. 90–114), how can (1) the removal of the TM from around much of the umbo (Fig. 1), and (2) extensive slits along the margins of the manubrium (Fig. 7) have only a small effect on H_{TOC} (for frequencies above 400 Hz)? Perhaps the coupling between the peripheral region of the TM through the superior part of the manubrium is of primary importance. Perhaps fluid filling a slit couples the TM to the manubrium nearly as well as the TM itself. In this case the microstructure of the TM seems unimportant to its function. This idea is consistent with clinical practice of reconstructing TMs with a variety of materials as well as the observation that the eardrums of normal-hearing ears can have tympanosclerotic plaques or be abnormally thinned as a result of past disease (Hunter, 1993). In general, the results of this one experiment suggest that further experimental investigation is needed to determine the importance of TM structural and mechanical features for its function.

D. Clinical application of results

Our results describe hearing loss caused by different sized perforations in otherwise normal ears. These results should aid clinicians in determining whether a specific hearing loss results only from a perforation or whether other middle-ear pathology should be expected. The following features are consistent with a hearing loss from a perforation only: (1) losses that decrease as frequency increases up to about 1000 Hz, (2) losses or gains near zero in the 1000–2000 Hz range, and (3) for frequencies above 2000 Hz, losses that do not exceed about 10 dB for perforations that are less than 50% of the TM area. Some of these features may be difficult to distinguish on standard audiograms taken at only five or six octavespaced frequencies, but audiometric results in perforated ears that do not fit this pattern are suggestive of additional pathology.

The result that the dominant mechanism for transmission loss is a reduction in the pressure difference across the TM provides experimental support for the well-known clinical ''paper patch test'' (Schuknecht, 1993a; Glasscock and Shambaugh, 1990).

> ''A patch of thin paper of appropriate size is coated on one surface with unguentum and placed over the perforation. Results of hearing tests before and after application of the paper patch provide a prediction of the functional outcome of surgery'' (Schuknecht, 1993a, p. 5).

In the ''paper patch test,'' the patch returns the pressure difference across the TM to near normal levels. The patch does not return to normal any of the structural modifications made to the TM by the perforation (e.g., disruption of fibers or changes in tension). Thus, the observation that the ''paper patch test'' successfully improves hearing in many cases of perforated TMs is consistent with the results that show the pressure difference is the primary mechanism of hearing loss with perforations.

Even though numerous clinical studies have examined hearing levels with perforations, a clear picture of middle-ear function with perforations has not emerged. Instead, it is commonly observed that similar appearing perforations result in dramatically different hearing levels. For example,

> ''In general, the larger the perforation, the greater the hearing impairment, but this relationship is not constant and consistent in clinical practice; seemingly identical perforations in size and location produce different degrees of hearing loss. The reasons for the variations in the hearing effects of simple perforations are not easily defined'' (Glasscock and Shambaugh, 1990, p. 337).

Some, if not all, of this variability may result from differences in middle-ear cavity air volumes. The theoretical treatment of the results (Voss *et al.*, 2001d) demonstrates that middle-ear function with perforations depends on the middle-ear cavity volume. Although Békésy (1936) alludes to the effect of the cavity volume, the relationship between hearing levels and middle-ear cavity volume has not previously been described quantitatively. We now have a theoretical structure that describes the interaction between the perforation and the cavity volume, and it appears that normal variations in cavity volume can lead to differences in low-frequency hearing levels of up to 20 dB in ears with otherwise similar perforations. Future clinical studies can test for the importance of middle-ear cavity volume by estimating this quantity via low-frequency impedance measurements and/or CT scans.

ACKNOWLEDGMENTS

This work was supported by training and research grants from the NIDCD. We thank Diane Jones of the Oto-Pathology Laboratory at the Massachusetts Eye and Ear Infirmary for obtaining temporal bones and Christopher Shera of the Eaton Peabody Laboratory at the Massachusetts Eye

and Ear Infirmary for helpful discussions. We also thank two JASA reviewers, Dr. Robert H. Margolis and an anonymous reviewer, for helpful suggestions.

Ahmad, S. W., and Ramani, G. V. (**1979**). ''Hearing loss in perforations of the tympanic membrane,'' J. Laryngol. Otol. **93**, 1091–1098.

Allen, J. B. (**1986**). ''Measurement of eardrum acoustic impedance,'' in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer, Berlin), pp. 44–51.

Anthony, W. P., and Harrison, C. W. (**1972**). ''Tympanic membrane perforations; effect on audiograms,'' Arch. Otolaryngol. **95**, 506–510.

Austin, D. F. (**1978**). ''Sound conduction of the diseased ear,'' J. Laryngol. Otol. **92**, 367–393.

Békésy, G. v. (**1936**). ''Zur Physik des Mittelohres and über das Hören bei fehlerhaftem Trommelfell,'' Akust. Z. **1**, 13–23 (English translation, pp. 104–115 in Békésy, 1960).

Békésy, G. v. (**1960**). *Experiments in Hearing*, edited by E. G. Wever (McGraw-Hill, New York).

Bigelow, D. C., Swanson, P. B., and Saunders, J. C. (**1996**). ''The effect of tympanic membrane perforation size on umbo velocity in the rat,'' Laryngoscope **106**, 71–76.

Bright, R. A., Moore, R. M., Jeng, L. L., Sharkness, C. M., Hamburger, S. E., and Hamilton, P. M. (**1993**). ''The prevalence of tympanostomy tubes in children in the United States, 1988,'' Am. J. Public Health **83**, 1026–1028.

Dallos, P. (**1973**). *The Auditory Periphery—Biophysics and Physiology* (Academic, New York).

Egolf, D. (**1977**). ''Mathematical modeling of a probe-tube microphone,'' J. Acoust. Soc. Am. **61**, 200–205.

Glasscock, M. E., and Shambaugh, G. E. (**1990**). *Surgery of the Ear*, 4th ed. (Saunders, Philadelphia).

Huang, G. T., Rosowski, J. J., Flandermeyer, D. T., Lynch, 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.

Hughes G. B. and Nodar, R. H. (**1985**). ''Physiology of hearing,'' in *Textbook of Clinical Otology*, ed. by. G. B. Hughes (Thieme-Strutton, Inc., New York), Chap. 7, pp. 71–77.

Hunter, L. L. (**1993**). ''Auditory sequelae of recurrent and persistent otitis media with effusion in children,'' Ph.D. thesis, University of Minnesota.

Kruger, B., and Tonndorf, J. (**1977**). ''Middle ear transmission in cats with experimentally induced tympanic membrane perforations,'' J. Acoust. Soc. Am. **61**, 126–132.

Kruger, B., and Tonndorf, J. (**1978**). ''Tympanic membrane perforations in cats: Configurations of losses with and without ear canal extensions,'' J. Acoust. Soc. Am. **63**, 436–441.

Lim, D. (**1970**). ''Human tympanic membrane an ultrastructural observation,'' Acta Oto-Laryngol. **70**, 176–186.

Lynch, 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.

McArdle, F. E., and Tonndorf, J. (**1968**). ''Perforations of the tympanic membrane and their effects upon middle-ear transmission,'' Arch. Ohren. Nasen Kehlkopfheilkd. **192**, 145–162.

Mehmke, V. S. (**1962**). ''Der Hörverlust bei der Perforation des Trommelfelles,'' Z. Laryngol. Rhinol. Otol. **41**, 677–682.

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.

Molvaer, O., Vallersnes, F., and Kringlebotn, M. (**1978**). ''The size of the middle ear and the mastoid air cell,'' Acta Oto-Laryngol. **85**, 24–32.

Nishihara, S., Aritomo, H., and Goode, R. L. (**1993**). ''Effect of changes in mass on middle ear function,'' Otolaryngol.-Head Neck Surg. **109**, 899–910.

Payne, M. C., and Githler, F. J. (**1951**). ''Effects of perforations of the tympanic membrane on cochlear potentials,'' Arch. Otolaryngol. **54**, 666–674.

Pickles, J. O. (**1987**). ''Physiology of the ear,'' in *Scott-Brown's Otolaryngology: Basic Sciences*, edited by D. Wright (Butterworths, London), Vol. 1, Chap. 2, pp. 59–60.

Rabinowitz, W. M. (**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., Merchant, S. N., Ravicz, M. E., Voss, S. E., Caradonna, D., Cunningham, M. J., and Peake, W. T. (**1996**). ''Analysis of acoustic mechanisms in middle-ear pathology and reconstruction,'' in *Middle Ear Mechanics in Research and Otosurgery*, edited by K.-B. Hüttenbrink (Dresden University of Technology, Dresden), pp. 183–190.

Sadé, J. (**1982**). ''Prologue,'' in *Proceedings of the Second International Conference of Cholesteatoma and Mastoid Surgery*, edited by J. Sadé (Kugler, Amsterdam), p. 1.

Schuknecht, H. F. (**1993a**). ''Office examination of the ear,'' in *Surgery of the Ear and Temporal Bone* (Raven, New York), Chap. 1, pp. 1–8.

Schuknecht, H. F. (**1993b**). *Pathology of the Ear*, 2nd ed. (Lea & Febiger, Malvern, PA).

Shambaugh, G. (**1967**). *Surgery of the Ear*, 2nd ed. (Saunders, Philadelphia).

Shambaugh, G. E., and Glasscock, M. E. (**1980**). *Surgery of the Ear*, 3rd ed. (Saunders, Philadelphia).

Tavin, M. E., Gordon, M., and Ruben, R. J. (**1988**). ''Hearing results with the use of different tympanostomy tubes: A prospective study,'' Int. J. Pediatric Otorhinolaryngology **15**, 39–50.

Terkildsen, K. (**1976**). ''Pathologies and their effect on middle ear function,'' in *Acoustic Impedance and Admittance. The Measurement of Middle Ear Function*, edited by A. S. Feldman and M. A. Wilber (The Williams and Wilkins Company, Baltimore), pp. 78–102.

Voss, S. E. (**1998**). ''Effects of tympanic-membrane perforations on middle-ear sound transmission: measurements, mechanisms, and models,'' Ph.D. thesis, Massachusetts Institute of Technology.

Voss, S. E., Rosowski, J. J., Merchant, S. N., and Peake, W. T. (**2000b**). ''Acoustic responses of the human middle ear,'' Hear. Res. **150**, 43–69.

Voss, S. E., Rosowski, J. J., Merchant, S. N., and Peake, W. T. (**2000c**). ''How do tympanic-membrane perforations affect human middle-ear sound transmission?'' Acta Oto-Laryngologica (in press).

Voss, S. E., Rosowski, J. J., Merchant, S. N., and Peake, W. T. (**2001d**). ''Middle-ear function with tympanic-membrane perforations. II. A simple model,'' J. Acoust. Soc. Am. **110**, 1445–1452.

Voss, S. E., Rosowski, J. J., Merchant, S. N., Thornton, A. R., Shera, C. A., and Peake, W. T. (**2000e**). ''Middle ear pathology can affect the ear-canal sound pressure generated by audiologic earphones,'' Ear Hear. **21**, 265–274.

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.

Voss, S. E., Rosowski, J. J., Shera, C. A., and Peake, W. T. (**2000a**). ''Acoustic mechanisms that determine the ear-canal sound pressures generated by earphones,'' J. Acoust. Soc. Am. **107**, 1548–1565.

Wever, E. G., and Lawrence, M. (**1954**). *Physiological Acoustics* (Princeton University Press, Princeton).

Middle-ear function with tympanic-membrane perforations. II. A simple model

Susan E. Voss[a)]
Picker Engineering Program, Smith College, 51 College Lane, Northampton, Massachusetts 01063; Eaton-Peabody Laboratory of Auditory Physiology and Department of Otolaryngology, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114; Speech and Hearing Sciences Program, Harvard–M.I.T. Division of Health Sciences and Technology, Cambridge, Massachusetts 02139; Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts; and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

John J. Rosowski
Eaton-Peabody Laboratory of Auditory Physiology and Department of Otolaryngology, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114; Speech and Hearing Sciences Program, Harvard–M.I.T. Division of Health Sciences and Technology, Cambridge, Massachusetts 02139; Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts; and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Saumil N. Merchant
Eaton-Peabody Laboratory of Auditory Physiology and Department of Otolaryngology, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114; Speech and Hearing Sciences Program, Harvard–M.I.T. Division of Health Sciences and Technology, Cambridge, Massachusetts 02139; and Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts 02114

William T. Peake
Eaton-Peabody Laboratory of Auditory Physiology and Department of Otolaryngology, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114; Speech and Hearing Sciences Program, Harvard–M.I.T. Division of Health Sciences and Technology, Cambridge, Massachusetts 02139; and Department of Electrical Engineering and Computer Science and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 27 December 2000; revised 1 March 2001; accepted 24 May 2001)

A quantitative model of the human middle ear with a tympanic-membrane (TM) perforation is developed. The model is constrained by several types of acoustic measurements made on human cadaver ears, which indicate that perforation-induced changes in transmission result primarily from changes in driving pressure across the TM and that perforation-induced change in the structure of the TM and its coupling to the ossicles contributes a substantially smaller component. The model represents the effect of a perforation on the pressure difference across the TM by inclusion of a path for sound coupling through the perforation from the ear canal to the middle-ear cavity. The model implies that hearing loss with perforations depends primarily on three quantities: the perforation diameter, sound frequency, and the volume of air in the middle-ear cavity. For the conditions that produce the largest hearing loss (low frequency and large perforation), the model yields a simple dependence of loss on frequency, perforation diameter, and middle-ear cavity volume. Predictions from this model may be useful to clinicians in determining whether, in particular cases, hearing losses are explainable by the observed perforations or if additional pathology must be involved. [DOI: 10.1121/1.1394196]

PACS numbers: 43.64.Ha, 43.64.Bt, 43.70.Bk, 43.70.Gr [BLM]

I. INTRODUCTION

Measured effects of tympanic-membrane (TM) perforations on sound transmission through the middle ear have been systematically described (Voss *et al.*, 2001d). The measurements indicate that, for perforations up to half the area of the TM, the major mechanism for perforation-induced changes is an altered pressure difference across the TM. In this paper we develop a measurement-based mathematical model that describes sound transmission in ears with perforations. The model assumes that the effect of the perforation on the pressure difference across the TM can be represented by inclusion of a path for sound coupling through the perforation from the ear canal to the middle-ear cavity. The properties of this path are described by an impedance. The primary goal of this paper is to test whether this simple model represents features of the measurements.

[a)]Author to whom correspondence should be addressed. Address for correspondence: Smith College, Picker Engineering Program, 51 College Lane, Northampton, MA 01063; electronic mail: svoss@email.smith.edu

II. MIDDLE-EAR ANALOG CIRCUIT MODEL WITH TM PERFORATIONS

A. Circuit topology

The equations that govern sound pressures and volume velocities in acoustic systems are analogous to the equations that describe voltages and currents in electric circuits. Here, we use an analog-circuit model (Fig. 1) to represent the middle ear with a TM perforation. To distinguish them from measurements, model values are indicated with a circumflex above the variable. The perforation is represented by a volume-velocity path of impedance $\hat{Z}_{\text{perf}}$ from the ear-canal side of the TM to the middle-ear cavity. $\hat{Z}_{\text{perf}}$ depends on the size of the perforation; when there is no perforation, $|\hat{Z}_{\text{perf}}|$ is infinite, and the middle-ear cavity impedance $\hat{Z}_{\text{cav}}$ is in series with an impedance $\hat{Z}_{\text{TOC}}$ that represent the effects of the TM, ossicles, and cochlea (as in, e.g., Onchi, 1961; Møller, 1961; Zwislocki, 1962; Kringlebotn, 1988; Peake *et al.*, 1992; Shera and Zweig, 1992; Whittemore *et al.*, 1998).

The impedance $\hat{Z}_{\text{TOC}}$ is the input impedance of the two-port network, representing the coupling of sound through the TM and ossicles, terminated by an impedance that represents the stapes and cochlea $\hat{Z}_{\text{SC}}$. Although the two-port network, and thus $\hat{Z}_{\text{TOC}}$, could depend on the state of the TM (e.g., normal versus perforated) and the perforation characteristics (e.g., location and size), our measurements show that $\hat{Z}_{\text{TOC}}$ is not much affected by perforations (Voss *et al.*, 2001d, Fig. 4),[1] thus supporting the simplification that $\hat{Z}_{\text{TOC}}$ is independent of perforations. Measurements also support the assumption that $\hat{Z}_{\text{perf}}$ does not depend on perforation location (Voss, 1998; Voss *et al.*, 2000c).

One prediction of the circuit topology has been tested. In the circuit, the magnitude of the impedance at the TM, $|\hat{Z}_{\text{TM}}|$, has a minimum at frequencies where the magnitude of the pressure-difference ratio

$$|\hat{H}_{\Delta\text{TM}}| \equiv \left| \frac{(\hat{P}_{\text{TM}} - \hat{P}_{\text{cav}})}{\hat{P}_{\text{TM}}} \right|, \tag{1}$$

is a maximum. Specifically, the circuit requires that

$$\hat{H}_{\Delta\text{TM}} = \frac{\hat{Z}_{\text{perf}} \| \hat{Z}_{\text{TOC}}}{\hat{Z}_{\text{TM}}}. \tag{2}$$

Equation (3) predicts that local maxima in $|H_{\Delta\text{TM}}|$ will occur at the frequencies of local minima in $|Z_{\text{TM}}|$. Figure 6 of the companion paper (Voss *et al.*, 2001d) shows that our data fit this prediction over 53 perforation conditions in ten ears.

B. Representation of model components

A model was constructed for each of ten ears from measurements on that ear along with a $\hat{Z}_{\text{perf}}$ (as a function of frequency and perforation diameter) based on a theoretical model supported by independent acoustic measurements. The ten ears are the ten human-cadaver ears on which impedance measurements were made (Voss *et al.*, 2001d; Voss, 1998). Superscripts norm and perf designate the normal and perforated states of the TM, respectively.

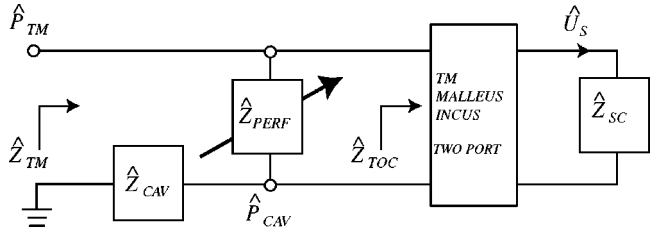


FIG. 1. Model of the human middle ear with a perforation. Voltages are analogous to sound pressures; nodes in the circuit are labeled with symbols for sound pressures, e.g., $\hat{P}_{\text{cav}}$. Currents are analogous to volume velocities; arrows on branches are labeled with volume-velocity symbols, e.g., $\hat{U}_S$. To distinguish them from measurements, model components are indicated with a circumflex to distinguish them from measured values. The TM, malleus, incus two-port network represents tympanic–membrane–ossicular coupling through the TM to the stapes and cochlea. $\hat{Z}_{\text{SC}}$ is the net impedance of the cochlea and the stapes. The input impedance $\hat{Z}_{\text{TOC}}$ of the two-port network (terminated by $\hat{Z}_{\text{SC}}$) is identical for the normal and perforated conditions. $\hat{Z}_{\text{cav}}$ is the impedance of the middle-ear cavity; $\hat{P}_{\text{TM}}$ is the sound pressure at the TM; $\hat{P}_{\text{cav}}$ is the sound pressure in the middle-ear cavity; $\hat{U}_S$ is the stapes volume velocity; and $\hat{Z}_{\text{perf}}$ is the impedance of a perforation. The arrow through $\hat{Z}_{\text{perf}}$ indicates that $\hat{Z}_{\text{perf}}$ depends on perforation size. The input impedance at the TM is $\hat{Z}_{\text{TM}}$.

In the individual-ear models used in this section, impedance measurements (Voss *et al.*, 2001d) directly determine each model middle-ear cavity impedance, i.e., $\hat{Z}_{\text{cav}} = Z_{\text{cav}}$ is taken from the measurement of cavity impedance in that ear. $\hat{Z}_{\text{TOC}}$ is obtained from measurements of Z_{cav} and $Z_{\text{TM}}^{\text{norm}}$. With a normal TM (i.e., $|\hat{Z}_{\text{perf}}| = \infty$), the model of Fig. 1 requires that

$$\hat{Z}_{\text{TOC}} = Z_{\text{TM}}^{\text{norm}} - Z_{\text{cav}}. \tag{3}$$

Therefore, for each ear's model, we determine $\hat{Z}_{\text{TOC}} \equiv Z_{\text{TM}}^{\text{norm}} - Z_{\text{cav}}$. Thus, $\hat{Z}_{\text{TM}} \equiv Z_{\text{TM}}$ for each ear.

Measurements of $Z_{\text{TM}}^{\text{norm}}$ and Z_{cav} and computations of $\hat{Z}_{\text{TOC}}$ are shown for the ten ears in Fig. 2. For most frequencies, $|Z_{\text{TM}}^{\text{norm}}| > 5 \times |Z_{\text{cav}}|$, and thus $\hat{Z}_{\text{TOC}} \approx Z_{\text{TM}}^{\text{norm}}$. This use of measurements to represent components of the model allows the analysis of the model to test the adequacy of the circuit topology for representing measurements with perforations.

To specify $\hat{Z}_{\text{perf}}$, we represent the perforation as a circular orifice in a thin plate. No theoretical derivation for the impedance of a circular orifice with dimensions that correspond to the thickness of the TM of 0.06 mm (Lim, 1995) and perforations on the order of 0.5–5 mm in diameter has been agreed upon (e.g., Sivian, 1935; Bolt *et al.*, 1949; Ingård and Labate, 1950; Kuckes and Ingård, 1953; Nolle, 1953; Thurston and Martin, 1953; Stinson and Shaw, 1985). However, Stinson and Shaw (1985) report measurements of the impedance of circular orifices in thin plates that are of comparable dimensions to our smaller perforations. Furthermore, they show that their measured impedances are in excellent agreement with expressions suggested by Thurston (1952), in which the impedance of a circular orifice is approximated by the input impedance of a cylindrical tube (''shorted'' at its other end) with a length that is extended by two ''end corrections.'' Applied to the perforations, the description is

$$\hat{Z}_{\text{perf}} = \frac{4\rho 2\pi f(t+\delta)}{\pi d^2}\phi, \tag{4}$$

where

$$\phi \equiv -j\frac{J_0(\kappa d/2)}{J_2(\kappa d/2)} \tag{5}$$

and f is frequency, $j=\sqrt{-1}$, d is the perforation diameter, ρ is the density of air, $t=0.06$ mm is the thickness of the TM (Lim, 1995), $\delta=8d/3\pi$, $\kappa^2=-j\rho 2\pi f/\mu$ where μ is the coefficient of viscosity of air ($\mu=1.82\times10^{-5}\,\text{kg s}^{-1}\,\text{m}^{-1}$), and J_0 and J_2 are zero- and second-order Bessel functions (of complex arguments), respectively.

Figure 3 shows, as a function of frequency, both the factor ϕ and the impedance $\hat{Z}_{\text{perf}}$, with perforation diameter as the parameter. At the lower frequencies, for the smaller perforations, ϕ varies substantially with frequency (in magnitude and angle) and the perforation impedance $\hat{Z}_{\text{perf}}$ has comparable resistive and mass components. As the perforation diameter increases, $|\phi|\rightarrow 1$ and $\angle\phi\rightarrow 0.25$, and the impedance becomes mass dominated. As frequency increases above 1000 Hz, for all diameters, $|\phi|\rightarrow 1$ and $\angle\phi\rightarrow 0.25$, and the impedance is mass dominated; for this range $|\hat{Z}_{\text{perf}}|$ is approximately inversely proportional to diameter.

One assumption inherent in our use of this model for $\hat{Z}_{\text{perf}}$ involves the relative velocities of the TM and the air moving through the perforation. The model is a circular orifice in a stationary baffle, whereas the TM is a membrane that vibrates. Our measurements, however, indicate that for perforations that include up to at least 25% of the TM, the TM velocity is generally at least ten times smaller than the particle velocity of air flowing through the perforation;[2] thus the model's assumption that the TM's velocity is small relative to the particle velocity of air moving through the perforation is accurate.

III. COMPARISON OF MODEL PREDICTIONS TO MEASUREMENTS

A. Organization

In general, the figures in this section compare the model to the measurements for one ear (Bone 24L) as a typical example. [Comparisons for all other ears are similar and can be found in the appendices of Voss (1998).[3]]

We start by comparing impedance at the TM with perforations (Sec. III B). Because the model is based on bone-specific measurements except for the impedance of the perforation, similarity between measurements and model

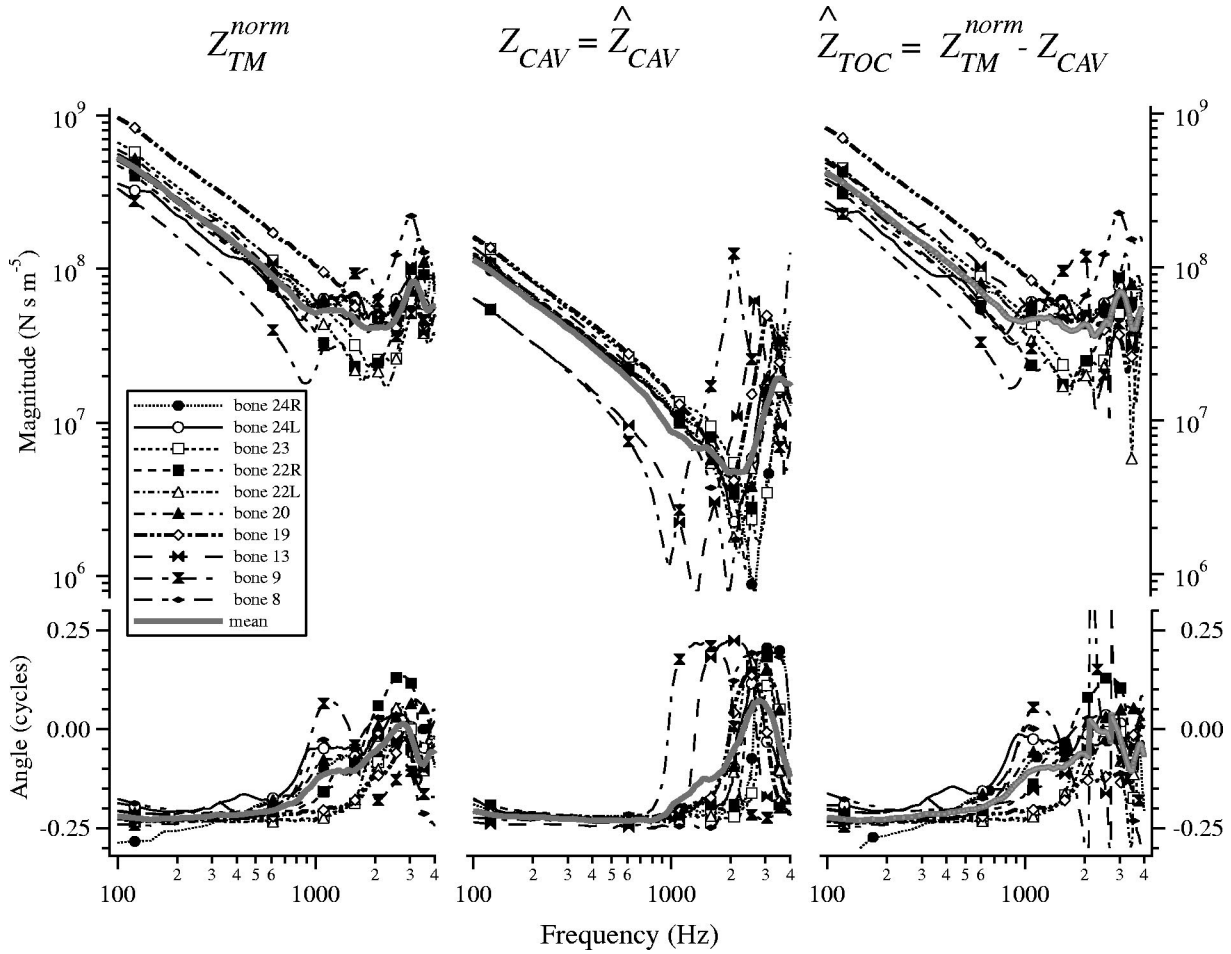


FIG. 2. Impedances from ten temporal-bone preparations as used to specify components in models for each ear. Left: Impedance measured at the normal TM, $Z_{\text{TM}}^{\text{norm}}$. Middle: Impedance measured with the TM removed Z_{cav}. Right: Impedance $\hat{Z}_{\text{TOC}}$ calculated from measurements as $\hat{Z}_{\text{TOC}}=Z_{\text{TM}}^{\text{norm}}-Z_{\text{cav}}$. Upper panel: Magnitude. The magnitude means were computed in the logarithmic domain. Lower panel: Angle. The equations at the tops of the three columns indicate how the measurements are used to determine each model's components. Symbols indicate every 20th point.

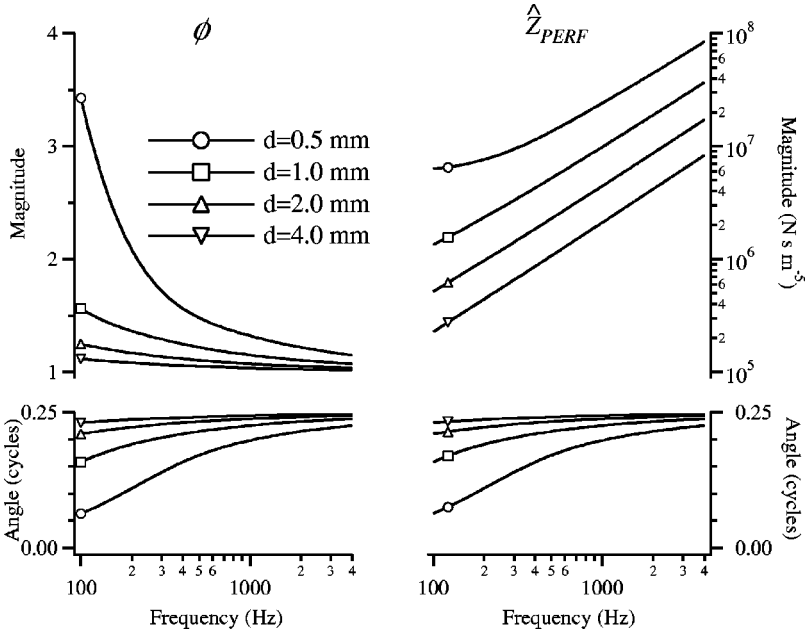


FIG. 3. Values of ϕ [Eq. (5)] and $\hat{Z}_{\text{perf}}$ [Eq. (4)]. The parameter d is perforation diameter.

indicates that (1) the inclusion of $\hat{Z}_{\text{perf}}$ in the model produces the measured effects of perforations and (2) the model for $\hat{Z}_{\text{perf}}$ is consistent with the measurements.

Next, in Sec. III C, we compare perforation-induced changes in (1) stapes velocity and (2) pressure differences across the TM. In this section, the data that are compared to the model were not used in any way to define the model parameters.

B. Impedance at the tympanic membrane

Figure 4 compares model and measurements for one ear (Bone 24L). All of the salient features of the measurements are seen in the model's behavior: (1) At the lower frequencies, the model and the measurements have similar magnitudes and decrease at 20 dB/decade with an angle that is approximately -0.20 cycles, (2) in both model and measurements the angle changes from negative to positive at a perforation-diameter-dependent frequency that varies from 700 Hz for the smallest perforation to 2000 Hz for the largest; (3) both model and measurements have a magnitude minimum whose frequency increases with increasing perforation diameter, and (4) near 3000 Hz, the model and the measurements have similar local magnitude maxima which decreases as perforation diameter increases. The similarity between the measured $Z_{\text{TM}}^{\text{perf}}$ and the model $\hat{Z}_{\text{TM}}^{\text{perf}}$ shows that the model for $\hat{Z}_{\text{perf}}$ [i.e., Eq. (4)] provides an accurate representation of the perforation's effects on input impedance for the entire frequency range and span of perforation diameters.

C. Transmission loss

With values for $\hat{Z}_{\text{TOC}}$, $\hat{Z}_{\text{perf}}$, and $\hat{Z}_{\text{cav}}$, we use the model of Fig. 1 to predict loss in sound transmission caused by perforations. The model's perforation-induced change in transmission (i.e., loss), $\Delta\hat{T}$, can be expressed as either (1) the ratio between the volume velocity through the element $\hat{Z}_{\text{SC}}$ with a normal TM and the volume velocity through the element $\hat{Z}_{\text{SC}}$ with a perforated TM or (2) the ratio between the pressure-difference transfer function [i.e., Eq. (2)] across

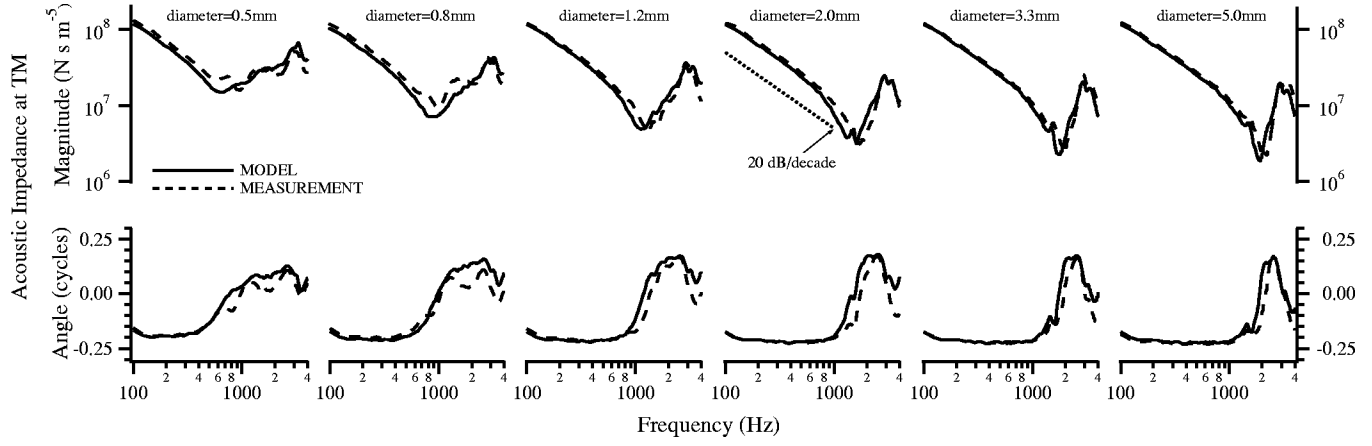


FIG. 4. Comparisons of the measured impedance at the TM $Z_{\text{TM}}^{\text{perf}}$ to the model prediction of impedance at the TM $\hat{Z}_{\text{TM}}^{\text{perf}}$ for six perforation diameters. Upper panel: Magnitudes. Lower panel: Angles. Note: The "normal" (no perforation) condition is not shown as the model is identical to the measurements by definition.

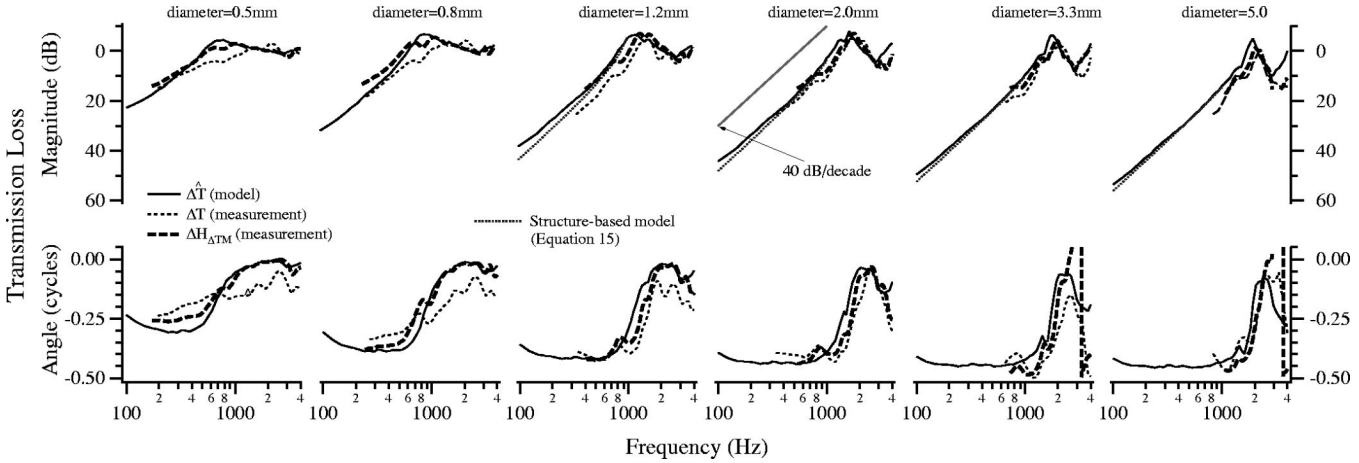


FIG. 5. Model and measurements of perforation-induced loss in transmission. The model (solid lines) is from Eq. (6) with the magnitude expressed in dB as $20\log_{10}|\Delta\hat{T}|$. Measurements from Voss *et al.* (2001d) are changes from normal in both stapes velocity $\Delta V_S/P_{\mathrm{TM}}$ (thin dashed lines) and pressure difference across the TM $\Delta H_{\Delta\mathrm{TM}}$ (thick dashed lines). Each column corresponds to the indicated perforation diameter. For perforations greater than 1 mm in diameter, the magnitude of the model simplification given by Eq. (12) is indicated by a dotted line which is hard to see in the larger perforations because it is obscured by the ''model''. Upper panel: Magnitudes. Lower panel: Angles.

the TM with a normal TM and the pressure-difference transfer function across the TM with a perforated TM: the two are identical because the only mechanism for perforation-induced change in the model is via the pressure difference across the TM,

$$\Delta\hat{T} \equiv \text{model transmission loss} \equiv \frac{\hat{U}_S^{\mathrm{norm}}}{\hat{U}_S^{\mathrm{perf}}} \equiv \frac{\hat{H}_{\Delta\mathrm{TM}}^{\mathrm{norm}}}{\hat{H}_{\Delta\mathrm{TM}}^{\mathrm{perf}}} \equiv 1 + \frac{\hat{Z}_{\mathrm{cav}} \| \hat{Z}_{\mathrm{TOC}}}{\hat{Z}_{\mathrm{perf}}}, \tag{6}$$

where the loss magnitude in dB equals $20\times\log_{10}(|\Delta\hat{T}|)$. Superimposed plots of the experimentally measured losses [i.e., perforation-induced changes in both stapes velocity and the pressure-difference transfer function (Voss *et al.*, 2001d, Fig. 3 ratio of perforated to normal)] and the model loss [$\Delta\hat{T}$, Eq. (6)] for the example ear can be compared in Fig. 5. Differences between the two measured losses are the measured loss in H_{TOC}, which is much less than the total loss for many conditions.

From Fig. 5 it is clear that the characteristics of the experimental data are present in the model: (1) At the lowest frequency, the loss magnitudes are greatest and they decrease with increasing frequency at about 40 dB/decade, (2) the magnitude of the loss at the lowest frequency increases with increasing perforation diameter, (3) small increases (negative loss magnitudes) in transmission occur around 1000 Hz, and (4) loss magnitudes in the 3000–4000 Hz region increase with perforation diameter. The lower plots (Fig. 5) show the angles of the model and measured losses. Again, many features of the measurements are seen in the model's behavior: (1) For low frequencies, the angle difference is negative (between -0.25 and -0.50 cycles) and relatively flat; thus, the perforations introduces a lag in the transmission. (2) The low-frequency angle becomes more negative as perforation diameter increases, and (3) the angle difference increases toward zero above 1000 Hz and decreases again around 4000 Hz. The model's prediction of transmission loss is not limited by the mechanical artifact, so $\Delta\hat{T}$ can be predicted for the lowest frequencies. For the larger perforations, the model predicts loss that exceeds 50 dB at the lower frequencies.

In summary, the model of Fig. 1 represents changes in sound transmission with perforations quite accurately and provides predictions beyond the range of the measurements. The results demonstrate that the model of Fig. 1 is a useful representation of sound transmission with perforations.

IV. APPLICATIONS OF THE MODEL TO PREDICTION OF HEARING LOSS

A. Goal

As discussed in the companion paper (Voss *et al.*, 2001d), a theory for middle-ear function with perforations can help physicians (1) determine whether a specific hearing loss results only from a perforation or whether other middle-ear pathology should be expected and (2) understand why similar-appearing perforations often result in different hearing losses. This section describes the model's prediction for transmission dependence on frequency, middle-ear cavity volume, and perforation diameter.

B. Dependence on middle-ear cavity volume

The impedance of the middle-ear cavity is an important part of the model for the perforated middle ear. A problem in applying this model of cadaver ears to live ears results from differences between the middle-ear cavities of cadaver and live ears. The measurements (Voss *et al.*, 2001d) were made on ears with middle-ear cavity volumes that were smaller than normal, because it is not possible to retain the entire mastoid cavity when the ear is removed from the cadaver. (The middle-ear cavity volume is the sum of the tympanic-cavity volume and the mastoid cavity volume. The tympanic-cavity volume, which is typically about 1 cm^3 and shows little inter-ear variation, is not affected by the removal process.) The volume of the middle-ear space inferred from acoustic measurements in the cadaver ears ranged from 1.5 to 3.5 cm^3, whereas volumes in normal ears range from 2 to

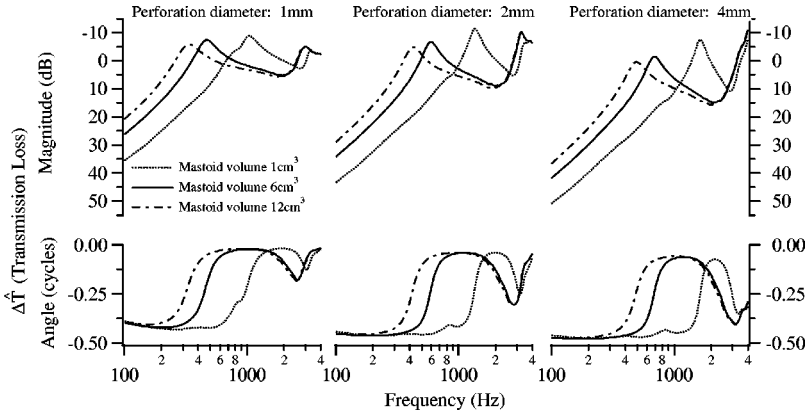


FIG. 6. Model calculations for $\Delta\hat{T}$ with variation of mastoid-cavity volume (1, 6, and 12 cm^3) and perforation diameter (1, 2, and 4 mm). The calculations use Fig. 1 with $\hat{Z}_{TOC}$ calculated by Eq. (3) from measurements on "Bone 24L", $\hat{Z}_{perf}$ from Eq. (4), and $\hat{Z}_{cav} \equiv \hat{Z}_{cav}^{SB}$ with $\hat{Z}_{cav}^{SB}$ from Voss *et al.* (2000a) with the mastoid-cavity volume having one of three indicated values. See Voss *et al.* (2000b, Fig. 12 and Table I) for details of the calculation of $\hat{Z}_{cav}^{SB}$; all element values were held constant except for the mastoid-cavity volume V_a. At frequencies above 2000 Hz, the curves for 6 and 12 cm^3 are essentially indistinguishable.

22 cm^3 (Molvaer *et al.*, 1978). To represent the effects of the smaller-than-normal middle-ear cavity volumes, we introduce a structure-based four-element acoustic model for the middle-ear cavity in which the mastoid-cavity volume is a parameter (Voss *et al.*, 2000b). $\hat{Z}_{cav}$ (in Fig. 1) is then the structure-based (SB) cavity impedance $\hat{Z}_{cav}^{SB}$ with a variable mastoid-cavity volume, i.e.,

$$\hat{Z}_{cav} \equiv \text{structure-based model} \equiv \hat{Z}_{cav}^{SB}. \tag{7}$$

Figure 6 shows how variation in mastoid-cavity volume affects the model prediction for transmission loss $\Delta\hat{T}$. For low frequencies, loss decreases as the mastoid volume increases. Variations of more than 20 dB occur in $|\Delta\hat{T}|$ when the mastoid-cavity volume is varied across the range of volumes found in normal ears (Molvaer *et al.*, 1978). Additionally, the frequency where the loss is a minimum (peak in Fig. 6) decreases as volume increases. These model results illustrate that identical perforations may not lead to similar hearing loss; the middle-ear cavity's mastoid volume plays an important role in determining hearing loss.

C. Low-frequency loss: Dependence on structure

At the lowest frequencies, Eq. (6) for $\Delta\hat{T}$ can be simplified to an analytic expression that describes how $\Delta\hat{T}$ depends on structure at low frequencies. At frequencies where $|\hat{Z}_{cav}| \ll |\hat{Z}_{TOC}|$ ($f < 1000$ Hz as in Fig. 2), Eq. (6) can be approximated as

$$\Delta\hat{T} \approx 1 + \frac{\hat{Z}_{cav}}{\hat{Z}_{perf}}. \tag{8}$$

Additionally, at the lowest frequencies, $\hat{Z}_{cav}$ is compliance dominated so that the structure-based cavity impedance $\hat{Z}_{cav}^{SB}$ can be approximated as

$$\hat{Z}_{cav}^{SB} \approx \frac{\rho c^2}{j2\pi f V}, \quad f < f_c, \tag{9}$$

where V is the total volume of the middle-ear cavity (the sum of the mastoid-cavity and tympanic-cavity volumes) and f_c is a frequency limit above which $\hat{Z}_{cav}^{SB}$ cannot be approximated by a compliance. The frequency f_c depends on V. As discussed in Voss *et al.* (2000b), the tympanic-cavity volume does not vary much from about 1 cm^3 across all ears, but the mastoid-cavity volume shows substantial variation (i.e., an order of magnitude). Because of the acoustic interaction of the two volumes (i.e., the tympanic-cavity volume and the mastoid-cavity volume), f_c decreases as V increases. For mastoid-cavity volume V_a such that $V_a < 2$ cm^3, $f_c \approx 1000$ Hz; for $V_a = 6$ cm^3, $f_c \approx 500$ Hz; and for $V_a = 12$ cm^3, $f_c \approx 250$ Hz. [See Fig. 13 of Voss *et al.* (2000b) for the specific dependence of $\hat{Z}_{cav}^{SB}$ on mastoid-cavity volume.]

For the larger perforations (i.e., $d > 1$ mm) $\hat{Z}_{perf}$ is mass dominated (Fig. 3) so that

$$\hat{Z}_{perf} \approx j2\pi f \hat{M}_{perf}, \quad d > 1 \text{ mm} \tag{10}$$

and

$$\hat{M}_{perf} \approx \frac{4\rho}{\pi d^2}\frac{8d}{3\pi}1.2, \tag{11}$$

where the term $(t+\delta)$ from Eq. (4) is approximated by $\delta = 8d/3\pi$ [since $\delta \gg t$], and the term ϕ in Eq. (4) is approximated by the constant 1.2. [The term ϕ—defined in Eq. (5)—is frequency dependent. As either perforation diameter or frequency increases, $|\phi| \to 1$ and $\angle\phi \to 0.25$ cycles (Fig. 3). We have approximated $|\phi|$ by 1.2, as 1.2 is a middle value of the total range that $|\phi|$ covers for the frequencies and perforation diameters of interest.]

Substitution of Eq. (9) and (10) into Eq. (8) leads to the approximation

$$\Delta\hat{T} \approx 1 - \frac{\alpha d}{f^2 V}, \quad f < f_c, \tag{12}$$

where $\alpha = 2.3 \times 10^6$ $cm^3/mm\,s^2$, diameter d has units mm and $d > 1$ mm, middle-ear cavity volume V has units cm^3, and frequency f has units Hz with Eq. (12) valid for frequencies where $\hat{Z}_{cav}$ is compliance dominated (i.e., $f < f_c$). The magnitude $|\Delta\hat{T}|$ predicted with the approximation of Eq. (12) is plotted in Fig. 5 for perforations with diameters greater than 1 mm, and differences between this approximation and $|\Delta\hat{T}|$ from the complete model of Eq. (6) are minimal. Thus, Eq. (12) approximates $\Delta\hat{T}$ [Eq. (6)] for frequencies below 500 Hz (with total cavity volume V of about 2 cm^3) and perforation diameters greater than 1 mm.

The accuracy of Eq. (12) in predicting the measured loss can be tested with data reported in the companion paper (Voss *et al.*, 2001d). Figure 7 shows results for a frequency of 680 Hz, which is below the $f_c = 1000$ Hz appropriate for

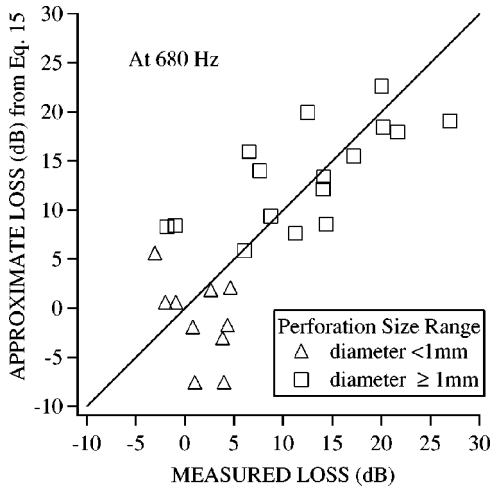


FIG. 7. A scatter plot of the measured perforation-induced transmission loss from the companion paper (Voss *et al.*, 2001d) vs the model-based approximation [Eq. (12)] at 680 Hz. The line indicates $y=x$. The plotted points correspond to all the measurements in which transmission loss at 680 Hz was accurately determined and the impedance measurements existed. The data include 26 perforations in 7 ears.

the volumes of the modified cadaver ears (range of 1.3 to 1.7 cm^3). The plot indicates that the approximate expression is consistent with the large trend in the measurements. The largest deviations from perfect prediction (the line in Fig. 7) are about 10 dB and the root-mean square error between predicted and measured is 5.9 dB. The errors for cases of small perforation are similar to those with larger perforations; an explanation for this is that for the frequency illustrated (680 Hz) the model for the perforation impedance (see Fig. 3) roughly satisfies the mass-like assumption of the approximation for all perforations larger than 0.4 mm in diameter. Figure 7 supports the use of Eq. (12) to estimate transmission loss, although it does not test the dependence of the loss on middle-ear volume, as the volumes in these ears are all within ± 0.2 cm^3 of 1.5 cm^3.

The approximation [Eq. (12)] may be useful clinically to estimate the effects of an existing perforation on hearing loss. Middle-ear cavity volume could be estimated via low-frequency impedance measurements, which will also determine the frequency range in which the approximation is expected to be accurate. For a particular ear, estimation of hearing loss caused by the perforation should allow clinicians to estimate the loss caused by other pathologies.

D. Future clinical studies of hearing with perforations

The theory presented here, coupled with the companion paper (Voss *et al.*, 2001d), provide a basis for future clinical studies of hearing with perforations. In particular, this work identifies the middle-ear cavity volume as an important variable in hearing with perforations. In future studies, the middle-ear cavity volume can be estimated via either low-frequency impedance measurements or CT scans (or a combination of the two). Additionally, future studies that rely on audiometrically measured hearing levels should include measurements of the sound pressure levels generated by the earphone in the ear canal, as recent measurements have shown that ears with lower-than-normal input impedances can reduce the ear-canal sound pressure relative to that produced in the ear canal of an ear with a normal input impedance (Voss *et al.*, 2000a,e).

V. CONCLUSIONS

A simple model for sound transmission with perforations in the tympanic membrane is tested. The perforation is represented by an impedance, which depends on frequency and perforation diameter, and is between the ear canal and the middle-ear cavity. The model's predictions for the input impedance of the middle ear and the transmission loss to the cochlea are similar to measurements of these quantities (Voss *et al.*, 2001d). Modification of the model to allow middle-ear cavity volume to be a parameter predicts that variations in middle-ear cavity volume can result in variations in hearing loss of up to 20 dB. Simplification of the model to a structure-based expression for low frequencies and perforation diameters greater than 1 mm should be useful to clinicians in relating hearing losses with perforation size, frequency, and cavity volume.

ACKNOWLEDGMENTS

This work was supported by training and research grants from the NIDCD. We thank Douglas H. Keefe of the Physical Acoustics Lab at the Boys Town National Research Hospital, Christopher Shera of the Eaton Peabody Laboratory at the Massachusetts Eye and Ear Infirmary, and Michael R. Stinson of the National Research Council (Ottawa, Canada) for helpful discussions. We also thank two JASA reviewers, Dr. Robert H. Margolis and an anonymous reviewer, for helpful suggestions.

[1]To see how Voss *et al.* (2001d, Fig. 4) support the statement that $\hat{Z}_{TOC}$ is not (much) affected by perforations, consider the transfer admittance of the two-port network of Fig. 1:

$$\hat{G}_{TOC} \equiv \frac{\hat{U}_S}{\hat{P}_{TM}-\hat{P}_{cav}} = \frac{\hat{\alpha}\hat{V}_S}{\hat{P}_{TM}-\hat{P}_{cav}} = \hat{\alpha}\hat{H}_{TOC}, \tag{13}$$

where the model's stapes volume velocity $\hat{U}_S$ is related to the model's stapes velocity $\hat{V}_S$ via $\hat{\alpha}$, which might depend on both frequency and perforation characteristics. The companion paper (Voss *et al.*, 2001d) discusses measurements of the transfer function $H_{TOC}=V_S/(P_{TM}-P_{cav})$, which in the model is $\hat{G}_{TOC}/\hat{\alpha}$. The measurements show that H_{TOC} is not a strong function of TM state. Specifically, measurements of $|H_{TOC}|$ with perforations that cover up to 25% of the TM are within 5 dB of $|H_{TOC}|$ measured with a normal TM; also, measurements show that changes from normal in $\angle H_{TOC}$ are generally less than 0.05 cycles. Measurements also support the assumption that $\hat{\alpha}$ does not depend strongly on TM state. Specifically, velocity measurements made at several locations on the stapes' suprastructure show that character of the stapes motion is not affected by perforations; the stapes primarily translates in one direction for both normal and perforated TMs up to at least 2000 Hz (e.g., Voss *et al.*, 2000b, 2001d; Voss, 1998). This experimental finding supports the assumption that $\hat{\alpha}$ does not depend on the state of the TM. In summary, our model assumption that $\hat{Z}_{TOC}$ is independent of the state of the TM is supported through measurements that show (1) H_{TOC} does not depend strongly on perforation state and (2) the mode of stapes motion is not affected by perforations.

[2]The two relevant velocities were estimated from measurements: (1) the TM velocity V_{TM} and (2) the particle velocity of air flowing through the perforation (V_{perf});

$$V_{TM}=\frac{P_{TM}}{Z_{TM}A_{TM}}, \tag{14}$$

where $A_{TM}=70$ mm^2 is the TM area (Wever and Lawrence, 1954), and

$$V_{perf}=\frac{P_{TM}-P_{cav}}{Z_{perf}A_{perf}}, \tag{15}$$

where A_{perf} is the measured cross-sectional area of the perforation. Estimates of the velocity magnitudes $|V_{TM}|$ and $|V_{perf}|$ from one of our ears are plotted in Fig. 4–5 of Voss (1998). We note that for perforation diameters of 3.3 and 5.0 mm, $|V_{TM}|$ approaches $|V_{perf}|$ for substantial frequency ranges above 800 Hz; however, $|V_{TM}|$ is generally less than $|V_{perf}|$.

[3]The model description of $\hat{Z}_{perf}$ in Voss (1998) is not identical to that used here; however, differences between the predictions from the two models are negligible.

Bolt, R. H., Labate, S., and Ingård, U. (**1949**). ''The acoustic reactance of small circular orifices,'' J. Acoust. Soc. Am. **21**, 94–97.

Ingård, U., and Labate, S. (**1950**). ''Acoustic circulation effects and the nonlinear impedance of orifices,'' J. Acoust. Soc. Am. **22**, 211–218.

Kringlebotn, M. (**1988**). ''Network model for the human middle ear,'' Scand. Audiol. **17**, 75–85.

Kuckes, A., and Ingård, U. (**1953**). ''Letter to the editor: A note on acoustic boundary dissipation due to viscosity,'' J. Acoust. Soc. Am. **25**, 798–799.

Lim, D. (**1995**). ''Structure and function of the tympanic membrane: A review,'' Acta otorhino-laryngologica Belg. **49**, 101–115.

Møller, A. R. (**1961**). ''Network model of the middle ear,'' J. Acoust. Soc. Am. **33**, 168–176.

Molvaer, O., Vallersnes, F., and Kringlebotn, M. (**1978**). ''The size of the middle ear and the mastoid air cell,'' Acta Oto-Laryngol. **85**, 24–32.

Nolle, A. (**1953**). ''Small-signal impedance of short tubes,'' J. Acoust. Soc. Am. **25**, 32–39.

Onchi, Y. (**1961**). ''Mechanism of the middle ear,'' J. Acoust. Soc. Am. **33**, 794–805.

Peake, W., Rosowski, J. J., and Lynch, T. J. (**1992**). ''Middle-ear transmission: Acoustic versus ossicular coupling in cat and human,'' Hear. Res. **57**, 245–268.

Shera, C. A., and Zweig, G. (**1992**). ''Middle-ear phenomenology: The view from the three windows,'' J. Acoust. Soc. Am. **92**, 1356–1370.

Sivian, L. (**1935**). ''Acoustic impedance of small orifices,'' J. Acoust. Soc. Am. **7**, 94–101.

Stinson, M. R., and Shaw, E. (**1985**). ''Acoustic impedance of small, circular orifices in thin plates,'' J. Acoust. Soc. Am. **77**, 2039–2042.

Thurston, G. B. (**1952**). ''Periodic fluid flow through circular tubes,'' J. Acoust. Soc. Am. **24**, 653–656.

Thurston, G. B., and Martin, Jr., C. E. (**1953**). ''Periodic fluid flow through circular orifices,'' J. Acoust. Soc. Am. **25**, 26–31.

Voss, S. E. (**1998**). ''Effects of tympanic-membrane perforations on middle-ear sound transmission: Measurements, mechanisms, and models'' Ph.D. thesis, Massachusetts Institute of Technology.

Voss, S. E., Rosowski, J. J., Merchant, S. N., and Peake, W. T. (**2000b**). ''Acoustic responses of the human middle ear,'' Hear. Res. **150**, 43–69.

Voss, S. E., Rosowski, J. J., Merchant, S. N., and Peake, W. T. (**2000c**). ''How do tympanic-membrane perforations affect human middle-ear sound transmission?'' Acta Oto-Laryngologica (in press).

Voss, S. E., Rosowski, J. J., Merchant, S. N., and Peake, W. T. (**2001d**). ''Measurements and mechanisms of middle-ear function with tympanic-membrane perforations,'' J. Acoust. Soc. Am. **110**, 1432–1444.

Voss, S. E., Rosowski, J. J., Merchant, S. N., Thornton, A. R., Shera, C. A., and Peake, W. T. (**2000e**). ''Middle ear pathology can affect the ear-canal sound pressure generated by audiologic earphones,'' Ear Hear. **21**, 265–274.

Voss, S. E., Rosowski, J. J., Shera, C. A., and Peake, W. T. (**2000a**). ''Acoustic mechanisms that determine the ear-canal sound pressures generated by earphones,'' J. Acoust. Soc. Am. **107**, 1548–1565.

Wever, E. G., and Lawrence, M. (**1954**). *Physiological Acoustics* (Princeton University Press, Princeton).

Whittemore, Jr., K. R., Merchant, S. N., and Rosowski, J. J. (**1998**). ''Acoustic mechanisms: Canal wall-up versus canal wall-down mastoidectomy,'' Otolaryngol.-Head Neck Surg. **118**, 751–761.

Zwislocki, J. (**1962**). ''Analysis of the middle-ear function. 1. Input impedance,'' J. Acoust. Soc. Am. **34**, 1514–1523.

Wideband reflectance tympanometry in chinchillas and humans[a)]

Robert H. Margolis,[b)] Saurav Paul, George L. Saly, and Patricia A. Schachern
Department of Otolaryngology, Otitis Media Research Center, University of Minnesota, Minneapolis, Minnesota 55455

Douglas H. Keefe
Boys Town National Research Hospital, Omaha, Nebraska

(Received 23 October 2000; revised 22 June 2001; accepted 25 June 2001)

Wideband reflectance tympanometry was performed on twelve chinchillas ears. The complex input impedance of the middle ear, multifrequency admittance tympanograms, reflectance patterns (reflectance versus frequency), and reflectance tympanograms (reflectance versus ear-canal air pressure) were analyzed and compared to human data. The complex impedance of the chinchilla ear has a lower stiffness reactance at low frequencies, a higher mass reactance at high frequencies, and a lower resistance compared to the human. Multifrequency admittance tympanograms from chinchillas follow the same sequence of patterns as humans for low frequencies (<2 kHz). At higher frequencies tympanograms from both species are poorly organized and do not follow a consistent sequence of patterns. Reflectance patterns of chinchillas and humans are different. However, both species show high reflectance at low frequencies, regions of lower reflectance in mid-frequencies (2–6 kHz), and high reflectance at high frequencies (>8 kHz). Reflectance tympanograms for the two species show a single, centrally located minimum at low frequencies (<2 kHz) and are substantially different at higher frequencies. Results are shown for two animals that underwent eustachian tube obstruction. Reflectance patterns obtained with different ear-canal air pressures are substantially different. Reflectance results at any single ear-canal pressure (including ambient pressure) do not completely characterize the effects of middle-ear pathology. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1394219]

PACS numbers: 43.64.Ha, 43.80.Lb, 43.80.Qf [BLM]

I. INTRODUCTION

Tympanometry, the measurement of acoustic immittance in the ear canal at various static air pressures, is widely used for the clinical assessment of middle-ear function. It is commonly performed at a single low frequency (226 Hz). The test has been shown to be sensitive to a wide variety of middle-ear pathologies. (See Margolis and Hunter, 1999 for a discussion.) By performing the measurement over a range of frequencies, multifrequency tympanometry (MFT) enhances the sensitivity of the test relative to single-frequency measurements.

Tympanometry is also useful for studying the effects of experimental ear pathology in animal subjects. Using a commercially produced instrument, we described tympanometric characteristics in normal chinchillas (Margolis *et al.*, 1995) and in chinchillas with experimentally produced ear pathologies that mimic some of the changes that occur in humans with chronic otitis media (Margolis *et al.*, 1998). With the commercial instrument that was employed, which is no longer in production, measurements were limited to frequencies at and below the resonant frequency of the chinchilla middle ear (c. 1.2 kHz). Above the resonant frequency, artifacts resulted from the instrument's inability to follow the large negative admittance phase angles that characterize the chinchilla ear. In pathologic ears, it was possible to obtain valid measurements up to 2 kHz. Above 2 kHz, the interaction between the ear canal impedance and the middle ear impedance becomes complex. In addition, the method used to calibrate these instruments, in which calibration cavities

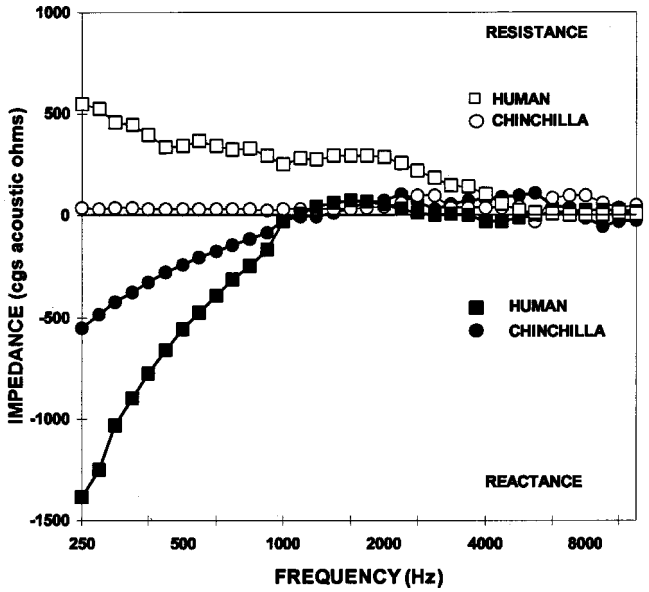


FIG. 1. Middle-ear impedance for chinchillas and humans. Human data are from Margolis *et al.* (1999).

[a)]Portions of this project were presented at the 1997, 1998, and 2001 Midwinter Meetings of the Association for Research in Otolaryngology.

[b)]Author to whom correspondence should be addressed; electronic mail: margo001@umn.edu

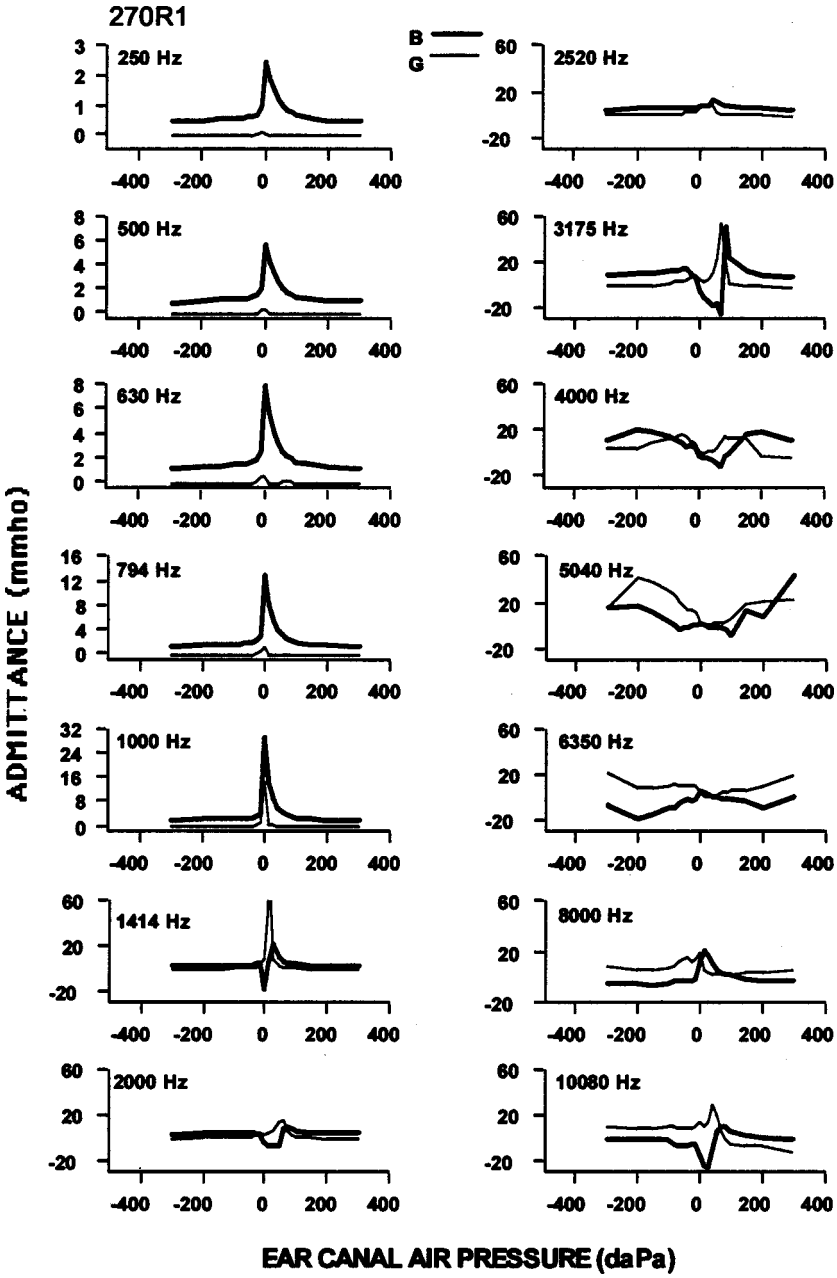


FIG. 2. Admittance tympanograms from a normal adult chinchilla. Below 2 kHz, patterns are well organized and consistent with the model of Vanhuyse *et al.* (1975). Above 2 kHz, the patterns are not predicted by the model and do not follow an orderly sequence.

are assumed to be pure acoustic compliances, is not valid above 2 kHz. For these reasons, multifrequency tympanometry with probe frequencies higher than 2 kHz has not been shown to be useful for identifying middle ear pathologies.

Experimental systems have been developed that are capable of measuring impedance over a wider frequency range. These systems are calibrated to measure the complex impedance of acoustic transmission lines (Lynch *et al.*, 1994; Huang *et al.*, 2000). By modeling the ear canal as a transmission line, the impedance at the tympanic membrane can be determined over a wider frequency range compared to clinical instruments. Using this approach we reported complex wideband impedance measurements at the tympanic membrane for normal adult humans (Margolis *et al.*, 1999). In cats, the transmission-line approximation of the ear canal yields valid estimates of the eardrum impedance up to 5 kHz (Huang *et al.*, 2000). Above that frequency, errors in ear canal parameter approximations resulted in significant errors in eardrum impedance estimates.

The energy reflectance of the ear, the ratio of the incident to reflected energy in the ear canal, can be estimated from the same ear canal sound pressure measurements and is less dependent on ear canal characteristics (Keefe, Ling, and Bulen, 1992; Keefe, Bulen, Arehart, and Burns, 1993). Keefe *et al.* (1992, 1993) reported wideband reflectance characteristics of human subjects ranging from infants to adults. The measurements were made at a single ear-canal air pressure (ambient pressure). In a subsequent paper, Keefe and Levi introduced the concept of *reflectance tympanometry*, reflectance measured as a function of ear-canal air pressure and calculated reflectance tympanograms from tympanometric recordings obtained with a commercial instrument (Keefe and Levi, 1996). We used the system developed by Keefe and colleagues (Keefe *et al.*, 1992, 1993) to record

TABLE I. Percent occurrence of each Vanhuyse pattern at each frequency.

Freq	Pattern				
	1B1G	3B1G	3B3G	5B3G	Irreg
250	*92*	0	0	0	8
500	*92*	0	0	0	8
561	*92*	0	0	0	8
630	*92*	0	0	0	8
707	*92*	0	0	0	8
794	*92*	0	0	0	8
891	*92*	0	0	0	8
1000	*92*	0	0	0	8
1120	*75*	*25*	0	0	0
1260	*17*	*75*	*8*	0	0
1410	0	*67*	*25*	*8*	0
1590	0	*25*	*50*	*25*	0
1780	0	*8*	*17*	*75*	0
2000	0	0	*8*	*83*	8
2245	0	0	0	*50*	50
2520	0	0	0	*33*	67
2828	0	0	8	*17*	75
3175	0	8	0	*8*	83
3564	0	0	0	0	100

wideband reflectance tympanograms in human ears (Margolis *et al.*, 1999). Wideband reflectance measurements have not been reported in chinchillas.

This study was undertaken to describe wideband reflectance tympanograms and wideband impedance measurements in normal chinchilla ears. The results are compared to those of a similar study of humans (Margolis *et al.*, 1999) that was performed with the same equipment and procedures. The results will provide a basis for comparison for future studies of chinchillas with experimental ear pathology.

A widely used animal model of middle-ear disease is the chinchilla ear after eustachian tube obstruction (ETO) (Canafax *et al.*, 1989). This model has relevance to human middle-ear disease because ETO is thought to be the condition that triggers otitis media pathogenesis in children and because ETO in chinchillas causes otitis media with effusion. Some examples of effects of ETO on middle ear reflectance are presented in this report.

II. METHODS

Adult chinchillas were anesthetized by intramuscular ketamine hydrochloride (HCl, 40 mg/kg) and xylazene (8 mg/kg). Measurements were made from twelve ears of six animals.

Energy reflectance was measured by a system similar to that described by Keefe (Keefe *et al.*, 1992, 1993). The measurement probe consisted of a small loudspeaker and microphone assembly coupled to the ear by commercial tips used for tympanometry. The probe was inserted into the entrance to the osseous portion of the ear canal and hand-held in that location. The stimulus was a 40 ms, digitally synthesized ''chirp'' with a bandwidth spanning the frequency range 0.25–11.3 kHz. The system was calibrated daily in a set of six rigidly terminated tubes with a diameter of 8.0 mm and varying in length from 47.5 to 95.4 cm. The calibration routine calculates the source impedance and (Thevenin) source pressure associated with the probe. The measured pressure

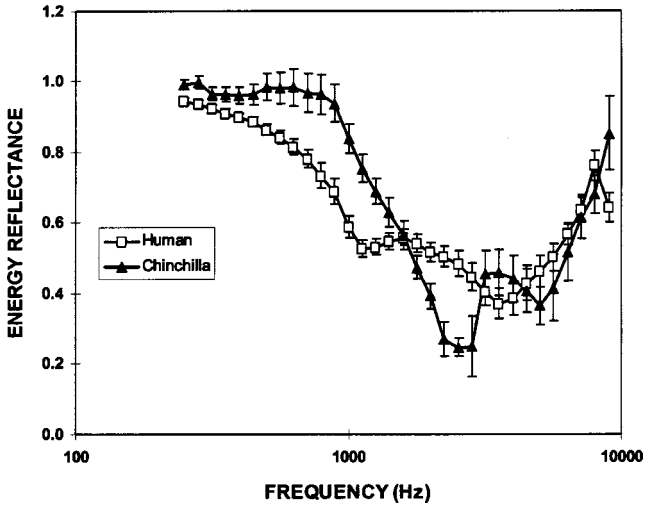


FIG. 3. Reflectance at ambient ear-canal air pressure for chinchillas and humans. Human data are from Margolis *et al.* (1999). The standard error of the mean is shown by error bars.

spectrum in the ear canal was converted into a measured impedance (or admittance) at the probe tip, and energy reflectance was calculated using standard transformations of the acoustic impedance (Keefe *et al.*, 1992). The ear-canal area is needed for this transformation, and was estimated from the size of the tip used to seal the ear canal. [For a description of the influence of ear canal area on reflectance measures, see Keefe *et al.* (1993), Appendix.] Twelve chirps were delivered into the ear canal at a rate of 12.5 chirps per second, and the acoustic pressure response to the final eight chirps was averaged. Including processing time, a wideband analysis was completed in approximately 1 s.

To vary the ear-canal air pressure, a microsyringe and digital manometer were coupled to the ear-canal probe assembly through a flexible tube. Air pressure was varied manually over the range +300 daPa to −300 daPa with the microsyringe. Measurements were made at 21 air pressures: 84 to −84 daPa in 14 daPa steps, ±100, ±150, ±200, and ±300 daPa.

To study the effects of middle-ear pathology on reflectance characteristics, the eustachian tubes were obstructed by inserting silastic sponges deeply into the nasopharyngeal orifices. One month after eustachian tube obstruction (ETO), reflectance measurements were made and the temporal bones were harvested for histologic analysis. (See Margolis *et al.*, 1998 for further details.) Results from two chinchilla ears are presented to illustrate the potential use of the technique for studying the effects of middle-ear pathology.

III. RESULTS

A. Middle-ear impedance

Middle-ear impedance was estimated by compensating the measured impedance for the ear-canal volume using the formula provided by Lynch *et al.* (1994). (See Margolis *et al.*, 1999 for further details.) Figure 1 shows the results for chinchillas and for humans obtained with identical methods (Margolis *et al.*, 1999). In both species, there is a relatively large negative reactance at low frequencies. With increasing frequency, the negative reactance approaches zero. The fre-

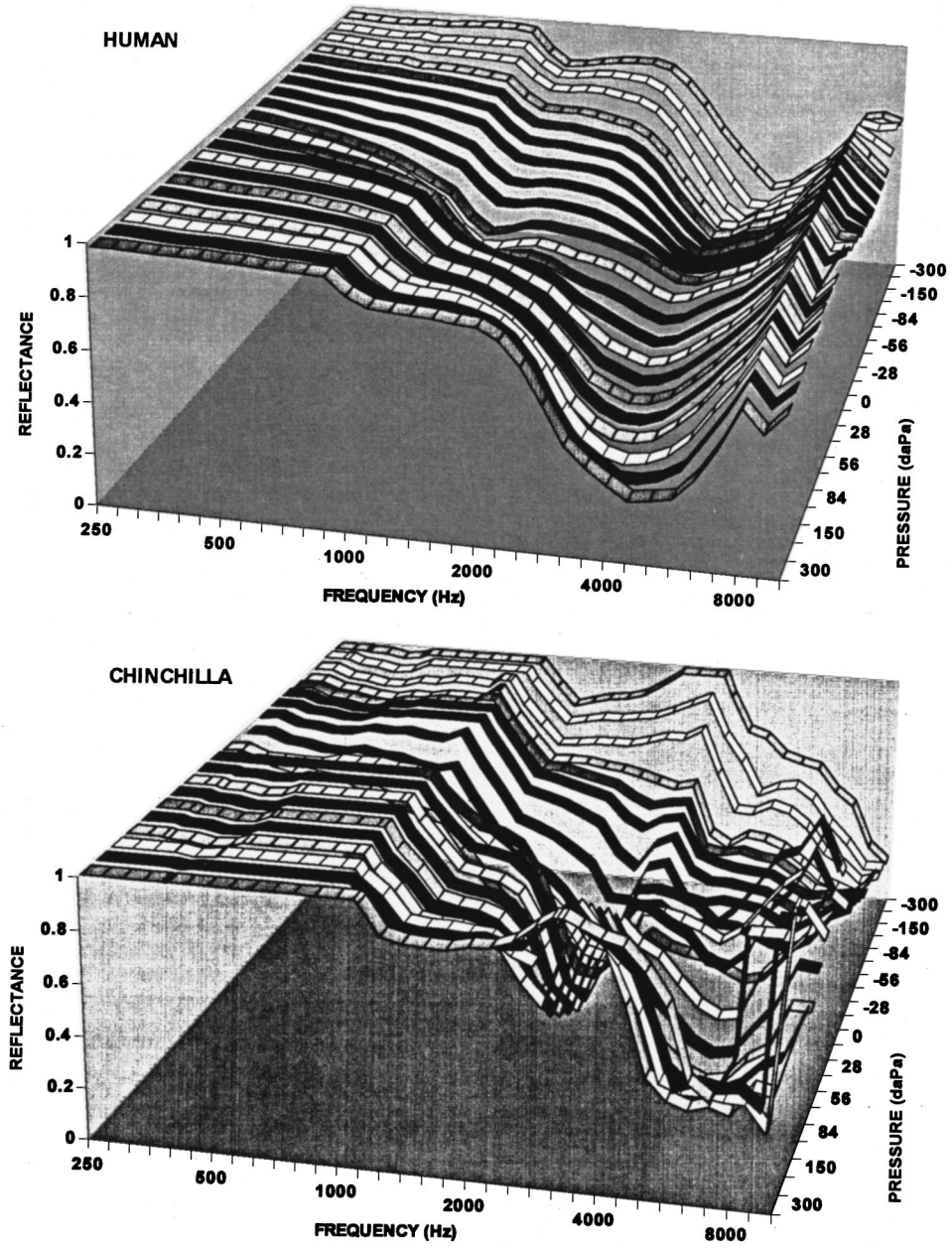


FIG. 4. Reflectance patterns for humans and chinchillas. Human data are from Margolis *et al.* (1999).

quency at which the compensated reactance is zero (phase angle $=0°$) is the resonant frequency of the middle ear. Remarkably, the resonant frequency is nearly identical in the two species (1.1 kHz). Compared to human ears, the reactance of the chinchilla ear is lower (less negative) at low frequencies (<1 kHz), higher (more positive) in the middle frequency region (2–6 kHz), and lower in the high frequencies (6–10 kHz). The resistance of the chinchilla ear is substantially lower than that of the human.

B. Multifrequency tympanograms

Figure 2 shows multifrequency tympanograms for a single chinchilla ear for frequencies 0.25–10.08 kHz. A model described by Vanhuyse *et al.* (1975) predicts a sequence of tympanometric patterns at increasing measurement frequencies and has been useful for the interpretation of tympanometric results in normal and pathologic ears. A summary of the model is presented in the appendix of a previous paper (Margolis *et al.*, 1999). The term ''irregular'' is used in this discussion to describe tympanometric patterns that are not predicted by the Vanhuyse model. The patterns in Fig. 2 follow the orderly sequence predicted by the Vanhuyse model for frequencies $\leqslant 3.175$ kHz. Above that frequency, the patterns are irregular and do not follow an orderly sequence.

Table I summarizes the proportion of Vanhuyse patterns at each frequency. The model predicts 100% occurrence of 1B1G patterns at low frequencies with an increasing occurrence of more complex patterns as frequency increases. At frequencies $\leqslant 1$ kHz, all animals showed 1B1G patterns except one which showed irregular patterns. As frequency increases above 1 kHz, the patterns shift toward the more complex patterns as the model predicts. Above 2 kHz, the patterns become predominantly irregular with 100% irregular patterns occurring for frequencies $\geqslant 3.564$ kHz.

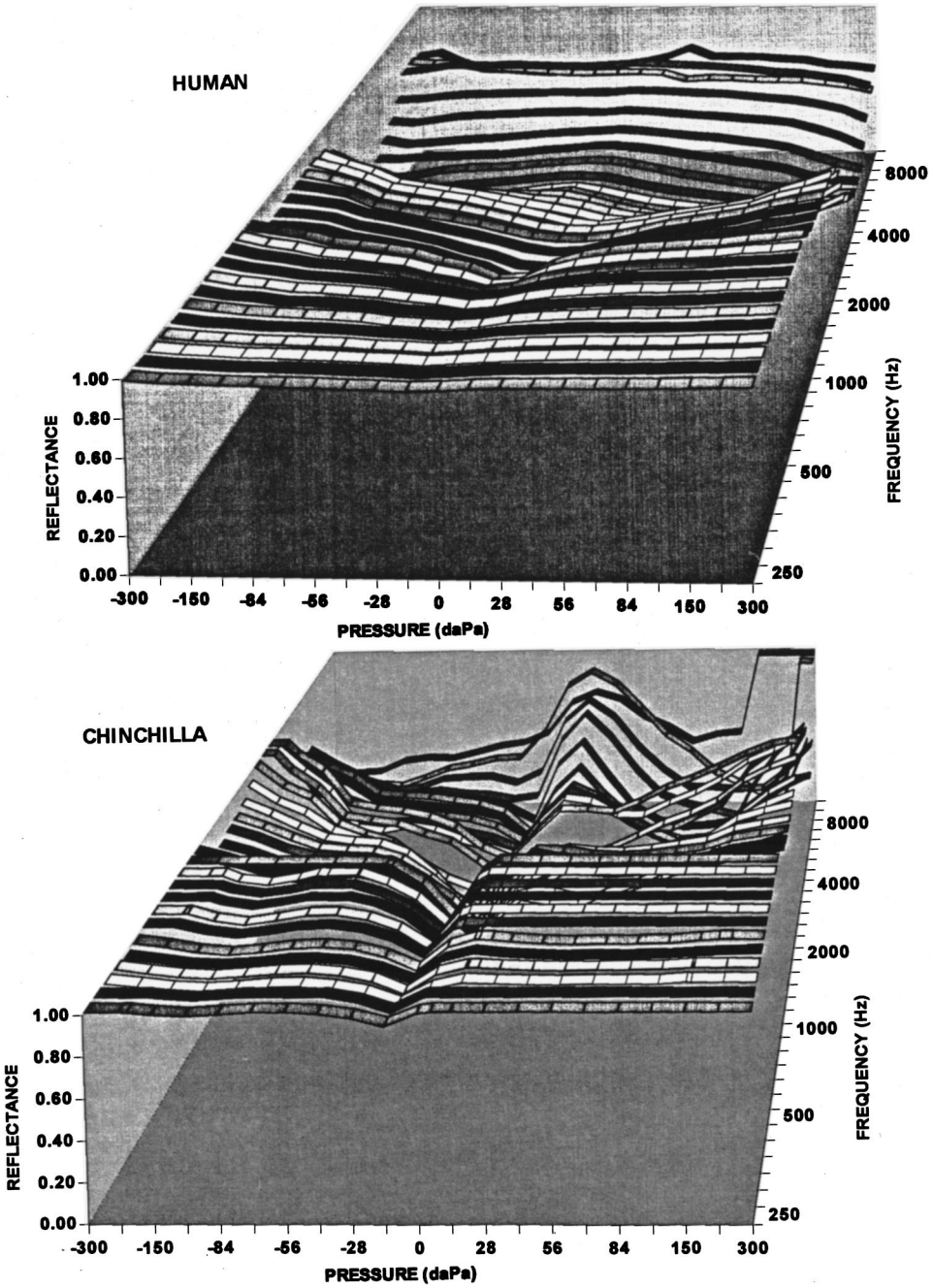


FIG. 5. Reflectance tympanograms for humans and chinchillas. Human data are from Margolis *et al.* (1999).

C. Reflectance patterns

The average reflectance at ambient ear-canal air pressure for chinchillas is shown in Fig. 3 along with similar data for humans. The human reflectance pattern, previously described by several investigators (Keefe *et al.*, 1993; Margolis *et al.*, 1999; Stinson *et al.*, 1982; Voss and Allen, 1994), is characterized by high reflectance at low frequencies that decreases to two broad minima at roughly 1 and 4 kHz, increasing to about 8 kHz and decreasing again for higher frequencies. The reflectance pattern from the chinchilla ear is a bit more complex with higher reflectance in the low frequencies and local minima at 3 and 6 kHz.

The standard error of the mean is shown as error bars in Fig. 3. The variability for the two species is similar.

Figure 4 shows average reflectance patterns for all ear-canal pressures for humans (top panel) and chinchillas (bottom panel). In humans, reflectance patterns are well organized with the largest changes occurring in the regions of the 1 and 4 kHz minima. At low frequencies there is an increase in reflectance with either positive or negative ear-canal air pressure. The minimum that occurs around 3 kHz becomes deeper and shifts upward in frequency as the ear is pressurized. In chinchillas, the patterns are not as well organized with complex pattern changes as the ear canal is pressurized. Like the human, there is an increase in reflectance at low frequencies when the ear is pressurized. At high pressures the distinct 3 kHz minimum is nearly obliterated by positive and negative pressure. These pressure effects are illustrated in more detail in Figs. 6–8, discussed later.

By rotating the plots in Fig. 4 reflectance tympanograms

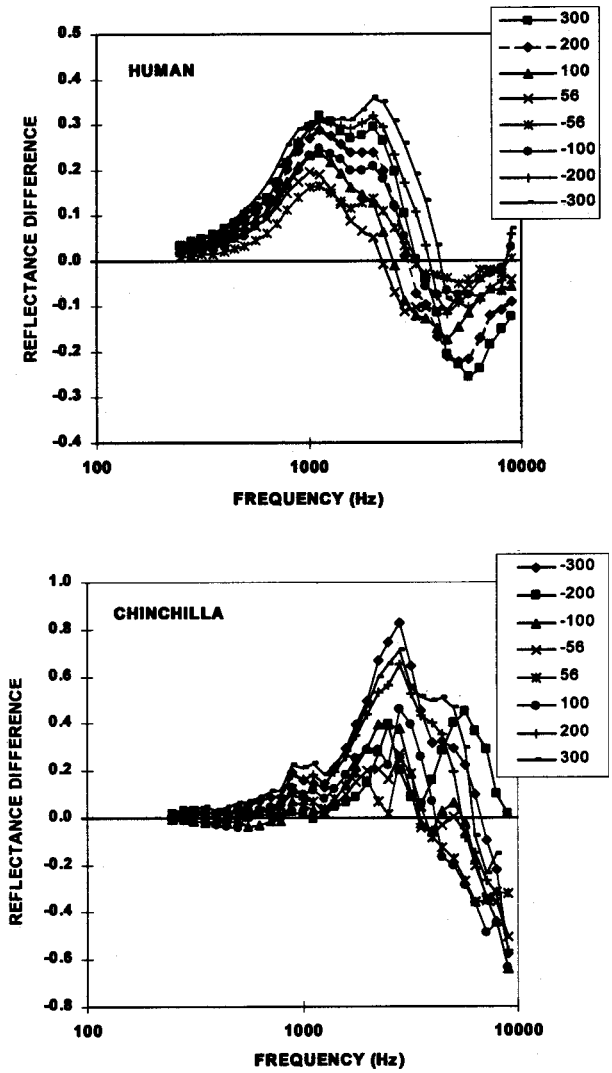


FIG. 6. Reflectance difference functions for humans and chinchillas showing the effect of pressurizing the ear canal on reflectance. Each point is the difference between reflectance at a given ear-canal pressure from the reflectance at ambient pressure. Human data are from Margolis *et al.* (1999).

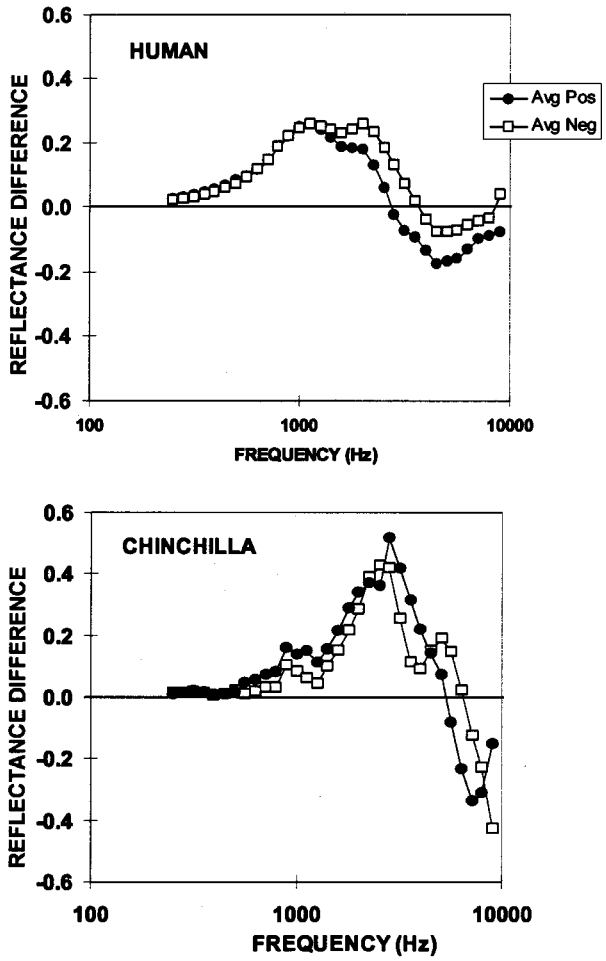


FIG. 7. Reflectance difference functions for humans and chinchillas comparing positive and negative ear-canal pressure. Human data are from Margolis *et al.* (1999).

can be examined. These are shown in Fig. 5 for humans (top panel) and chinchillas (bottom panel). Reflectance tympanograms for adult humans (from Margolis *et al.*, 1999) progress through a sequence of three patterns as frequency increases. At low frequencies they are "V" shaped; at intermediate frequencies they form an inverted "V," and at high frequencies they are flat.

Chinchilla reflectance tympanograms progress through a different series of shapes. At low frequencies ($\leqslant$1 kHz) they are "V" shaped. Above that frequency a second minimum appears giving the pattern a "W" appearance. As frequency increases the second minimum becomes deeper and the "W" is asymmetrically positioned about 0 on the pressure axis, shifted toward positive pressures.

Another way to visualize the effect of ear-canal air pressure is to examine the reflectance relative to the reflectance at ambient pressure. The difference functions shown in Fig. 6 illustrate that in humans the effect of pressurizing the ear canal is an increase in reflectance for frequencies below 4 kHz and a decrease in reflectance in the 4–8 kHz region. The largest differences occur in the 1–2 kHz region where pressurizing the ear produces large increases in reflectance. In chinchillas, the effect is more complex. Pressurizing the ear canal produces a large increase in reflectance in the 1–4 kHz region. Above 4 kHz there is either an increase or decrease in reflectance depending on the magnitude of the ear-canal pressure.

To compare the effects of positive and negative ear-canal air pressures, the reflectance was averaged for four positive and four negative pressures ($\pm$56, $\pm$100, $\pm$200, and $\pm$300 daPa) and expressed as a difference from the reflectance at ambient pressure. The comparison is shown in Fig. 7 for humans and chinchillas. In humans (top panel), reflectance was changed similarly for negative and positive pressures in the low frequencies ($<$1 kHz). For higher frequencies, reflectance was higher for negative than for positive pressure, resulting in a larger pressure-related increase in reflectance in the 1.0–3.5 kHz region and a smaller decrease in the 3.5–9.0 kHz region.

In chinchillas, reflectance differences for positive and negative ear-canal air pressure was more complex (Fig. 7, bottom panel). Below 2 kHz, the increase in reflectance was slightly greater for positive pressure. At higher frequencies there was a tendency for a greater increase in reflectance for positive pressure in the 2.8–4.5 kHz region and a greater

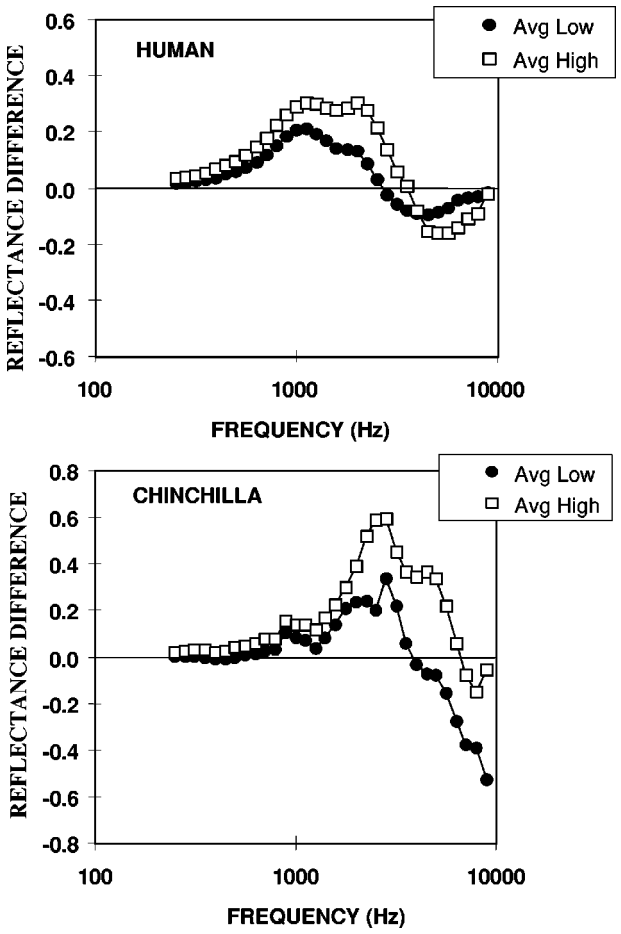


FIG. 8. Reflectance difference functions for chinchillas comparing low pressure (−100 to +100 daPa) with high pressure (±100 and ±200 daPa). Above 4 kHz, low pressure decreases reflectance and high pressure increases reflectance.

decrease for positive pressure in the 4.5–8.0 kHz region.

Figure 8 shows the effect of absolute pressure magnitude on reflectance changes for humans (top panel) and chinchillas (bottom panel). Low pressures were ±56 and ±100 daPa and high pressure were ±200 and ±300 daPa and the reflectance is expressed as a difference from ambient pressure. In humans, the effect of pressure magnitude appears linear with high pressures producing a greater increase in reflectance below 4 kHz and a greater decrease above 4 kHz. In chinchilla, the effect is nonlinear with high pressure causing a greater increase in reflectance below 4 kHz. Between 4 and 6 kHz, high pressure increases reflectance while low pressure decreases it. Above 6 kHz, there is a greater decrease in reflectance for low pressures than for high pressures.

D. Pathologic ears

After ETO, a variety of middle-ear pathologies was observed on histologic analysis. Two cases were selected to illustrate the effects of localized lesions in the middle ear. In the first case, after ETO there was a mass on the tympanic membrane. In the second case there was a localized lesion at the stapes footplate and negative middle ear pressure.

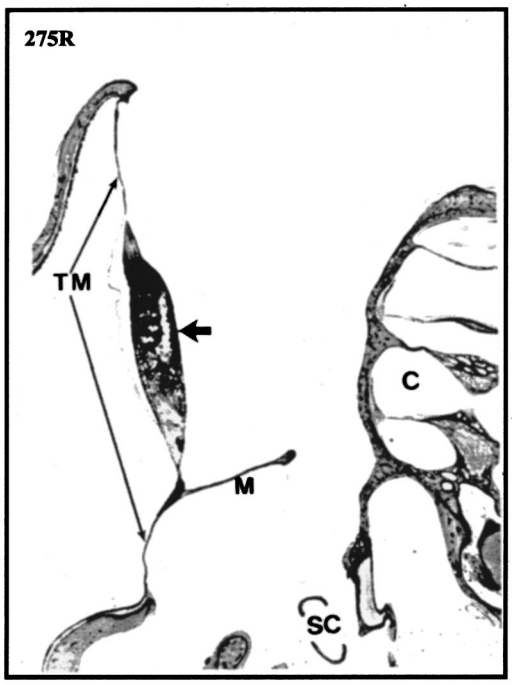


FIG. 9. Transverse section through the middle ear of a chinchilla one month after eustachian tube obstruction. There was a mass of hemorrhagic and fibrotic material on the tympanic membrane (arrow). TM=tympanic membrane; M=malleus; SC=stapes crura; C=cochlea.

1. Case 1: Tympanic membrane mass

Figure 9 shows a transverse section through the temporal bone taken one month after ETO. There was a mass of hemorrhagic and fibrotic material on the tympanic membrane (arrow). Figure 10 shows multifrequency admittance tympanograms that were obtained before the surgical procedure (left-hand panel) and one month later, immediately before the temporal bones were harvested for histologic analysis (right-hand panel). Before ETO tympanometric patterns were normal and consistent with the Vanhuyse model. One month after the procedure, patterns were abnormal at all frequencies and the patterns were not those predicted by the Vanhuyse model. The notched patterns at low frequencies (≤561 Hz) are consistent with mass loading of the middle-ear system, an effect that we observed in a previous experiment (Margolis *et al.*, 1998).

Figure 11 shows reflectance patterns obtained before (top panel) and after (bottom panel) ETO along with normative data averaged for all 12 ears. The post-ETO data were obtained with an ear-canal air pressure of 84 daPa to compensate for the positive middle-ear pressure that is evident on the tympanograms. Before ETO reflectance patterns are in the normal range (defined by the 5th–95th percentiles). One month after ETO, reflectance was abnormally reduced in the low frequencies (<2 kHz) suggesting that the eardrum pathology has an energy absorbing effect. It should be noted that a decrease in reflectance does not necessarily result in improved transmission to the cochlea. A pathologic condition can absorb energy, decreasing the reflectance without increasing sound transmission though the middle ear.

Figure 12 shows reflectance tympanograms obtained one month after ETO. The patterns were abnormal at all frequen-

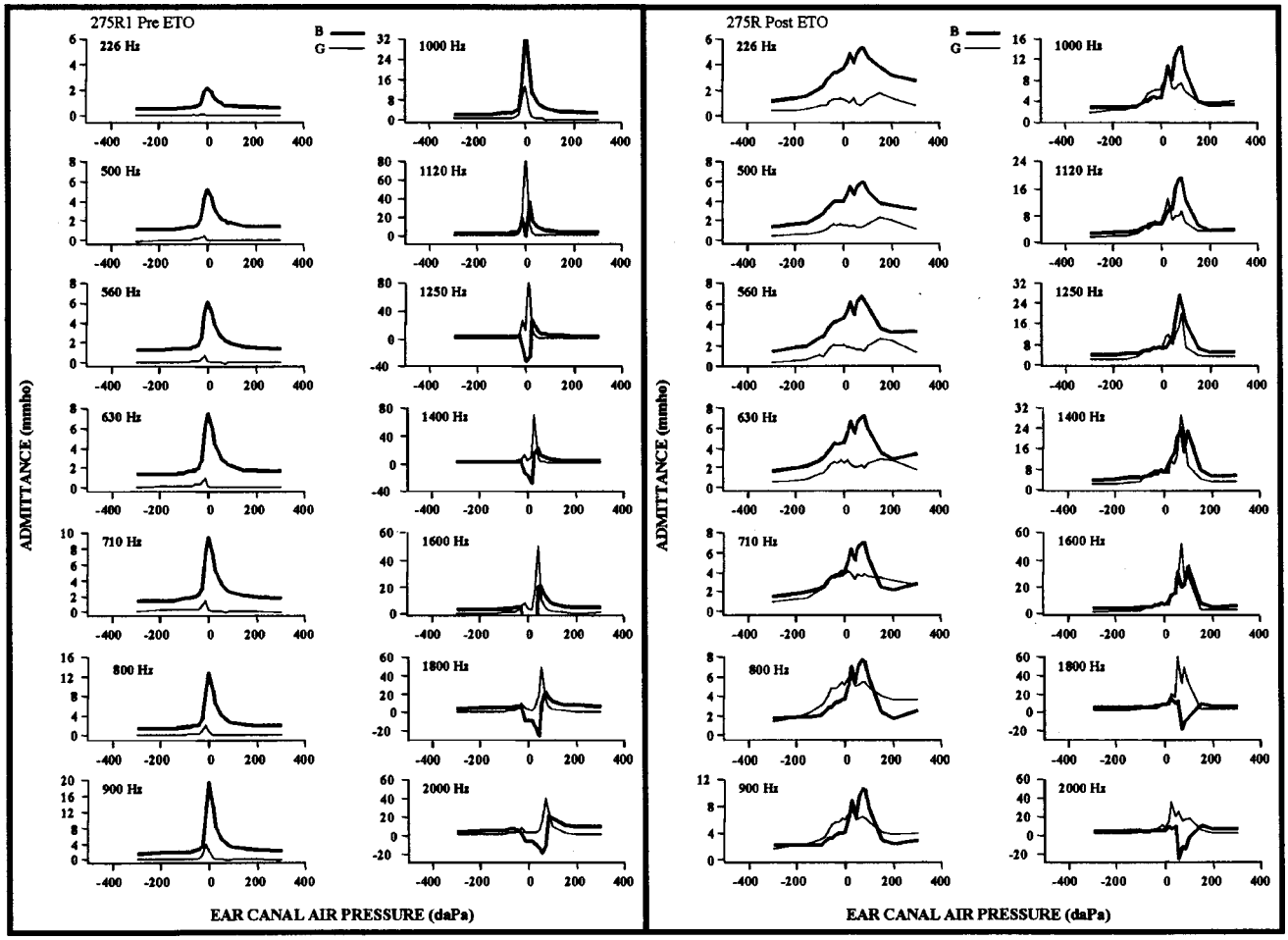


FIG. 10. Admittance tympanograms from an adult chinchilla before and one month after eustachian tube obstruction (ETO). Before ETO the patterns are normal and consistent with the Vanhuyse model. After ETO, the patterns are more complex. The notched patterns at low frequencies are consistent with mass loading of the middle ear.

cies and did not follow the orderly sequence seen in normal ears (compare to Fig. 5, bottom panel).

2. Case 2: Stapes adhesion with negative middle-ear pressure

Figure 13 shows a transverse section through the temporal bone of another animal taken one month after ETO. There was blood surrounding the stapes (arrows) with an otherwise normal middle ear.

Figure 14 shows multifrequency tympanograms obtained before and after ETO. Before ETO the patterns are normal and consistent with the Vanhuyse model. The peaks are shifted in the negative pressure direction indicating a slight negative middle-ear pressure, a normal physiologic condition. After ETO the patterns are similar to the pre-ETO data except at the high frequencies where they become irregular (non-Vanhuyse). The peak that occurs at all frequencies at 0 daPa (arrow on 1000 Hz tympanogram) is probably an artifact.

Figure 15 shows pre-ETO and post-ETO reflectance patterns for two pressure conditions, ambient ear-canal pressure [panels (a) and (c)] and at pressures that compensate for the middle-ear pressure that was estimated from the tympanograms, −100 daPa for the pre-ETO condition and −150 daPa for the post-ETO condition [panels (b) and (d)]. At ambient pressure reflectance for both the pre- and post-ETO condition was high for low frequencies (<1 kHz), with a narrow minimum just above 2 kHz, and within the normal range for higher frequencies. When the middle-ear pressure was compensated, the pre- and post-ETO patterns were very different indicating a significant effect of ETO. This suggests that an ambient pressure measurement does not always characterize the reflectance under other pressure conditions and may not reflect pathological changes.

Figure 16 shows post-ETO reflectance tympanograms. The patterns are similar to the normal patterns shown in Fig. 5 (bottom panel), except that they are shifted toward negative pressures indicating negative middle-ear pressure.

IV. DISCUSSION

A. Middle-ear impedance: Species difference

The impedance differences between humans and chinchillas are likely the result of anatomical differences, notably in the thickness of the tympanic membrane, the mass of the ossicles, and the volume of the middle ear and mastoid air spaces. The thickness of the chinchilla eardrum ranges from 18 to 30 μm in various locations of the pars tensa (Hsu *et al.*, 2000). The human eardrum thickness, which varies over a wider range than chinchillas, is 30–500 μm thick (Ruah

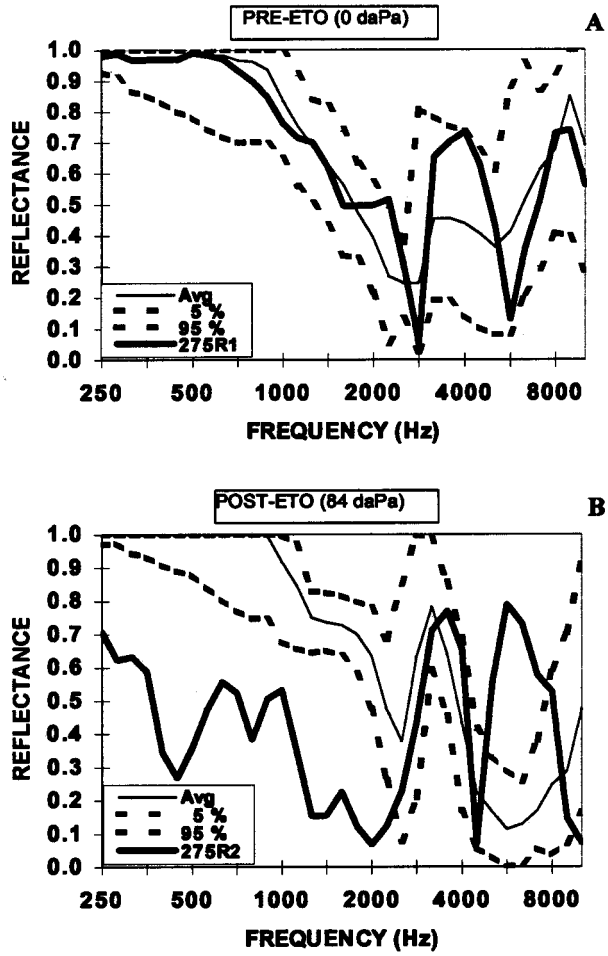


FIG. 11. Reflectance patterns for a chinchilla before (a) and after (b) eustachian tube obstruction (bold lines). Dashed lines show means and 90% ranges for normal adult chinchillas.

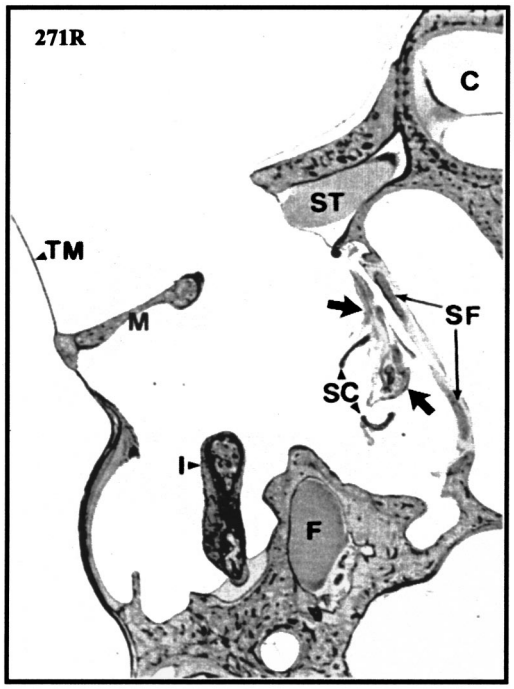


FIG. 13. Transverse section through the middle ear of a chinchilla one month after eustachian tube obstruction. There was blood surrounding the stapes (arrows). TM=tympanic membrane, M=malleus; I=incus; F=facial nerve; SC=stapes crura; SF=stapes footplate; ST=stapedius tendon; C=cochlea.

et al., 1991). The thicker human eardrum probably accounts for much of the difference in reactance, particularly at low frequencies where the stiffness of the eardrum dominates the impedance. The greater mass of the human ossicles partially offsets the greater thickness, probably accounting for the nearly identical resonant frequency. The smaller mastoid volume of the chinchilla adds stiffness to the eardrum impedance, offsetting some of the effect of the thinner eardrum.

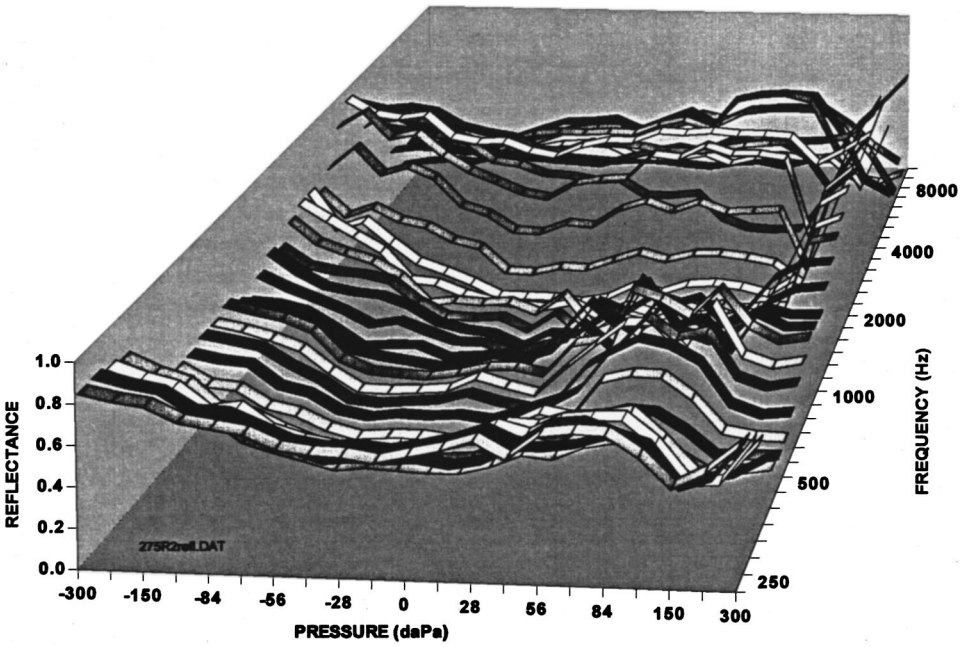


FIG. 12. Reflectance tympanograms from a chinchilla one month after eustachian tube obstruction. The patterns are abnormal (compare to Fig. 5, lower panel) and about the same at all frequencies.

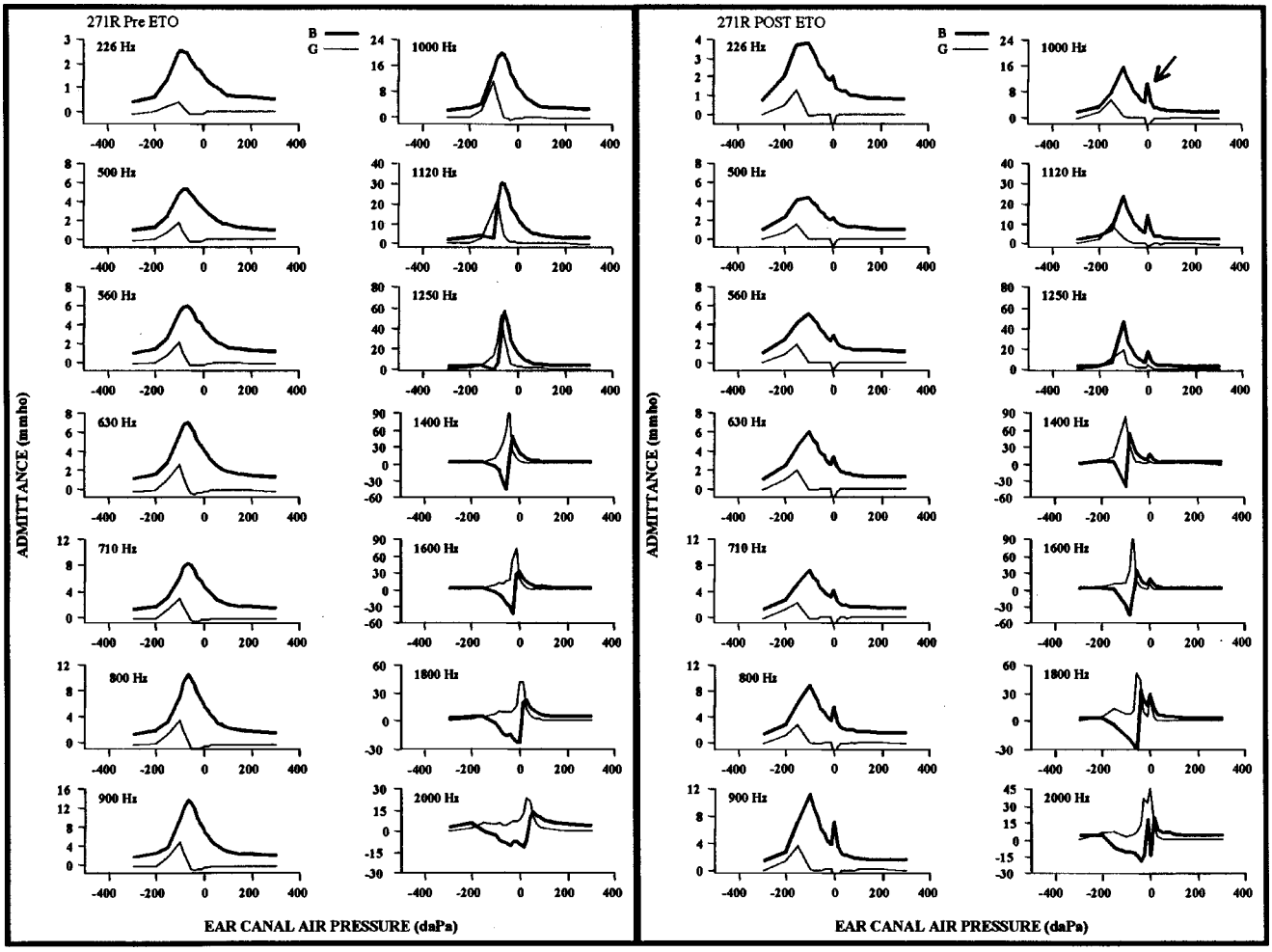


FIG. 14. Admittance tympanograms for a chinchilla before and one month after eustachian tube obstruction (ETO). Except for the artifact that occurred at about 0 daPa (arrow) post-ETO tympanograms were essentially identical to pre-ETO tympanograms.

B. Multifrequency tympanograms

The multifrequency tympanograms of normal adult chinchillas are predicted well by the Vanhuyse model for frequencies $\leqslant 2$ kHz but not for higher frequencies. At the resonant frequency and below, the patterns are similar to those reported in our earlier study that used a commercial instrument (Margolis *et al.*, 1995). The resonant frequency can be identified by the frequency at which the susceptance notch crosses the tail value or the frequency at which the input reactance to the middle ear is zero (Fig. 1). (See Margolis and Hunter, 2000 for a discussion.) For frequencies between the resonant frequency and 2 kHz the patterns obtained with the wideband reflectance system show orderly patterns, in contrast to the artifactual results obtained with the commercial system used in our previous study (Margolis *et al.*, 1995).

A comparison of Vanhuyse patterns for chinchillas and adult humans (from Margolis *et al.*, 1999) is shown in Fig. 17. These data were obtained by assigning a value of one to four to the four Vanhuyse patterns and averaging the values. Single-peaked susceptance and conductance tympanograms (1B1G) are assigned a value of one; the most complex pat-

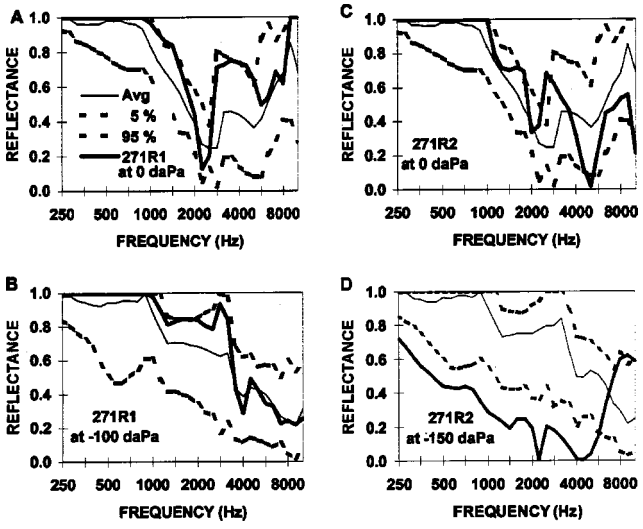


FIG. 15. Reflectance patterns for a chinchilla before and after eustachian tube obstruction (bold lines). Dashed lines show means and 90% ranges for normal adult chinchillas at the indicated ear canal pressure. (a) and (c) Reflectance at ambient ear-canal pressure is shown. (b) and (d) Reflectance patterns at the ear-canal pressure set to compensate for middle-ear pressure are shown.

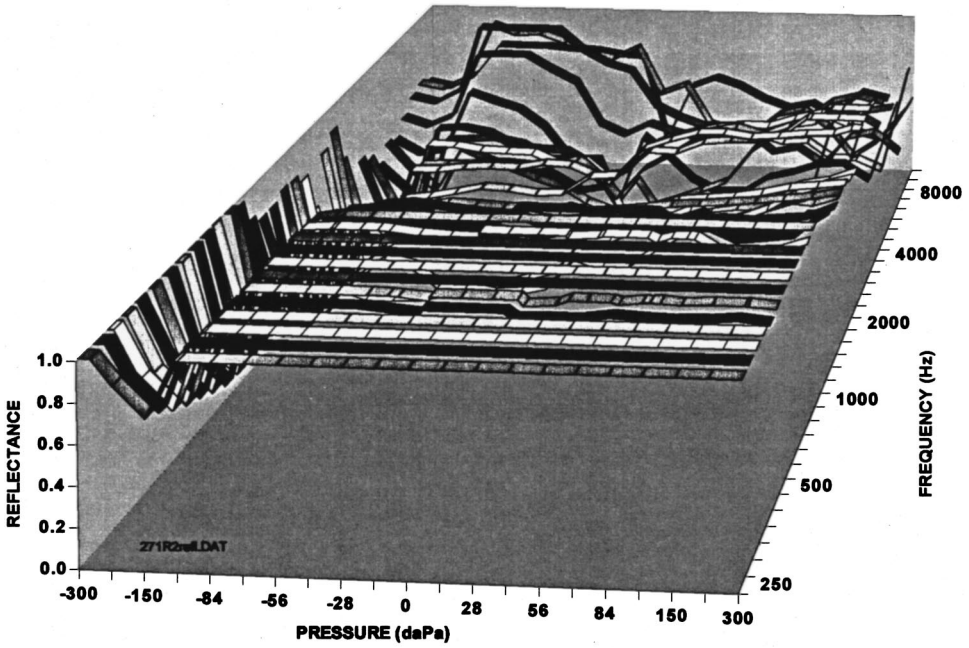


FIG. 16. Reflectance tympanograms from a chinchilla one month after eustachian tube obstruction. Except for a shift toward negative pressure the patterns are similar to those of normal chinchillas. Compare to Fig. 5, bottom panel.

tern (5B3G) is assigned a value of four. While both species progress through the patterns in the manner predicted by the model, there are differences in the form of the progression. Tympanograms from chinchillas remain single peaked (1B1G) over a wider frequency range at low frequencies and show a higher proportion of more complex patterns at higher frequencies.

Above 2 kHz, the tympanometric patterns were poorly organized and not consistent with the Vanhuyse model. Similar results were obtained from adult human ears (Margolis *et al.*, 1999). It is possible that this disorganized behavior results from the fact that the reactance and resistance are near zero in this frequency region so that small changes result in large fluctuations in conductance and susceptance. Like the human, chinchilla MFT patterns for frequencies above 2 kHz probably are not useful for characterizing normal or pathologic middle-ear conditions.

C. Reflectance patterns

Reflectance patterns from chinchillas (Figs. 3 and 4) are different from those of adult humans and not as well organized. However, like the human, they are characterized by nearly complete reflectance at low frequencies which probably limits hearing sensitivity in that region. The effect of air pressure on reflectance patterns is complex. As shown in Figs. 6–8, in both humans and chinchillas there is a frequency region in which reflectance is systematically increased by air pressure, with positive and negative pressure affecting reflectance similarly.

To the extent that air pressure can be considered a pathologic condition of the ear, the data in Fig. 6 suggest that in both humans and chinchillas there is an optimal frequency region for detecting that condition from reflectance measurements, 1–2 kHz for humans and a narrower region around 2 kHz for chinchillas. If other pathologies produce similar effects, an efficient test may be a single-frequency reflectance measurement at 2 kHz. This could be accomplished very quickly providing a test that could be performed on difficult to test patients like infants and young children. The two pathologies reported here would be effectively detected by reflectance at 2 kHz if the measurement was made at an ear-canal pressure that compensates for the nonzero middle-ear pressure [Figs. 11(b) and 15(d)].

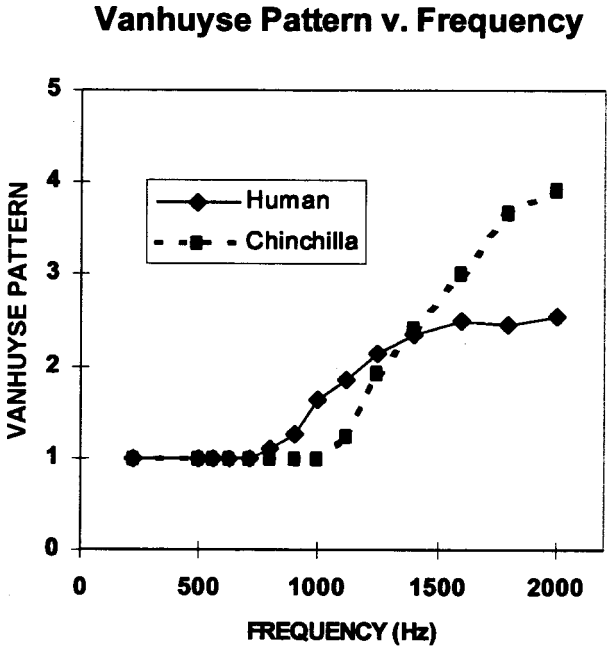


FIG. 17. "Average" Vanhuyse patterns for humans and chinchillas. The four Vanhuyse patterns were given values of 1–4 and averaged at each frequency. The most striking difference is the larger average values at frequencies above 1500 Hz indicating that there is a higher prevalence of more complex patterns at higher frequencies, suggesting a more mass dominated ear.

D. Possible clinical applications

The cases shown by Figs. 9–16 suggest some possible clinical applications for wideband reflectance/impedance measurements. There are many ways to examine the data, e.g., admittance tympanometry, reflectance patterns, reflectance tympanograms. The challenge is to determine the optimal method for distinguishing between normal and pathologic conditions. A test that does not require manipulation of the ear-canal air pressure would provide significant advantages over currently used tympanometric methods. The robust effect of ear-canal air pressure in the 1–2 kHz region of both humans and chinchillas (Fig. 6) suggests an interesting possibility that a single-frequency reflectance measurement may be sensitive to middle-ear pathology. On the other hand, the large differences in reflectance patterns shown in Fig. 15 [(b) and (d)] suggest that measurements at any single ear-canal air pressure do not completely characterize the reflectance characteristics of an ear. Impedance measurements did not become a clinically useful tool until the ear-canal air pressure dimension was added—tympanometry. It remains to be seen if wideband reflectance measurements made at ambient ear-canal pressure can improve on the efficiency and accuracy of our clinical tests.

V. SUMMARY AND CONCLUSIONS

Wideband reflectance measurements are described for normal adult chinchillas and compared to similar data from humans. Species differences include the following. Compared to the adult human

(1) the chinchilla ear impedance has a lower stiffness reactance at low frequencies, a higher mass reactance at high frequencies, and a lower acoustic resistance;
(2) admittance tympanograms from chinchillas are similarly well organized at low frequencies (<2 kHz) and become progressively disorganized at higher frequencies;
(3) reflectance patterns (reflectance versus frequency) of chinchilla ears show similarly high reflectance at low frequencies and are somewhat less organized at higher frequencies;
(4) reflectance tympanograms of chinchillas progress through a different sequence of patterns as frequency is increased;
(5) the chinchilla reflectance shows a similar robust effect of ear-canal pressure at 2 kHz where there is a large increase in reflectance;
(6) there is a similar complex effect of middle-ear pathology on reflectance patterns obtained at different ear-canal pressures, suggesting that measurements at a single pressure do not completely characterize the reflectance response of an ear.

ACKNOWLEDGMENTS

This work was supported by Grant No. P50-DC03093 from the National Institutes of Health. We are grateful to Grason-Stadler, Inc. for their assistance in the design and construction of the probe used for the impedance/reflectance system. Sherry Fulton provided the histopathologic processing of the temporal bones and Moses Quartey performed the eustachian tube obstruction surgery. Their expertise contributed greatly to this study.

Canafax, D. M., Nonomura, N., Erdman, G., Le, C. T., Juhn, S. K., and Giebink, G. S. (**1989**). ''Experimental animal models for studying antimicrobial pharmokinetics in otitis media,'' Pharm. Res. **6**, 279–285.

Hsu, G. S., Margolis, R. H., and Schachern, P. A. (**2000**). ''The development of the middle ear in neonatal chinchillas. I. Birth to 14 days,'' Acta Oto-Laryngol. **120**, 922–932.

Huang, G. T., Rosowski, J. J., Puria, S., and Peake, W. T. (**2000**). ''A noninvasive method for estimating acoustic admittance at the tympanic membrane,'' J. Acoust. Soc. Am. **108**, 1128–1146.

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 reflectance coefficient,'' J. Acoust. Soc. Am. **91**, 470–485.

Lynch, 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 Hunter, L. L. (**1999**). ''Tympanometry—Basic principles and clinical applications,'' in *Contemporary Perspectives on Hearing Assessment*, edited by F. E. Musiek and W. F. Rintelmann (Allyn and Bacon, Boston), pp. 89–130.

Margolis, R. H., and Hunter, L. L. (**2000**). ''Acoustic immittance measurements for audiologic diagnosis,'' in *Audiology: Diagnosis*, edited by R. J. Roeser, M. Valente, and H. Hosford-Dunn (Thieme, New York).

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., Schachern, P. A., and Fulton, S. (**1998**). ''Multifrequency tympanometry and histopathology in chinchillas with experimentally-produced middle-ear pathologies,'' Acta Oto-Laryngol. **118**, 216–225.

Margolis, R. H., Schachern, P. L., Hunter, L. L., and Sutherland, C. (**1995**). ''Multifrequency tympanometry in chinchillas,'' Audiology **34**, 232–247.

Ruah, C., Schachern, P., Zelterman, D., Paparella, M., and Yoon, T. (**1991**). ''Age-related morphologic changes in the human tympanic membrane. A light and electron microscopic study,'' Arch. Otolaryngol. Head Neck Surg. **117**, 627–634.

Stinson, M. R., Shaw, E. A. G., and Lawton, W. B. (**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.

Maturation of the human cochlear amplifier: Distortion product otoacoustic emission suppression tuning curves recorded at low and high primary tone levels

Carolina Abdala[a)]
House Ear Institute, Children's Auditory Research and Evaluation Center, 2100 West Third Street, Los Angeles, California 90057

(Received 20 December 2000; revised 12 February 2001; accepted 18 May 2001)

The cochlear amplifier shows level-dependent function and works optimally at low levels. For this reason, manipulation of stimulus level is a route through which the human cochlear amplifier can be investigated in a noninvasive manner. Distortion product otoacoustic emissions (DPOAEs) evoked as a function of stimulus level provide a tool for exploration of human cochlear amplifier function and, when applied to neonates, for investigation of cochlear maturation. The current experiment generated $2f1-f2$ DPOAE ipsilateral suppression tuning curves (STCs) at three primary tone levels and five $f2$ frequencies in a large group of premature and term neonates and adults. The differences between tuning generated with low- and high-level primary tones was measured to provide a gross estimate of the ''tuning enhancement effect'' attributed to the cochlear amplifier. Other features of the DPOAE suppression tuning curves were measured as well. Consistent with previous reports, at 1500 and 6000 Hz, STCs were narrower, with a steeper slope on the low-frequency flank of the tuning curve in premature neonates versus adults. Additionally, only DPOAE STCs from adults and term neonates became markedly broader and more shallow when recorded with high-level primary tones. It has been hypothesized that the excessive narrowness of suppression tuning and the absence of a level effect on DPOAE STCs recorded in premature neonates reflects a subtle immaturity in cochlear amplifier function just prior to term birth. [DOI: 10.1121/1.1388018]

PACS numbers: 43.64.Jb, 43.64.Kc [BLM]

I. INTRODUCTION

In the last 30 years our understanding of cochlear function has changed rather drastically. It is now accepted that the cochlea does not work in a passive manner, as a simple receptor of acoustic energy. Rather, the cochlea is thought to include an active component that mechanically augments motion of the basilar membrane around the region of maximum displacement. This activity is the motility of the outer hair cells. The physical augmentation of membrane motion enhances frequency resolution and auditory sensitivity (Dallos, 1992; Neely and Kim, 1986). It is not clear when the human cochlea becomes fully active and mature in its ability to enhance audition. This issue of human cochlear maturation has most recently been studied using otoacoustic emissions (Abdala *et al.*, 1996; Abdala, 1998; Abdala, 2000; Brown *et al.*, 1995; Popelka *et al.*, 1995; Lasky, 1998). Otoacoustic emissions are considered to be by-products of the cochlear amplification process and thus appear to reflect outer hair cell and cochlear integrity in a noninvasive manner.

One way to study cochlear amplifier function and maturation of cochlear amplifier function in humans is to manipulate stimulus level while evoking otoacoustic emissions (OAEs). The ''cochlear amplifier'' saturates at moderate-to-high levels and ceases to enhance sensitivity and frequency resolution. It is most effective at low levels (Dallos, 1992). Therefore by generating OAEs at various stimulus levels, and observing changes in the response as a function of level, we may be able to define the boundaries and characteristics of the cochlear amplifier. That is, we may be able to estimate the ''gain'' in sensitivity and the magnitude of ''tuning enhancement'' provided by the amplifier.

The distortion product OAE (DPOAE) growth function has provided an effective index of cochlear amplifier gain in laboratory animals. The growth function is a graph of DPOAE amplitude as a function of primary tone level and has been recorded in several species including rodents and rabbit (Brown *et al.*, 1989; Mills and Rubel, 1996; Whitehead, 1992a, b). In small mammals, the DPOAEs generated with low-level stimuli reflect cochlear amplifier function; when the cochlear amplifier is disabled by ototoxins such as loop diuretics or aminoglycocides, these low-level DPOAEs are eliminated. In contrast, DPOAEs generated with high-level stimuli remain largely unchanged when ototoxins are introduced (Whitehead, 1992a, b; Mills and Rubel, 1996; Brown *et al.*, 1989). Thus distortion generated with low-level primary tones is thought to reflect active cochlear motion and distortion generated with high-level primary tones is thought to reflect primarily passive cochlear motion. A plateau, or a deep notch in the DPOAE growth function typically separates these two modalities of distortion and represents the general range of levels at which the cochlear amplifier saturates and the cochlea transitions from active to primarily passive function (Whitehead *et al.*, 1992a, b).

It is less clear how stimulus level influences the cochlear

a)Electronic mail: Cabdala@hei.org

amplifier and DPOAEs in humans. A recent paper from our laboratory on DPOAE growth functions in human adults and neonates suggests that stimulus level can also be used to define saturation characteristics of the cochlear amplifier (Abdala, 2000). This study showed that the human adult cochlea generates distortion that generally saturates around 70–73 dB sound pressure level (SPL). Using animal results as a model, this finding suggests that active processes do not enhance tuning or sensitivity beyond this stimulus level. In premature neonates, the saturation plateau occurred at elevated levels for high-frequency primary tones, suggesting that the effective range of cochlear amplifier function differs in adults and neonates.

In the current experiment, primary tone level was also manipulated to study maturation of cochlear amplifier function; however, DPOAE suppression tuning curves (rather than DPOAE growth functions) were recorded at three primary tone levels from adults and neonates at various maturational stages to study this question. The DPOAE suppression tuning curve (STC) has been generated and measured in laboratory animals and humans and appears to reflect features of cochlear tuning and traveling wave motion (Abdala *et al.*, 1996; Frank and Kossl, 1995; Mills, 1999). The basic premise of the present study is that a DPOAE STC generated with low-level primary tones reflects cochlear mechanical function that has been enhanced by cochlear amplifier influence. In contrast, a DPOAE STC generated with high-level primary tones will reflect primarily passive motion of the basilar membrane due to a saturated cochlear amplifier. The difference in suppression tuning between these two level conditions should grossly estimate the strength of the sharpening or ''tuning enhancement'' provided by the cochlear amplifier. By studying this estimate of tuning enhancement during development, it may be possible to make statements about maturation of active cochlear processes in humans.

The purpose of this experiment was to determine: (1) the extent of tuning enhancement provided by the cochlear amplifier and, (2) the developmental stage at which the cochlear amplifier becomes adultlike as an effective mechanism for tuning enhancement as measured by DPOAE suppression.

II. METHODS

A. Subjects

Thirty normal-hearing adults with a mean age of 25 years served as subjects for this study (11 male; 19 female). Twelve right ears and 18 left ears were tested. Adult subjects had an audiometric air-conduction threshold of <15 dB HL from 500 to 8000 Hz and a negative history of otologic disease. They were recruited from House Ear Institute employee pool and were tested in a sound-treated booth in the Children's Auditory Research and Evaluation Center while sitting quietly in a comfortable arm chair.

The neonatal subjects were born at Women and Children's Hospital, Los Angeles County+University of Southern California Medical Center and tested prior to discharge in the Infant Auditory Research Laboratory, a quiet room away from the neonatal intensive care unit (NICU) and postpartum rooms. Three-hundred neonates (202 premature and 98 term-born) served as subjects for this study. 142 of the neonates were female and 158 were male. 151 right ears and 149 left ears were tested. The premature subjects comprised three groups categorized by conceptional age at test (CA): P1=31–33 weeks; P2=34–36 weeks; and P3=37–41 weeks. Conceptional age is defined as gestational age at birth plus weeks between birth and test. Average birthweight for premature neonates was 2073 g and average 5 and 10 min APGAR scores (indicator of neonatal health at birth) were 7.3 and 8.6, respectively (10 reflects optimal conditions).

Term neonates were born between 37 and 41 weeks of gestation and tested on average within 84 h of birth. Average birthweight for term infants was 3344 g and 5 and 10 min APGARS were 7.94 and 8.93, respectively. None of the neonates included in this study had high-risk factors for hearing loss and all of them passed a hearing screening with a click-evoked ABR recorded at 30 dB HL.

B. Instrumentation and signal analysis

An Ariel DSP16+ signal processing and acquisition board housed within a Compaq Prolinea 590 personal computer with a Pentium processor was used to generate stimuli and acquire data. The Ariel board was connected to an Etymotic Research ER-10C probe system and to an analog high-pass filter (12 dB/oct; 710 Hz high-pass cutoff). The ER-10C probe contains two output transducers and a low-noise microphone. The two primary tones and the suppressor tone were generated by the DSP processor. The primary tone at $f1$ was generated by one digital-to-analog converter (DAC) and delivered via one transducer. The primary tone at $f2$ and the suppressor tone (fs) were produced by the second DAC and output through the second transducer.

The signal at the probe microphone was high-pass filtered and sampled at a rate of 50 kHz with a sweep length of 4096 samples, giving a frequency resolution of 12.2 Hz. Twenty-five sweeps of the microphone signal were added and comprised one block for $f2=3000$, 6000, 9000, and 12 500 Hz. Due to elevated noise in the low-frequencies, 50 sweeps were added to make up one block at 1500 Hz.

C. Data acceptance criteria

Acceptance criteria were as follows: (1) Noise measurements for three frequency bins (12.2 Hz wide) on either side of the $2f1-f2$ frequency had to be <0 dB SPL to assure appropriate subject state and (2) the measured DPOAE level must be at least 5 dB above the average noise measured in the same six bins around the distortion product frequency to be accepted into the grand average.

The program attempted up to six blocks of either 25 or 50 sweeps to achieve the absolute noise criteria of 0 dB SPL and the signal-to-noise ratio of 5 dB. If both of these criteria were not met after six attempted blocks, no data was collected and the next condition was initiated. In addition, sweeps were accepted into a block of data only when the estimated rms level in that sweep did not exceed a user-controlled artifact rejection threshold. This level was set for

each subject based on observations of baseline activity level determined early in the test session, and modified if necessary during the experiment.

Intermodulation distortion produced by the recording system at $2f1-f2$ was measured with the probe in a Zwislocki coupler for all test conditions. The mean level of distortion was -21 dB SPL. In no case did the level exceed -17 dB SPL. The recording system noise floor was determined using a similar method with no tones present. The level of system noise floor ranged between -22 and -27 dB SPL depending on frequency.

D. Calibration

An *in situ* calibration procedure was conducted on both output transducers before each subject was tested. Tones of fixed voltage were presented to the transducers at 250 Hz intervals from 500 to 15 000 Hz and the resulting SPL of these tones recorded in each ear canal. Based on this information, an equalization of output levels was performed for each subject to achieve target stimulus levels across test frequencies.

Because the sensitivity of the ER-10C microphone decreases at frequencies above 10 000 Hz, it was calibrated using the ER7-C tube microphone which has a flat frequency response beyond 10 000 Hz. Both the ER-10C and ER7-C probes were coupled to a 9-ft long vinyl, open-ended tube with a 1/4 in. internal diameter. Tones of varying frequency were delivered by one of the ER10-C transducers. The resulting sound pressure level of these tones was measured by both the ER10-C and ER7-C at the same position in the tube. The results were compared over the test frequency range and a 10 dB correction was introduced for frequencies from 10 000 to 12 000 Hz (approximately 5 dB/kHz). The error in this equalization process was ± 2.5 dB.

It was evident prior to initiating this study that standing waves in the ear canal could produce significant misestimations of sound level arriving at the tympanic membrane and, subsequently, at the cochlea. A preliminary effort was initiated to measure the difference between levels recorded at the entrance of the ear canal with the ER10C probe microphone and levels recorded at the tympanic membrane with an ER-7 tube microphone in a small group of adult subjects. The objective of this measurement was to derive an accurate correction template by defining the pattern of standing wave interference in your ''average'' adult ear. However, after scrutinizing the data from eight adult subjects, it was concluded that the variable manner in which standing waves impact acoustic energy in each ear canal makes this correction nearly impossible. These data were consistent with the reports of Siegel (1994). Therefore DPOAE suppression tuning curves were collected at 9000 and 12 500 Hz with the understanding that results might be variable due to inaccurate estimates of sound level for both primary tones arriving at the cochlea and distortion coming from the cochlea.

E. Procedure

Adult subjects were tested in a sound-treated IAC booth at the House Ear Institute while reading or sitting quietly. Neonates were tested after obtaining informed consent from one of the parents. They were tested within the hospital, in a room away from the NICU if they were premature, or away from the mother's postpartum room if they were term. They were swaddled and fed if necessary then placed in an acoustically treated infant isolette (Eckels ABC-100 acoustic isolette).

For the collection of DPOAE suppression tuning curves the ratio between primary tone $f1$ and $f2$ was kept constant at 1.21 and the level separation at 10 dB (L1>L2). Five $f2$ frequencies were presented: 1500, 3000, 6000, 9000, and 12 500 Hz. The low-level primary tones were presented at 55–45 dB SPL for 3000 and 6000 Hz; however, DPOAEs evoked by these levels at 1500, 9000, and 12 500 Hz were not easily recorded above the noise floor when suppressed by 6 dB or more. For this reason, the low-level stimulus category for these three frequencies had to be modified to include primary tones at 65–55 dB SPL.

The high-level condition was uniform across the five $f2$ frequencies and included primary tones at 75–65 dB SPL. Based on investigations of the human DPOAE growth function, these levels produce distortion that is just beyond the point of response saturation for most subjects (Abdala, 2000). Including primary tone levels higher than 75–65 dB was not possible. Pilot work has shown that suppressor tones >85 dB SPL are required to suppress these very high-level emissions. We were not able to present pure tones higher than 85 dB SPL because of concerns about patient discomfort. We chose primary tone level categories that could be implemented safely and consistently across subjects and best represent a low-level category (where the cochlear amplifier is known to work most effectively) and a high-level category (where the cochlear amplifier is saturated).

Custom-designed software for the collection of DPOAE suppression tuning curves was developed at the Children's Auditory Research and Evaluation Center, House Ear Institute. An unsuppressed DPOAE was initially measured for a given $f2$ frequency. A suppressor tone (fs) was then presented simultaneously with the primary tones and its level increased in 5 dB steps over a range of intensities. 12–15 suppressor tones with frequencies ranging from one octave below to 1/4 octave above $f2$ were presented at intervals between 25 and 150 cents (one octave=1200 cents). To generate isosuppression tuning curves, the fs level that reduced DPOAE amplitude by 6 dB was determined from the suppression growth function (DPOAE amplitude $\times$suppressor level) using linear interpolation and plotted as a function of fs.

F. Data analysis

DPOAE suppression data were analyzed in the following manner. (1) The tuning curve width was quantified with a $Q10$ value. In the high-level condition, there were many cases where $Q10$ could not be calculated because there was not a 10 dB range between the tip of the STC and the low-frequency boundary of the tuning curve. These tuning curves were uniformly assigned a low Q value (3 standard deviations below the mean of that group) that would clearly denote their broadness. (2) Q difference score (Q DS) was cal-

culated by subtracting the Q value in the high level condition from Q in the low level. Q difference score estimates the enhancement of tuning effect. In order to serve this purpose, however, it is critical that the high-level condition actually reflect primarily passive motion of the basilar membrane; that is, the point of cochlear amplifier saturation must be reached and the recorded distortion must result from primarily passive cochlear motion. (3) Slope of the low- and high-frequency sides of the tuning curve was determined as follows: a regression line was fit to all points from the tip or center frequency of the STC to the data point representing the lowest or highest frequency suppressor tone. If a ''tail'' was present on the low-frequency side, the regression line was fit from the tip to the beginning of the tail portion (prior to the flattened region). A dB/octave slope value was determined from the regression equation. (4) Tuning curve tip frequency/level: (a) Tip frequency refers to the suppressor frequency at which criterion suppression was achieved with the lowest suppressor level. It is the lowest value on the suppression tuning curves (STCs) frequency axis; (b) tip level refers to the level of fs at the tip frequency. (5) Suppression growth slope (or rate) was measured from the suppression growth function. This is a plot of DPOAE amplitude as a function of suppressor level. The slope of suppression growth was determined at each fs by fitting a regression line to the linear portion of the function.

III. RESULTS

A. Suppression tuning curves

Ninety-nine suppression tuning curves (STC) were recorded from adult subjects. 407 tuning curves were recorded from neonates for a total of 506 DPOAE suppression tuning curves. The Appendix displays these 506 tuning curves at 5 $f2$ frequencies for adults, term neonates, and three categories of premature neonate. For adults (bottom row of graphs), STCs appear to broaden when recorded with high-level primary tones. At 1500 and 6000 Hz, the low-frequency side of the adult STC flattens considerably when high-level versus low-level stimuli are used. The adult low-frequency slope at 1500 Hz (high level) appears to be flat but the average slope value is actually 21 dB/oct (low value=0.82 dB/oct). At 6000 Hz, there is more scatter, but the low-frequency side of the STC also flattens considerably when the tuning curves are generated with high-level primary tones. Mean low-frequency slope at 6000 Hz (high level) is 19 dB/oct. As can be seen from the other panels, this flattening of the low-frequency side and related broadening of tuning curve width at high primary tone levels was not clearly present in the neonatal data.

It is also evident from data in the Appendix that the intersubject variability of tuning curve shape is greatest for $f2$ frequencies of 9000 and 12 500 Hz. Neonates in the P2 age category at $f2=12\,500$ Hz are a particularly good example of this scatter. These two $f2$ frequencies were probably maximally affected by standing wave interference in the ear canal. Therefore the actual levels of stimuli reaching the cochlea most likely varied considerably from subject to subject, thus accounting for the increased variability in STC shape.

B. STC width

1. Absolute Q10 values

Level effects. Frequency×level (5×2) analysis of variances (ANOVAs) were conducted separately for each age group to investigate the effect of primary tone level on tuning curve width. Adults and term neonates were the only age groups to show a consistent level effect on $Q10$ (df=1; $f=7.18$; $p=0.008$: df=1; $f=7.34$; $p=0.0079$, respectively); the high-level primary tones produced broader tuning curves than the low-level primary tones for adults and term-born neonates. Primary tone level did not significantly impact the width of the STC in any group of premature neonates and there were no interactions between level and frequency for any age group. This result is shown in Fig. 1 for all age groups. The separation between black lines and dashed gray lines depicts the effect of primary tone level on tuning curve width. The effect is present and consistent across frequencies for adults and term neonates only.

Age effects. Low-Level: An initial age×frequency (5×5) ANOVA for a $Q10$ value found age and frequency effects. The influence of age on $Q10$ (generated with low-level primary tones) was analyzed at each $f2$ frequency independently using a Bon feroni adjustment to the significance level (0.05/5 tests). *Post hoc* analyses revealed that there were no significant effects of age on low-level $Q10$ at any of the five $f2$ frequencies.

High-Level: An initial age×frequency ANOVA for $Q10$ value found age and frequency effects as well as an interaction between age and frequency ($p=0.041$). When the influence of age on Q (generated with high-level primary tones) was evaluated independently at each $f2$ frequency, only 1500 and 6000 Hz showed age effects on tuning curve width (df=4; $f=3.299$; $p=0.01$; df=4; $f=8.069$; $p=0.0001$, respectively). Additional age contrasts between adults and each neonatal group at these two frequencies found that premature neonatal groups (P1, P2, and P3) had narrower tuning curves than adults at these two frequencies. Term neonates had narrower tuning than adults at 6000 Hz, but borderline results at 1500 Hz ($p=0.048$; alpha level set at 0.01 with Bon feroni adjustment). The youngest neonates in the study (P1) also showed borderline differences with adults at 1500 Hz, most likely due to the reduced number of subjects in the P1 age category ($n=6$). The mean $Q10$ data for high-level primary tones are graphically presented in Fig. 2. This figure clearly illustrates the lower Q values (broader STC width) observed in adult data at both 1500 and 6000 Hz.

Consistent with previous work published from our laboratory, frequency effects were present for STCs generated at both low and high primary tone levels. Q generally got larger, indicating sharper tuning, as $f2$ frequency increased.

2. Q10 difference score

The *Q difference score* is calculated by subtracting $Q10$ values recorded at high levels from $Q10$ values recorded at low levels. It is thought to estimate the magnitude of the

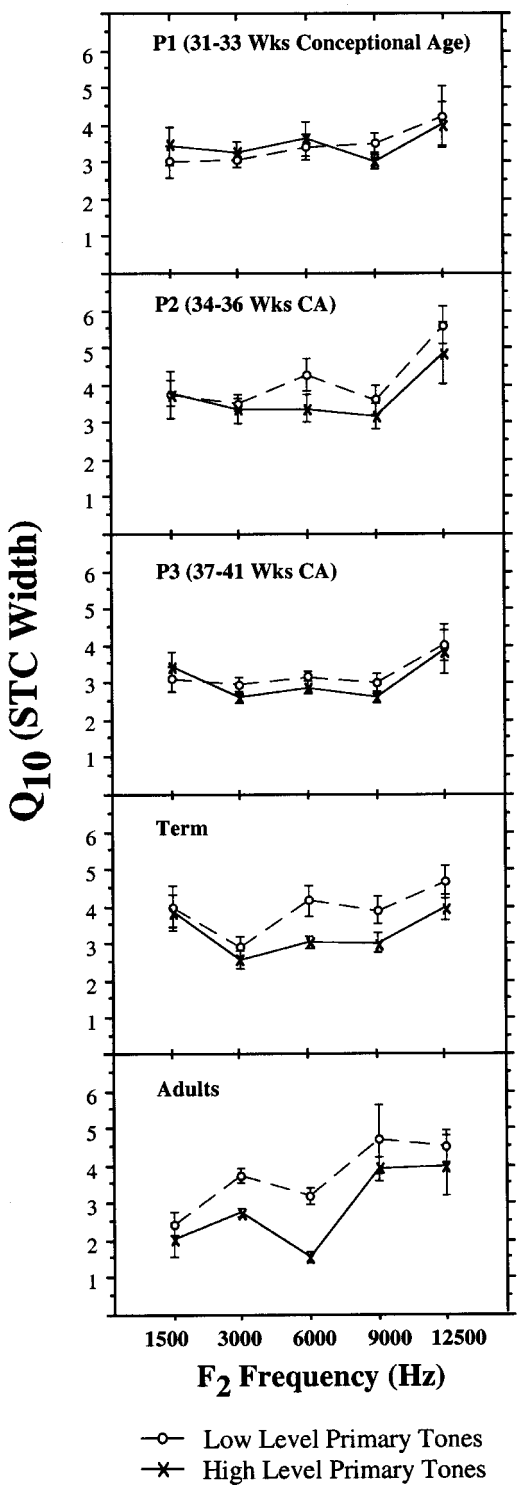


FIG. 1. Mean $Q10$ value as a function of $f2$ frequency for adults, term, and premature neonates. Q values for STCs recorded with both low- (--○--) and high- (--×--) level primary tones are displayed. Adults and term neonates show a level effect and have poorer tuning with high-level primary tones. Premature neonates do not show any level effect on DPOAE suppression tuning.

enhancement effect the cochlear amplifier has on cochlear tuning. However, this simplistic estimate of cochlear amplifier impact on frequency resolution can only be effective if the distortion generated with high-level primary tones reflects generally passive cochlear motion; that is, cochlear motion in the absence of cochlear amplification. Given the limited 20 dB difference between low- and high-level conditions, it is likely that we are not accessing optimal cochlear amplifier function and only observing a portion of its tuning enhancement effect.

A $Q10$ difference score (Q DS) could not be calculated for each individual subject since most participants had only one STC recorded in either the low- or high-level primary tone condition. Therefore the Q DS was calculated by subtracting the *mean* high-level Q score for a given age group and $f2$ frequency from the mean low-level Q score for that same group and frequency.

Figure 3 displays mean Q DS data. There is a general trend for Q DS to increase with age for all $f2$ frequencies except 12 500 Hz. It is largest in adults, however, mean Q difference scores from term neonates are comparable to adult scores at 1500, 9000, and 12 500 Hz. Two out of three of the premature age groups had mean Q DS that fell below 0, indicating no influence of stimulus level on tuning curve width. The largest difference score in any of the premature groups was below 0.5. In contrast, adults and term neonates had mean Q difference scores that ranged from 0.5 to 1.75. These data confirm that adults and term neonates have broader STC width when high-level (versus low-level) primary tones are used, whereas premature neonates do not.

There also appears to be a frequency effect on Q DS. The average amount of tuning enhancement is approximately 1.75 Q units for adults at 6000 Hz and just under 0.5 at 1500 Hz. These units, although unrelated in any concrete manner to cochlear physiology, give an indication of the impact stimulus level (and subsequently, cochlear amplifier function) has on tuning at these two frequencies in the mature cochlea. At 1500 Hz there is little difference in Q between low and high-level primary tone recordings for any age group. It must be noted that the difference between low- and high-level primary tones used to generate the STCs at 1500 Hz was only 10 dB (low=65–55 vs high=75–65 dB SPL). At 3000 and 6000 Hz, in contrast, the primary tone level difference was 20 dB (low=55–45 dB SPL vs high=75–65 dB SPL). This factor likely contributed to the smaller Q DS observed at 1500 Hz. However, it can only provide a partial explanation since there was a robust mean Q DS at 9000 Hz and only 10 dB separates the level conditions at this $f2$ frequency as well.

Figure 4 illustrates the level effect on STC width and shape in six individual subjects. ''Upside-down'' STCs are plotted with the peak up to provide better comparison with cochlear excitation patterns. The lowest frequency point on each STC was normalized to 0 dB and all other points reflect a shift from this value. These six subjects are among the relatively few subjects having contributed both low- and high-level STCs. As is evident, the adult subjects show gross degradation of tuning, a broadened curve, and the loss of the tuning curve tip. The term neonate shown in the left middle panel shows loss of the tip of the STC but retains most of the shape. Premature subjects retain both shape and tip of the tuning curve at high levels of stimulus presentation.

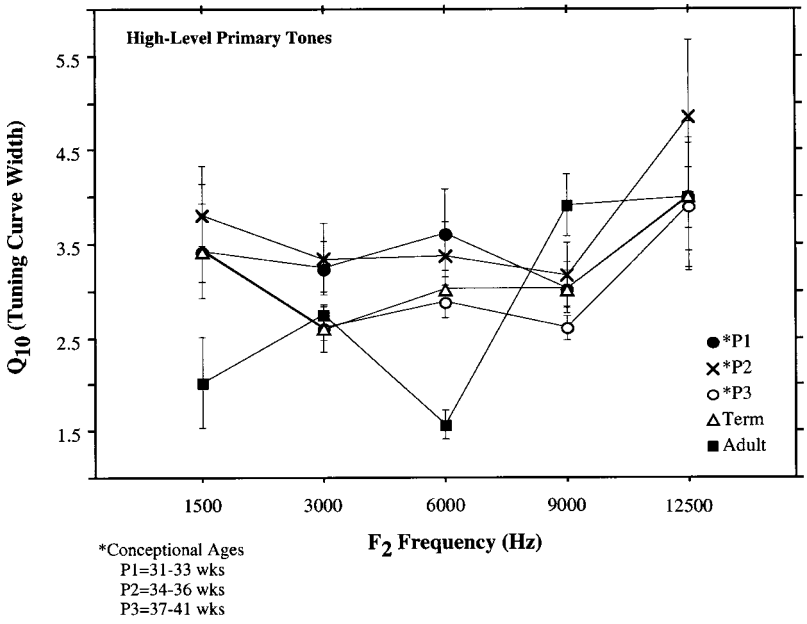


FIG. 2. Mean $Q10$ values for DPOAE STCs recorded with high-level primary tones. There is an age effect on tuning curve width at 1500 and 6000 Hz; premature neonates have narrower DPOAE suppression tuning than adults at these two frequencies.

3. Slope

Only the low-frequency side of the STC showed significant age effects on slope at both low and high primary-tone levels (df=4; $f=2.64$; $p=0.0338$: df=4; $f=4.55$; $p=0.0015$). When further tested at each frequency separately, only slope for 1500 and 6000 Hz tuning curves showed significant age effects on the low-frequency side; slope was shallower in adults than neonates. These results are consistent with the $Q10$ findings described in the preceding section and suggest that age differences in STC width can be explained primarily by a shallower slope on the low-frequency side of the tuning curve.

4. Tip frequency/level

The suppressor frequency at which criterion suppression was achieved with the lowest suppressor level formed the *tip* of the STC tuning curve. *Tip frequency* was always centered around $f2$ but varied with primary-tone level. Age×level (5×2) ANOVAs were conducted at each $f2$ frequency separately. The tip was lower in frequency when the STC was generated with high-level primary tones for all age groups and for all frequencies except 12 500 Hz. Figure 5(a) shows an example of this level trend for $f2$ of 3000 Hz.

There was an age effect only at 6000 Hz but also an interaction between age and level; *post hoc* analyses showed that at 6000 Hz the age effect was restricted to the high-level condition; adults had a lower STC tip frequency than all neonatal groups at 6000 Hz [Fig. 5(b)].

The *tip level* is the primary tone level associated with the suppressor frequency forming the tip of the tuning curve. Past research has shown that STC tip level is always equal to or near the lower-level primary tone (Harris *et al.*, 1992; Kummer *et al.*, 1995). In the present study mean tip level was within 1.5 dB of L2 for all ages and frequencies. Tip

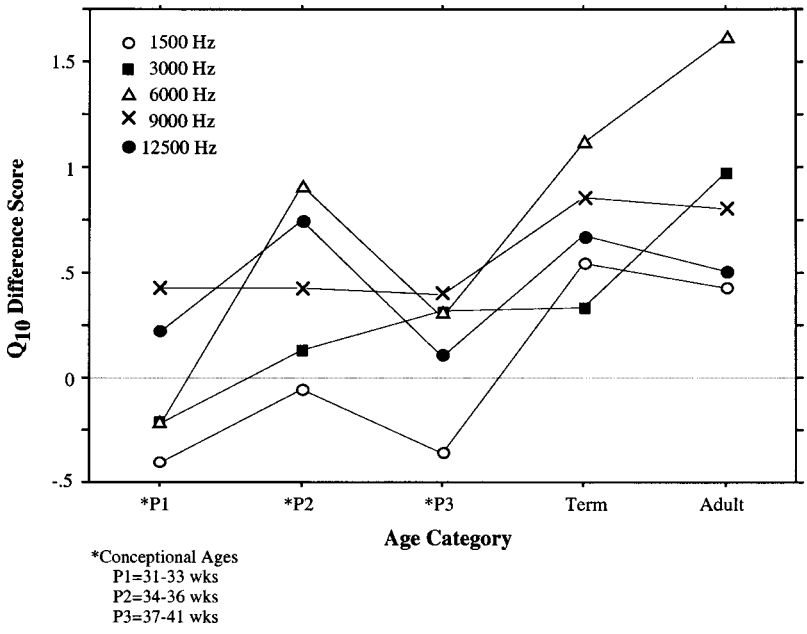


FIG. 3. Mean $Q10$ difference score (low-level Q minus high-level Q) for all age groups and $f2$ frequencies. This score is thought to provide a gross estimate of the "tuning enhancement effect" provided by the cochlear amplifier. As is evident from these data, Q difference score increases with age.

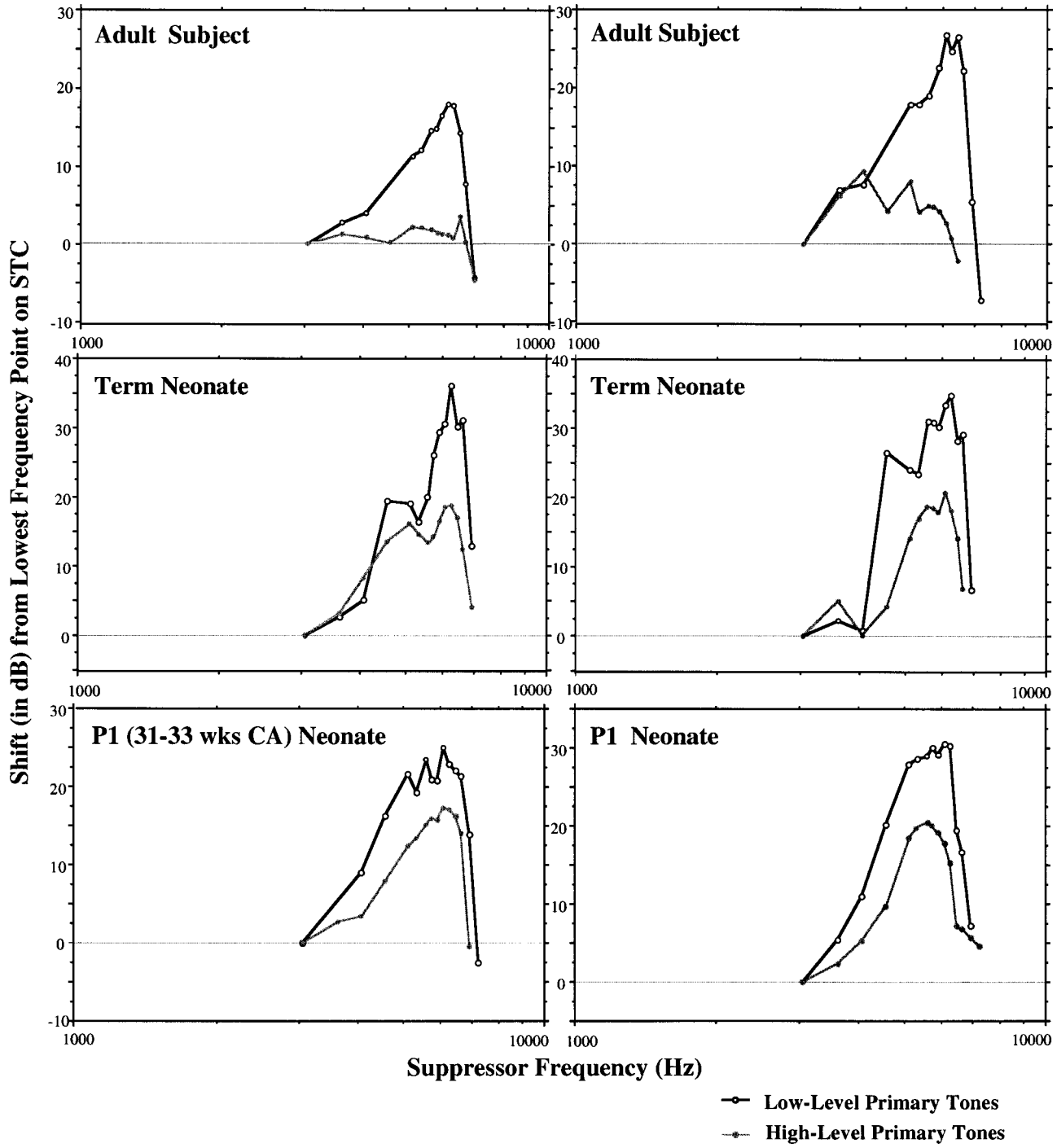


FIG. 4. Suppression tuning curves from six individual subjects who had both low- and high-level data (two adults, two term, and two premature neonates). These suppression tuning curves are plotted with the peak up for better comparison with cochlear excitation patterns. The lowest frequency point on each curve was normalized to 0 dB and all other points reflect a shift from this value. Neonates retain tuning curve sharpness and shape at high primary tone levels whereas adults do not.

level was analyzed separately for low- and high-level primary tone conditions with frequency×age (5×5) ANOVAs. At low primary-tone levels there was no age effect; tip level was comparable among ages. At high primary tone levels there was an age effect on STC tip level (df=4; $f=2.878$; $p=0.0236$), but also an interaction between age and frequency. *Post hoc* analyses of tip level at each $f2$ frequency separately showed that only $f2$ frequencies of 1500, 6000, and 9000 Hz had age effects. Tip level was generally elevated in adult subjects at these frequencies.

5. Summary of STC results

To summarize the primary results of analyses on DPOAE STCs thus far: (1) Only adults and term neonates show broadened tuning in the high- compared to low-level primary tone conditions. (2) At 1500 and 6000 Hz, premature neonates have STCs that are narrower and steeper on the low-frequency side than adult tuning curves. (3) Neonates have higher frequency tip than adults at 6000 Hz only. (4) Tip level of STCs is lower for neonates (elevated for adults) at 1500 and 6000 Hz.

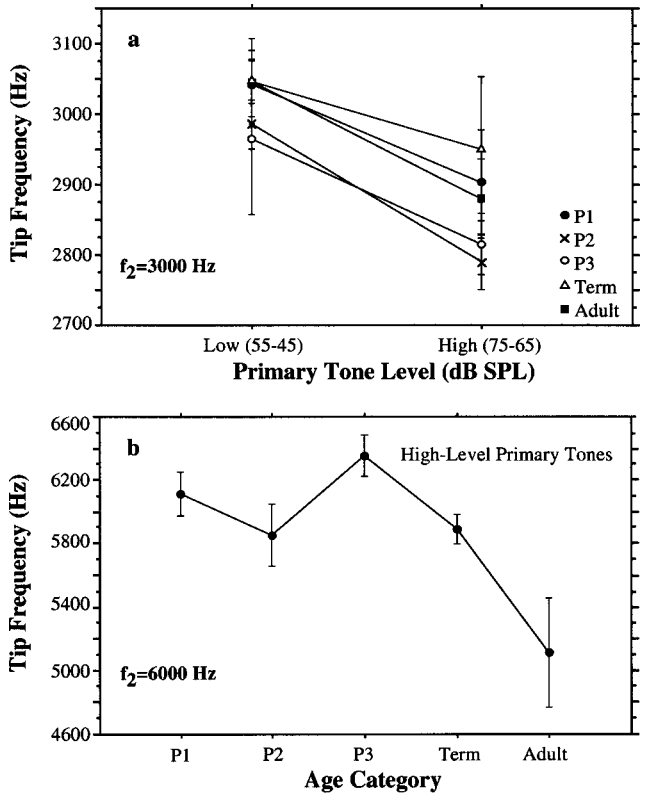


FIG. 5. (a) The mean tip frequency for DPOAE STCs recorded with low- and high-level primary tones at $f2=3000$ Hz. High-level stimuli produced a tuning curve tip shifted toward lower frequencies for all ages. (b) Mean tip frequency for DPOAE STCs recorded with high-level primary tones at $f2=6000$ Hz. Adults had significantly lower tip frequency than neonates at this frequency.

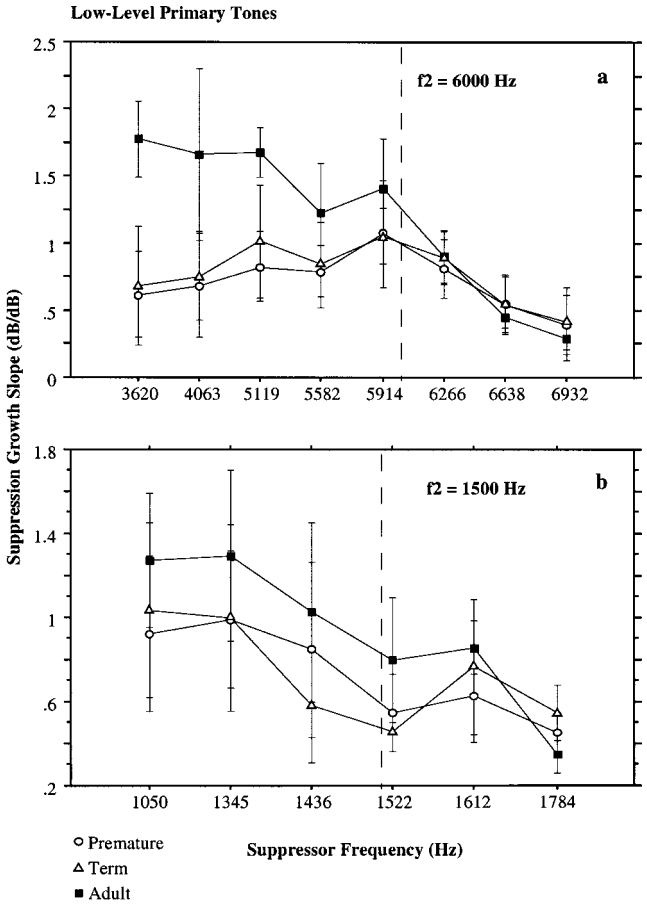


FIG. 6. Suppression growth recorded as a slope value (dB/dB) at $f2$ frequencies of 1500 and 6000 Hz in the low-level primary tone condition. Adults show steeper suppression growth slope than neonates for suppressor tones lower in frequency than $f2$ (to the left of the dashed vertical line).

C. Suppression growth rate

Suppressor frequency×age ($6-8\times5$) ANOVAs were conducted for each $f2$ frequency separately. Although the suppressor frequency variable could be considered a repeated variable, it could not be treated as such because each subject did not provide data at all suppressor frequencies. In a repeated measures design, any row without the full complement of data would simply be eliminated, leaving too few observations for analysis. Therefore a factorial design ANOVA was used. Data from low- and high-level conditions were analyzed separately. Suppression rate varied across suppressor frequency for all age groups and $f2$ frequencies. This has been shown previously and it is expected that low-frequency side suppressors ($<f2$) produce more rapid suppression than high-frequency side suppressors (Abdala, 1998; Kummer *et al.*, 1995), thus resulting in larger slope values.

At low-level primary tones there was an age effect for all $f2$ frequencies and an interaction between age and frequency at 3000 and 6000 Hz only. At high-level primary tones there was an age effect for all $f2$ frequencies except 1500 Hz and there was an interaction between age and frequency at 6000 Hz. The interactions reflect the fact that the age effect is restricted to the low-frequency suppressors in all instances.

Where an age effect was present (nine out of ten conditions), adults had steeper slope of suppression than neonates. This is most clearly evident at 6000 Hz. Figure 6(a) shows the low-level condition for an $f2$ frequency of 6000 Hz. All premature neonates (31–37 weeks conceptional age) have been collapsed into one age group for display purposes. At 1500 Hz [Fig. 6(b)], adults also show consistently larger slope value, indicating more rapid growth of suppression than either premature or term neonates.

IV. DISCUSSION

A. Absolute *Q*10

The absolute values of $Q10$ recorded in this study are generally in good agreement with data reported from this laboratory in the past. Previous experiments have shown: (1) that premature neonates have sharper, narrower tuning curves than adults at 1500 and 6000 Hz using moderate-level primary tones and (2) that term neonates have generally adultlike tuning (Abdala *et al.*, 1996; Abdala and Sininger, 1996; Abdala, 1998). In the present study, this same age effect was observed at 1500 and 6000 Hz. Additionally, the most robust age differences in STC width are between adults and premature neonates and they are most pronounced at 6000 Hz. The lack of age difference at 3000 Hz was a consistent finding in previous work from our laboratory and has been hypothesized to reflect the early maturation of this region on the basilar membrane (see Abdala, 1998).

In working with neonatal subjects born up to 9 weeks premature, it is unreasonable to assume that maturity of the cochlea is the only factor that contributes to age differences in DPOAE suppression tuning. There are undoubtedly unaccounted for variables influencing cochlear maturation and function in these subjects. Gestational age at birth may influence cochlear function and, in turn, influence DPOAE suppression results. Perhaps the earlier a baby is born, the more likely there are medical complications and interventions that can impact cochlear function and/or acquisition of DPOAE data. Despite these uncontrolled variables that can cloud existing age differences, sharper DPOAE suppression tuning in premature neonates is evidently a robust enough result that it has been observed in two independent experiments.

These consistently observed age differences in DPOAE suppression tuning suggest that there is something different in the cochlear function of adults and premature neonates at low (1500 Hz) and high (6000 Hz) frequencies. Interestingly, animal studies suggest that the most apical regions of the cochlea are the last to become anatomically, and thus, functionally, mature in mammals (Bredberg, 1968). Considering this, it would not be surprising to see some subtle, residual immaturity of cochlear function in the lowest frequency tested, during the first part of the third trimester (in 31–33 week old subjects). We have hypothesized in a previous work that the immaturity may reflect a less regulated, somewhat overactive cochlear amplifier (Abdala, 1998). Overshoot of amplifier gain has been previously observed in neonatal gerbils and correlated with fluctuations of the endocochlear potential during auditory maturation (Mills *et al.*, 1993). Tuning may not be *better* in the immature neonates tested here but simply ''sloppier,'' with less precise boundaries of function. Evidence suggests that maturity of the efferent system may be involved in this sequence of maturational events (Abdala *et al.*, 1999).

There may be a second mechanism contributing to the age results at 6000 Hz because adult tuning curve width differs from all neonatal groups at this frequency. It is difficult to account for an immaturity in cochlear physiology for term-born neonates since studies of human cochlear anatomy and function have generally suggested a functionally mature cochlea by 40 weeks gestation (Abdala *et al.*, 1996; Abdala, 1996; Abdala, 1998; Abdala, 2000; Eggermont *et al.*, 1996; Brown *et al.*, 1995; Bargones and Burns, 1988; Lavigne-Rebillard and Pujol, 1987, 1988). A second factor that might contribute to the age effects observed at 6000 Hz is slight hair cell loss in normal hearing adults. This could weaken cochlear amplifier function and consequently broaden the DPOAE suppression tuning curves slightly (relative to neonates) but not impact the audiogram or basic speech perception. Noise exposure-related hair cell damage first occurs in this frequency range (Clark and Bohne, 1978). Ototoxins commonly taken throughout life such as salicylates also selectively affect the high frequencies first (Ballantyne, 1973) and finally, most presbycusis first manifests as a sloping high-frequency SNHL (Schukneckt and Gacek, 1993). Thus the age differences observed at 6000 Hz between adults and all neonatal groups tested in this experiment may be due partially or completely to normal progression of hair cell loss in the adult cochlea rather than immaturity of neonatal cochlear function.

B. *Q* difference score

It was initially considered that the *Q* DS reflects an estimate of tuning enhancement effect provided by the cochlear amplifier; however, there are some conditions required for the difference score to serve this purpose. The *Q* DS reflects tuning enhancement only if the DPOAE STC recorded with high level primary tones reflects a cochlear amplifier that has saturated in its ability to sharpen cochlear tuning. In contrast, the low-level DPOAE STC reflects tuning when the cochlear amplifier is functional and impacting frequency resolution.

However, if any of these conditions is not met, the *Q* DS cannot estimate cochlear amplifier enhancement of tuning. This appears to have occurred with premature neonatal subjects in this experiment. For premature neonates, it is not possible to estimate the magnitude of the tuning enhancement effect from *Q* DS data. Results indicate that passive cochlear function was not accessed in this age group. In premature neonates, tuning in the high-level condition was not markedly worse than in the low-level condition. This is confirmed by the absence of a level effect on *Q* for all premature neonates (see data presented in the Appendix and Fig. 3). In contrast, adults show grossly affected tuning when high-level curves are compared to low-level curves.

Therefore although adults have a larger *Q* DS (as shown in Fig. 3), it is inaccurate to conclude from these data that adults have greater tuning enhancement than neonates (i.e., a more effective amplifier). The adults show a larger *Q* DS in this experiment because their high level STC width is broad (i.e., *Q* value is low), not because their low-level tuning is sharper. At low levels, in fact, there was no real age difference in *Q* between adults and neonates at any $f2$ frequency. Therefore the actual age difference identified by using the *Q* DS metric in this experiment is that adults require a different stimulus level to ''turn off'' the cochlear amplifier than do premature neonates.

This finding may suggest a cochlear immaturity in premature neonates that produces a nonadultlike operational range for cochlear amplifier function. The cochlear amplifier may function over a larger range of levels in neonates at frequencies ranging from 1500 to 9000 Hz as shown by the inability of our high-level primary tones to impact cochlear amplifier function and broaden suppression tuning. A previous paper published from our laboratory is consistent with the hypothesis that neonates may have a different range of cochlear amplifier function (Abdala, 2000). It is thought the saturation plateau consistently observed in the DPOAE growth function reflects the level at which the cochlear amplifier ceases to function efficiently. Our study found that DPOAE growth functions recorded from premature neonates shows amplitude saturation less often than adults and term neonates. When saturation is present in the premature data, it occurs at higher primary tone levels for 1500 and 6000 Hz only. If the saturation plateau defines the upper limit of cochlear amplifier function, these results also suggest that the amplifier works over a larger range of levels in premature neonates than in adults.

One additional observation about *Q* DS data is that there is an apparent difference between the P3 premature group (mean age at birth=34 weeks) and term neonates (mean age at birth=39 weeks). Both of these neonatal groups were tested at 37–41 weeks conceptional age. The difference between P3 and term suggests that maturational status at testing is not the only factor that influences DPOAE suppression; gestational age at birth may contribute. Interestingly, although both P3 and term neonates were in the same conceptional age category at testing, the P3 neonates were tested an average of 2.5 weeks earlier than term babies (DPOAE STC measured at 37.2 vs 39.8 weeks). Therefore it is possible that, although they fit into the same experimental age category, maturational state of the cochlea may, in fact, be different in each of these neonatal groups.

C. Frequency effect

Interestingly, 1500 Hz showed the smallest *Q* DS for all ages. This is a risky comparison since not all $f2$ frequencies had the same level separation between what were defined as ''low-level'' and ''high-level'' primary tones. The difference between low and high levels was only 10 dB for 1500 Hz (20 dB for 3000 and 6000 Hz). However, animal data have shown that the cochlea is less active and shows less nonlinearity near the apical end (Cooper and Rhode, 1995; Robles *et al.*, 1986). This finding is consistent with the *Q* difference score (i.e., tuning enhancement provided by cochlear amplifier) being minimal at the lowest frequency tested.

D. Additional indicators of age differences in cochlear function

Two additional findings are consistent with neonate/adult differences in cochlear function. (1) Tuning curve tip level was consistently lower in neonates than in adults at three $f2$ frequencies. This result could be restated as adults showing elevated tuning curve tips; (2) Neonates showed more shallow growth of suppression than adults. Premature neonates clearly show the most shallow growth.

Mills (1998) found that at low primary tone levels the difference between tip level and level at the low-frequency ''shoulder'' of the tuning curve (called ''TE'') roughly approximated the gain provided by the cochlear amplifier in gerbil. Adults in this study had elevated tip levels at 1500 and 6000 Hz. An elevated tip level would effectively reduce the TE measurement and indicate reduced cochlear amplifier gain in these adults. This idea is consistent with the premise that neonates may have excess cochlear amplifier influence during a stage of ''overshoot'' and/or adults may have diminished amplifier gain due to the cumulative effects of exposure to noise, ototoxins, and aging hair cell loss.

The shallow suppression growth seen in premature neonates, primarily for apical suppressors, has been observed previously (Abdala, 1998). This effect is restricted to premature neonates; in adults, whenever the suppressor tone was lower in frequency than $f2$, it was a very effective suppressor and produced rapid reduction of the DPOAE amplitude. The probable mechanism for suppression is that the apical suppressor tone produces basalward spread of energy that overlaps with the DPOAE generation site at $f2$. In neonates, and more markedly so in premature neonates, suppression of the DPOAE was slower (for suppressor tones lower in frequency than $f2$) and required higher levels of suppressor. This may indicate that the upward spread of masking produced by the suppressor tone is not as pronounced in premature neonates as suppressor level is raised. This may represent a fundamental difference in the cochlea of adults and neonates and is consistent with the differences in tuning curve width, *Q* DS, and tip level.

E. Middle ear involvement

It is difficult to ignore the fact that all of the age effects observed could be adequately accounted for by stimulus level differences; that is, if neonates are receiving lower stimulus levels, tuning would be expected to be sharper and steeper, cochlear amplifier gain greater, and the spread of energy on the basilar membrane more restricted. The inefficient transmission of energy through the neonatal middle ear may attenuate stimulus tones entering the cochlea in babies (Keefe *et al.*, 1993, 1994). This suggests that the conductive pathway could be involved in producing the age differences observed in suppression tuning.

However, at least two arguments discount the middle ear as the primary source of age differences observed in suppression tuning. The first argument is elucidated in a recent publication from the author's laboratory (Abdala, 2000) and is rooted in the fact that saturation threshold of the DPOAE growth function is elevated in premature neonates *more* than we could expect the immature middle ear to attenuate primary tones in this population (Keefe *et al.*, 1993, 1994). Additionally, it is difficult to conceive that the very lowest primary tone levels on the DPOAE growth function (30–20 dB SPL) could evoke a response above the noise floor in humans if they were actually attenuated by 10 dB (20 dB–10 dB SPL) due to middle ear immaturity.

The second argument comes from a test using subsets of data collected in the current experiment. Assuming that the immature middle ear might attenuate primary tone levels in babies by approximately 10 dB, 1500 Hz STCs generated in adults with stimulus levels of 65–55 were compared with STCs generated in neonates with levels of 75–65 dB SPL. If the neonatal cochlea received these pure tones at levels closer to 65–55 dB due to middle ear attenuation, we might expect the resulting STCs to be more similar between adults and neonates. However, STCs from neonates were significantly narrower than those seen in adults at 1500 Hz, even with this attempt to equalize primary tone levels.

Considering the above arguments, results of the present study are most consistent with some kind of subtle immaturity in cochlear function rather than simple attenuation factors stemming from inefficiency of middle ear transmission in neonates. However, it is unwise to disregard the fact that the adult–neonatal differences observed could be rather elegantly accounted for by simple level effects. At present, however, it is not clear what the anatomical or functional origin of these level effects might be.

V. SUMMARY

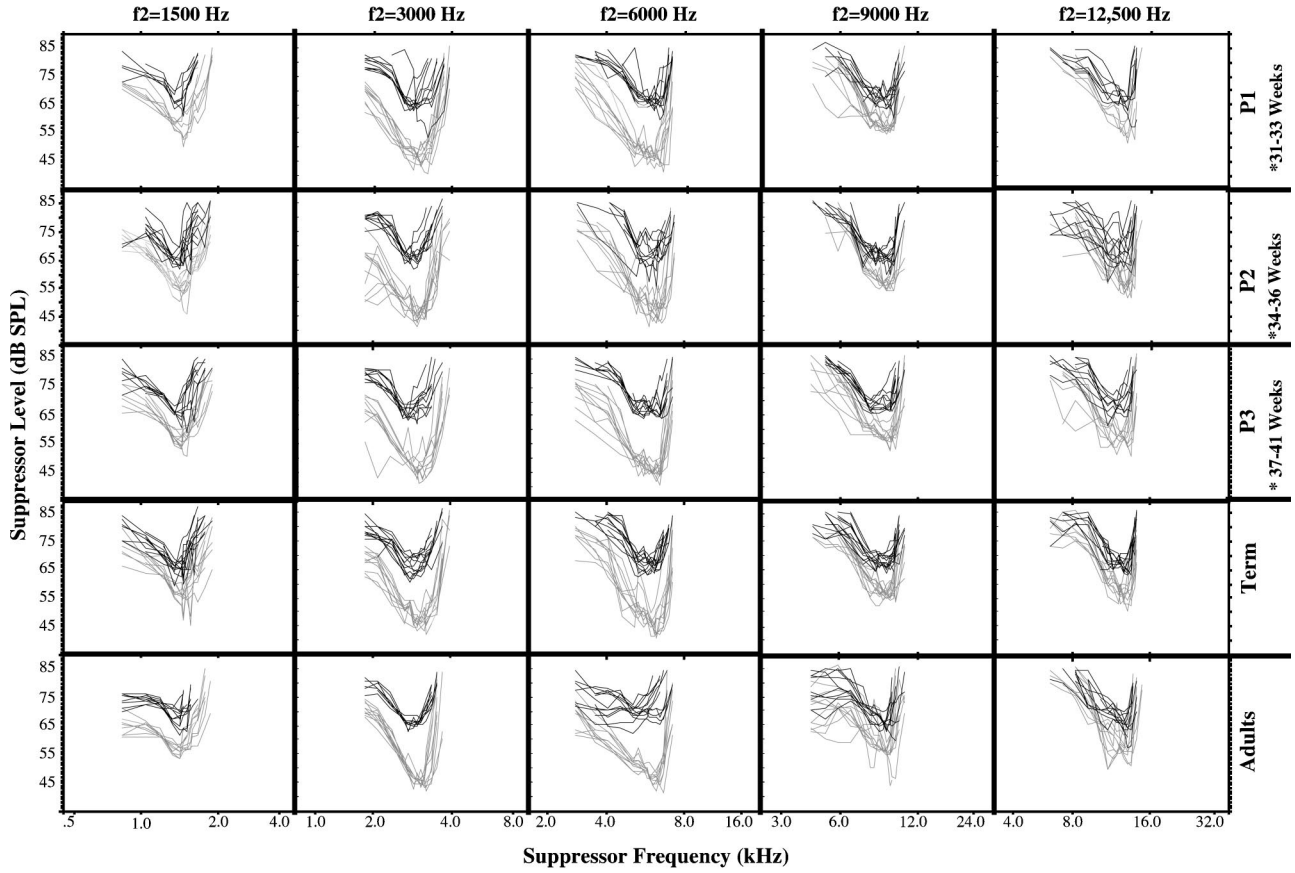
In this study, a simple estimate of cochlear amplifier tuning enhancement effect was applied to human adults and neonates. The results suggest that there are some fundamental differences between suppression tuning of premature neonates and adults. Term neonates, in contrast, are either adultlike or approximate adultlike values in most ways. The primary immaturity identified appears to involve the functional range of the cochlear amplifier. Apparently, we cannot drive the amplifier into saturation at the same stimulus levels in adults and premature subjects. Nonadultlike responses were also present across various indices of suppression tuning ($Q10$, Q difference score, tip level, and suppression growth) and were present at low (1500 Hz) and high (6000 Hz) frequencies. It is hypothesized that the basis of these age differences is cochlear immaturity in the youngest premature neonates.

ACKNOWLEDGMENTS

This work was supported by research grant No. 1 R29 DC03552 from the National Institutes of Deafness and Other Communication Disorders, National Institutes of Health and by the House Ear Institute. The author would like to thank Leslie Visser-Dumont and Ellen Ma for extensive data collection and management.

APPENDIX

The diagrams in this Appendix show 506 DPOAE suppression tuning curves recorded with high-level (black) and low-level primary tones (gray) in 300 term and premature neonates and 30 normal hearing adults. The tuning curves were recorded at five $f2$ frequencies. The asterisk denotes conceptual age.



Abdala, C. (**1996**). ''DPOAE amplitude ($2f1-f2$) as a function of $f2/f1$ frequency ratio and primary tone level separation in human adults and neonates,'' J. Acoust. Soc. Am. **100**, 3726–3740.

Abdala, C. (**1998**). ''A developmental study of DPOAE ($2f1-f2$) suppression in human premature neonates,'' Hear. Res. **121**, 123–138.

Abdala, C. (**2000**). ''Distortion product otoacoustic emission ($2f1-f2$) amplitude growth in human adults and neonates,'' J. Acoust. Soc. Am. **107**, 446–456.

Abdala, C., Ma, E., and Sininger, Y. (**1999**). ''Maturation of medial efferent system function in humans,'' J. Acoust. Soc. Am. **105**, 2392–2402.

Abdala, C., and Sininger, Y. (**1996**). ''The development of cochlear frequency resolution in the human auditory system,'' Ear Hear. **17**, 374–385.

Abdala, C., Sininger, Y., Ekelid, M., and Zeng, F.-G. (**1996**). ''Distortion product otoacoustic emission suppression tuning curves in human adults and neonates,'' Hear. Res. **98**, 38–53.

Ballantyne, J. (**1973**). ''Ototoxicity: A clinical review,'' Audiology. **12**, 325–336.

Bargones, J., and Burns, E. (**1988**). ''Suppression tuning curves for spontaneous otoacoustic emissions in infants and adults,'' J. Acoust. Soc. Am. **83**, 1809–1816.

Bredberg, G. (**1968**). ''Cellular pattern and nerve supply of the human organ of Corti,'' Acta Oto-Laryngol., Suppl. **236**, 1–135.

Brown, A., McDowell, B., and Forge, A. (**1989**). ''Effects of chronic gentamicin treatment on hair cells can be monitored using acoustic distortion products,'' Hear. Res. **42**, 143–156.

Brown, A., Sheppard, S., and Russell, P. (**1995**). ''Differences between neonate and adult cochlear mechanical responses,'' Aud. Neurosci. **1**, 169–181.

Clarke, W., and Bohne, B. (**1978**). ''Animal model for the 4-kHz tonal dip,'' Ann. Otol. Rhinol. Laryngol. Suppl. **87**, 1–16.

Cooper, N., and Rhode, W. (**1995**). ''Nonlinear mechanics at the apex of the guinea pig cochlea,'' Hear. Res. **82**, 225–243.
Dallos, P. (**1992**). ''The active cochlea,'' J. Neurosci. **12**, 4575–4585.
Eggermont, J., Brown, D., Ponton, C., and Kimberley, B. (**1996**). ''Comparison of DPE and ABR traveling wave delay measurements suggests frequency specific synapse maturation,'' Ear Hear. **17**, 386–394.
Frank, G., and Kossl, M. (**1995**). ''The shape of $2f1-f2$ suppression tuning curves reflects basilar membrane specialization in the mustached bat, Pteronotus parnellii,'' Hear. Res. **83**, 150–160.
Harris, R., Probst, R., and Xu, L. (**1992**). ''Suppression of the $2f1-f2$ otoacoustic emission in humans,'' Hear. Res. **64**, 133–141.
Keefe, D., Bulen, J., Arehart, K., and Burns, E. (**1993**). ''Ear-canal impedance and reflection coefficient in human infants and adults,'' J. Acoust. Soc. Am. **94**, 2617–2638.
Keefe, D., Bulen, J., Arehart, K., and Burns, E. (**1994**). ''Pressure transfer function and absorption cross section from the diffuse field to the human infant ear canal,'' J. Acoust. Soc. Am. **95**, 355–371.
Kummer, P., Janssen, T., and Arnold, W. (**1995**). ''Suppression tuning characteristics of the $2f1-f2$ distortion product otoacoustic emission in humans,'' J. Acoust. Soc. Am. **98**, 197–210.
Lasky, R. (**1998**). ''Distortion product otoacoustic emissions in human newborns and adults: II. Level effects,'' J. Acoust. Soc. Am. **103**, 992–1000.
Lavigne-Rebillard, M., and Pujol, R. (**1987**). ''Surface aspects of the developing human organ of Corti,'' Acta Oto-Laryngol., Suppl. **436**, 43–50.
Lavigne-Rebillard, M., and Pujol, R. (**1988**). ''Hair cell innervation in the fetal human cochlea,'' Acta Oto-Laryngol. **105**, 398–402.
Mills, D. (**1998**). ''Interpretation of distortion product otoacoustic emission measurements. II. Estimating tuning characteristics using three stimulus tones,'' J. Acoust. Soc. Am. **103**, 507–523.
Mills, D. (**1999**). ''Calibration of distortion product emission suppression tuning curves by comparison to masked CAP tuning curves,'' Abstr. Assoc. Res. Otolaryngol. St. Petersburg Beach, FL **388**, 98.
Mills, D., Norton, S., and Rubel, E. (**1993**). ''Vulnerability and adaptation of distortion product otoacoustic emissions to endocochlear potential variation,'' J. Acoust. Soc. Am. **94**, 2108–2122.
Mills, D., and Rubel, E. (**1996**). ''Development of the cochlear amplifier,'' J. Acoust. Soc. Am. **100**, 428–441.
Neely, S., and Kim, D. (**1986**). ''A model for active elements in cochlear biomechanics,'' J. Acoust. Soc. Am. **79**, 1472–1480.
Pavolic, C. (**1987**). ''Derivation of primary parameters and procedures for use in speech intelligibility predictions,'' J. Acoust. Soc. Am. **82**, 413–422.
Popelka, G., Karzon, R., and Arjmand, E. (**1995**). ''Developmental characteristics of the $2f1-f2$ distortion product otoacoustic emission (DPOE) in human neonates,'' Ear Hear. **16**, 159–165.
Robles, L., Ruggero, M., and Rich, N. (**1986**). ''Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves, and response phases,'' J. Acoust. Soc. Am. **80**, 1364–1374.
Schuknecht, H., and Gacek, M. (**1993**). ''Cochlear pathology in presbycusis,'' Ann. Otol. Rhinol. Laryngol. Suppl. **102**, 1–16.
Siegel, J. (**1994**). ''Ear-canal standing waves and high-frequency sound calibration using otoacoustic emission probes,'' J. Acoust. Soc. Am. **95**, 2589–2597.
Whitehead, M., Lonsbury-Martin, B., and Martin, G. (**1992a**). ''Evidence for two discrete sources of $2f1-f2$ distortion-product otoacoustic emission in rabbit: I. Differential dependence on stimulus parameters,'' J. Acoust. Soc. Am. **91**, 1587–1607.
Whitehead, M., Lonsbury-Martin, B., and Martin, G. (**1992b**). ''Evidence for two discrete sources of $2f1-f2$ distortion product otoacoustic emission in rabbit. II: Differential physiological vulnerability,'' J. Acoust. Soc. Am. **92**, 2662–2682.

Origin of cubic difference tones generated by high-intensity stimuli: Effect of ischemia and auditory fatigue on the gerbil cochlea

Thierry Mom
Laboratory of Sensory Biophysics (EA 2667), P.O. Box 38, School of Medicine, 63001 Clermont-Ferrand, France

Pierre Bonfils
Laboratoire de Recherche sur l'Audition, Formation Associée Claude-Bernard and Unité CNRS UPRESA 7060, Hôpital Européen Georges-Pompidou, Paris, France

Laurent Gilain and Paul Avan[a)]
Laboratory of Sensory Biophysics, P.O. Box 38, School of Medicine, 63001 Clermont-Ferrand, France

(Received 22 February 2001; revised 6 June 2001; accepted 7 June 2001)

Cubic difference tone (CDT) otoacoustic emissions are thought to arise from the feedback loop allowing outer hair cells to enhance the sensitivity and tuning of the organ of Corti. The existence of residual CDTs during complete cochlear ischemia is therefore disturbing. That stimulus intensities must exceed 50–60 dB SPL for residual CDTs to be recorded and for level notches to be present in CDT growth functions is often cited as evidence for a two-component, ''active/passive'' model: one component, the residual one, would originate from a passive, hardly vulnerable mechanism and thus be unsuitable for hearing screening purposes. This model was probed in gerbil ears after complete interruption of the cochlear blood flow. Cochlear potentials and CDTs were controlled simultaneously through continuous monitoring of CDT level and phase for 50 and 60 dB SPL stimuli and group-delay measurements. After a clear initial decay, CDT levels elicited at 60 dB SPL plateaued for several minutes at about 20 dB below initial level, and when early level notches were observed, CDT phase changes remained minor. The CDT group delays decreased by less than 30%. Later CDT level notches were associated with sharp phase reversals but the similarity between CDT characteristics before and after a notch was hardly consistent with a two-component interpretation. When mild sound overexposure (pure tone, 90–95 dB SPL, 15–30 min) had been performed prior to ischemia, little or no ischemic CDT came from the frequency bands where auditory fatigue had been detected (within 1 kHz), irrespective of the stimulus intensity. It suggests that instead of being passive, residual ischemic CDTs were vulnerable and produced according to a near-normal tonotopy by the same mechanisms that were sensitive to auditory fatigue. All the results lined up with a simple feedback model of cochlear function assuming a single CDT source related to mechano-electrical transduction in outer hair cells. More parsimonious than a two-component model, it posits that although early stages of ischemia dramatically impair the overall performance of the cochlea, the nonlinear mechanical stages responsible for the existence of CDTs keep working albeit at higher intensities. [DOI: 10.1121/1.1390337]

PACS numbers: 43.64.Jb, 43.64.Kc, 43.64.Bt [BLM]

I. INTRODUCTION

The normal mechanics of the mammalian cochlea is characterized by exquisite sensitivity and tuning (Khanna and Leonard, 1982; Sellick *et al.*, 1982). Such performances are thought to stem from the operation of a feedback loop, as initially suggested by Gold (1948), and to involve stimulus-related force generation by the electromotile outer hair cells (OHC; Brownell *et al.*, 1985; Dallos and Evans, 1995). The basilar-membrane (BM) displacement is altered in such a way that the response to the external stimulus gets enhanced in a narrow frequency interval where the phase of the cycle-by-cycle feedback is appropriate. The feedback is more effective at low than high stimulus intensities, so that the growth of the BM response to sounds at increasing intensities tends to slow down in the intermediate range of 30–80 dB SPL (Patuzzi *et al.*, 1984; Yates, 1990; Ruggero, 1992).

The feedback loop is intrinsically nonlinear (Patuzzi *et al.*, 1989b; Kros *et al.*, 1992; Jaramillo *et al.*, 1993; Withnell and Yates, 1998). As a direct consequence, intermodulation distortion products are generated inside the cochlea when two simultaneous pure tones, called primaries, are presented at frequencies $f1$ and $f2$ ($f2>f1$). The most prominent distortion product in mammals is observed at $2f1-f2$ and is termed cubic difference tone (CDT). Distortion is present in the basilar membrane movements (Robles *et al.*, 1991). It is recorded as a pressure wave in scala vestibuli (Avan *et al.*, 1998) as well as an otoacoustic emission in the external auditory meatus (Kim *et al.*, 1980; review in Probst *et al.*, 1991). Whenever the feedback loop is disrupted, there

[a)]Author to whom correspondence should be addressed; electronic mail: paul.avan@u-clermont1.fr

is cogent evidence that sensitivity, tuning, and compression disappear, and so does the CDT as the nonlinearities either vanish or are no longer fed back to BM movements. This is why CDT measurements have become an effective clinical tool to detect OHC dysfunction noninvasively (Kemp *et al.*, 1990; Probst *et al.*, 1991; Trautwein *et al.*, 1996).

To make sure that CDTs provide a reliable diagnosis of cochlear impairment, one must establish whether their characteristics relate to cochlear micromechanics and how they relate to it. This requires identifying what nonlinear process generates a CDT and where it comes from. The latter issue has been partly answered: when two primary tones at $f1$ and $f2$ are sent to the cochlea at low or moderate intensities, i.e., $\leqslant$60 dB SPL, it is widely held that the nonlinear interaction giving rise to the CDT at $2f1-f2$ only involves a restricted area of the organ of Corti, close to the place tuned to $f2$ (Brown and Kemp, 1984; Martin *et al.*, 1987; Avan *et al.*, 1998). When higher primary intensities are used, the origin of the resulting high-intensity CDTs is less clear and its relation to cochlear micromechanics has been called into question, e.g., by Mills *et al.* (1993) or Whitehead *et al.* (1992a, b). That high-intensity CDTs persist despite drastic impairment of cochlear micromechanics (due to furosemide injection—Mills and Rubel, 1994, exposure to loud sound—Puel *et al.*, 1995, or ischemia—Schmiedt and Adams, 1981; Mom *et al.*, 1997) possibly means that they are generated by different, hardly vulnerable nonlinear processes. Besides, the possibility that several places contribute to CDT generation in normal cochleas is supported by evidence of a secondary source basal to $f2$ at high primary intensities (Martin *et al.*, 1998) and of a contribution from the place tuned to $2f1-f2$ that may become dominant in certain stimulus conditions (Brown *et al.*, 1996; Fahey and Allen, 1997; Talmadge *et al.*, 1999). Although the source at $f2$ normally predominates, the balance between contributions might be disrupted in pathological cochleas and lead to unexpected CDT patterns, with unconventional relations to the degree and location of pathological processes.

The apparent robustness of high-intensity CDTs prompted Mills to propose a simple two-component model of CDT generation with so-called ''active'' CDTs elicited by tones well below 60 dB SPL (''active'' with reference to the concept of cochlear amplifier coined by Davis in 1983), reflecting the vulnerable cochlear micromechanics in a restricted BM region around the peak at $f2$, while ''passive,'' less vulnerable CDTs would be generated basal to $f2$ by higher-intensity primaries as a reflection of cochlear macromechanics. The overall ear-canal CDT would be the vector sum of these two components and the normally negligible passive component would be revealed by the impairment of the active one. Clinical procedures using stimulus intensities $>$60 dB SPL might lead to false results if the detected CDTs were of basal, passive origin, in which case these procedures should be avoided as not conservative. The two-component models of this kind account for the observations of persistent high-intensity CDTs for some time after certain severe cochlear insults: they would represent the residual passive contribution. They can also explain the frequently observed notches in CDT growth functions as resulting from the interference between the two discrete, out-of-phase contributions when their levels happen to be similar. However, several features of the ''active'' versus ''passive'' models are not supported by recent experimental data. Lonsbury-Martin *et al.* (1993) have pointed out that high-intensity CDTs are indeed sensitive to cochlear pathology so that their cochlear origin cannot be denied. Hamernik *et al.* (1996) and Trautwein *et al.* (1996) have reported that noise-induced hearing loss or ototoxic drugs do affect high-intensity CDTs as much as low-intensity ones. Frolenkov *et al.* (1998) have shown that once OHC electromotility is destroyed, CDTs vanish whatever the stimulus intensity up to 80 dB SPL. Finally, a recent single-component mathematical model by Lukashkin and Russell (1999) accounts for several features that at first sight were thought to require a two-component model, notably level notches and phase reversals.

The goal of this report is to perform a thorough analysis of CDT characteristics in the course of a complete ischemia achieved by selective mechanical blockage of the gerbil labyrinthine artery on one side, while the animal is kept alive. Ischemia undoubtedly affects cochlear sensitivity, tuning and compressive behavior, although CDTs persist for some time when high-intensity primaries are used (Schmiedt and Adams, 1981; Mom *et al.*, 1997). Evidence for a ''passive'' origin, lack of vulnerability and basal origin of these ischemic CDTs was sought for with the help of a generic two-component model of CDT generation. To this aim, the levels and phases of CDTs were repeatedly collected at various frequencies and stimulus intensities, every few seconds over the course of a complete ischemia, so as to build plots of CDT levels against frequency, CDT growth functions and CDT phase changes at a given frequency $f2$. Special attention was paid to CDT phase changes after versus before temporary CDT cancellations, since a two-component model specifically requires a half-cycle phase difference and identical levels for the two hypothetical components to cancel each other. The round-window cochlear microphonic potential (CM) in response to the primary tone $f1$ was extracted and monitored in parallel with the CDT. Furthermore, by varying the ratio of $f2/f1$, it was possible to derive the group delay of CDTs following the phase-gradient procedure described in Kimberley *et al.* (1993) and O'Mahoney and Kemp (1995), thereby gaining additional information on the cochlear-filter characteristics at the place of generation of CDTs. Actually, early stages of ischemia may not be the best means for affecting CDT sources because they are likely to affect only the endocochlear dc potential through stria-vascularis anoxia, and not directly the OHCs (Billett *et al.*, 1989). Overexposure to sound was expected to be more effective at wiping out high-intensity CDTs as its target, i.e., OHC stereocilia bundles, present well-documented nonlinearities which make them good candidates as CDT sources (Jaramillo *et al.*, 1993; Lukashkin and Russell, 1999). The goal of the present work was to submit a subset of gerbils to mild pure-tone overexposure so as to induce controlled OHC dysfunction in a narrow interval whose characteristic frequency obeyed the well-known half-octave shift law (e.g., Cody and Johnstone, 1981; McFadden, 1983). Exposure level and duration were such that CDTs remained present

even in the frequency range of maximum fatigue. The effect of subsequent ischemia on fatigued CDTs was compared to that on CDTs from unexposed sites of the cochlea.

II. MATERIAL AND METHODS

The 26 Mongolian gerbils (Meriones unguiculatus, males, 3–4 months, 55–75 g) used for this study came from a local breeding. Their care and use complied with the rules promulgated by the French Department of Agriculture and the principles of the declaration of Helsinki (Grant No. EA-2667/1999, University of Auvergne, ''cochlear ischemia in gerbils'').

All experiments were performed in a sound-proof booth. We chose to perform selective unilateral cochlear ischemia rather than sacrificing the gerbils. Although more demanding, this procedure ensured that the measurements were not influenced by middle-ear muscle rigor nor any general postmortem change. Ten minutes after i.p. premedication with atropin sulfate (0.1 mg) and levomepromazin chlorhydrate (1 mg), gerbils were anaesthetized by an i.p. injection of chloral hydrate (480 mg/kg, renewed as needed, i.e., half-dose every hour). Their rectal temperature was kept at $38\pm0.1\,^{\circ}\mathrm{C}$ using a thermoregulated heating blanket. They were prepared according to a previously described protocol (Mom *et al.*, 1999). Briefly, the tympanic membrane was punctured in order to prevent any middle-ear pressure gradient to build up during the experiment. The right auditory bulla was opened dorsally and the tip of a thin flexible optic fiber probe (Perimed PF 418, B500-0, 0.5 mm diameter) was placed against the bone of the first cochlear turn, in front of stria vascularis, after gentle resection of the middle-ear mucosa. A silver-wire electrode was placed against the round-window membrane. Once the optic fiber and electrode were secured, the bulla was closed with dental cement. A posterior transcranial approach allowed the eighth-nerve complex to be exposed at the porus of the internal auditory meatus, after partial suction of the cerebellum. One or several branches of the labyrinthine artery were visible at high magnification, running along the eighth-nerve bundle. Cochlear blood flow (CBF) was continuously monitored with a laser Doppler velocimeter (LDV, Perimed PF 4000) connected to the optic fiber. Finally, the tapered tip of an acoustic probe containing a sensitive microphone (ER10B, Etymotic Research) and two plastic tubings connected to the outputs of two earphones (Etymotic, ER2) were sealed in the external auditory canal of the right ear with cyanoacrylate glue.

The tip of a fire-shaped glass pipette was blocked in the porus of the internal auditory meatus under microscopic control. When the pressure exerted on the tip was large enough to counteract the arterial blood pressure, a sharp decrease of the CBF signal from the LDV was observed. For the ischemia to be considered as complete, the LDV signal had to drop to a stable value, likely corresponding to residual Doppler-shifted light diffusion from vessels in the bony labyrinth. It was necessary to check that this contribution remained stable and did not change after the end of every experiment, when the whole eighth-nerve bundle including the labyrinthine artery was severed to provide a control floor value of CBF.

A. Continuous monitoring

This procedure was performed in 12 gerbils. A computer-controlled Cub^{e}Dis system collected the CDTs (Allen, 1990; software Cubdisp and Cubdisc v2.43, Mimosa Acoustics). Two calibrated isointensity primary tones at frequencies $f1$ and $f2$ ($f2/f1=1.20$) were continuously emitted in the gerbil's right ear canal by the two ER2 earphones (Etymotic Research) connected to the two digital-to-analog outputs of an Ariel DSP16+ board plugged in a personal computer. The calibration procedure ensured that the primary tones were always played at the target intensities and with fixed phases, so that in fixed stimulus conditions, the CDT kept a constant level and phase as long as the cochlear and middle-ear physiologies remained stable. The analog-to-digital input of the DSP board recorded the sound pressure measured in the ear canal by the ER10B microphone (Etymotic Research). The spectral component at $2f1-f2$ was extracted, and the noise floor was estimated from the mean value of a few measurements of the acoustic signal at frequencies close to $2f1-f2$. Instrumental distortion was defined as the level recorded at $2f1-f2$ when the tip of the probe housing the two earphones and the microphone was placed in a 0.25 cm^3 plastic cavity meant to represent a dummy gerbil's ear.

Initially, primary intensities were set at a common value L, and $f2$ was swept from 16 to 2 kHz (ten steps per octave, 1 s per step). A CDT frequency profile (or DPgram) was obtained by plotting the sound level at $2f1-f2$ against $f2$. DPgrams were collected with $L=50$ dB SPL, then with L increasing from 55 to 80 dB SPL in 5 dB steps. They served as initial references and the next DPgrams collected at various stages of the ongoing experiment were meant to control the functional status of the OHCs along the organ of Corti. The monitoring protocol itself consisted in continuously playing the primaries with fixed frequencies ($f2=8$ kHz) and intensity ($L=60$ dB SPL) so as to register the time course of CDT level and phase (one sample every 10 s). Our strategy to probe the two-component model was to search for notches in the plot of CDT level changes versus time and check whether or not the corresponding CDT phase change was about half a cycle: this is to be expected if the notch were due to interference between two out-of-phase components, when the large vulnerable one has decreased down to the level of the small robust one just before vanishing. When such a 180° phase change was found, the characteristics of the residual, postnotch CDT were compared with those of the prenotch, allegedly ''active'' CDT in order to identify possible changes to ''passivity.''

Some indirect insight into the cochlear electrochemical gradients was useful and for this purpose, the output of the round-window electrode was amplified (Grass P511, filter bandwidth 0.3–30 kHz, gain $\times5000$) and sent to the input of a lock-in amplifier (two-channel EG&G 5201, analysis time 1 s). The sinusoidal voltage at $f1$ coming from one of the digital-to-analog outputs of the DSP board was adjusted to 1 V rms through a power amplifier and fed to the reference input of the lock-in amplifier. In this manner, the rms value of CM at $f1$ was extracted in real time and stored in parallel with the CDT at $2f1-f2$.

B. Discontinuous monitoring

Eight gerbils were submitted to discontinuous controls following three purposes. The frequency dependence of the effects of ischemia was checked in three ears, in which DPgrams were collected with $L=60$ dB SPL every minute for 30 min following CBF interruption.

In five gerbils, $f2$ was fixed at 8 kHz and $f1$ was swept in such a way that $f2/f1$ varied from 1.15 to 1.25 with a step of approximately 0.02. The phases ϕ of the CDTs were unwrapped in order to remove one-cycle ambiguities and the group delay τ of the CDT was derived from the frequency dependence of ϕ as

$$\tau = 1/2\pi d\phi/2df1.$$

This procedure could be repeated every 30 s in the course of an experiment. The intensity of the two primary tones was set at $L=60$ dB SPL for odd recordings and 50 dB SPL for even ones, so that the influence of primary intensity on ischemia-induced changes in CDT characteristics could also be established. The levels of the CDT elicited by a ratio $f2/f1=1.20$ were retrieved offline and plotted against time, with a sampling period of 1 min for each primary intensity, thereby providing two intertwined plots somewhat comparable to that of the continuous CDT monitoring procedure.

C. Pre-exposure to a loud tone

In six gerbils, the ischemia experiment was performed in a slightly different manner. After the DPgrams serving as initial references were collected as explained previously, the right cochleae were exposed to a relatively loud pure tone at a fixed frequency f_{expo}, with $f_{\text{expo}}=5$ kHz in three gerbils and 8 kHz in the three other ones. The tone intensity was set at 90 dB SPL and the overall exposure duration was 15 min. A complete series of postexposure DPgrams was recorded every 5 min for 30 min. If the first one did not reveal a clear decrease of at least a few CDT levels, another exposure was performed for 15 min at 95 dB SPL. Once the effects of auditory fatigue on CDTs were stabilized after initial partial recovery, the thresholds of eighth-nerve compound action potential (CAP) were determined using a conventional stimulus and averaging protocol (see Mom *et al.*, 1997). Stimuli were calibrated tone bursts at frequency f (f ranging from 4 to 20 kHz; trapezoidal envelope; rise and decay time: 1 ms, plateau duration 4 ms; 5 dB intensity steps) generated by a Wavetek 70 synthesizer connected to the acoustic probe sealed in the ear canal. They were repeatedly played and the output from the round-window electrode was synchronously averaged (CED 1401+, 10 ms epochs, sampling rate 20 kHz; repetition rate 19/s, 64 averages). The CAP threshold at frequency f was defined as the lowest SPL giving rise to a visually detectable wave N1. Next, the cochlear blood supply was blocked and CDTs were monitored with the help of DPgrams. Over the first 20 min of ischemia, they were recorded every min only at 60 and 70 dB SPL and between 4 and 16 kHz, then after 20 min, i.e., when CDT levels varied little with time, complete series of DPgrams were collected every 5 min for 40 min.

III. RESULTS

A. Initial cochlear responses

1. Ears unexposed to the loud tone

The CDT levels in response to 60 dB SPL primary tones with $f2=8$ kHz were around 20 dB SPL (range 15–26 dB SPL), whereas the noise floor around 8 kHz was close to -20 dB SPL. The residual CDT level found in a dummy cavity (0.25 cm^3) using the same primary characteristics did not significantly differ from the noise floor. When $f2$ was swept from 16 to 2 kHz, the DPgrams at 60 dB SPL as well as other primary intensities were relatively flat, within ± 7–8 dB. The noise floor did not change much between 4 and 16 kHz, and rose to -10 dB SPL when $f2$ decreased from 4 to 2 kHz.

Whatever $f2$, the slopes of CDT growth functions ranged between 0.55 and 0.85 below 60 dB SPL. They could even reach negative values between 60 and 65 dB SPL. Between 65 and 70 dB SPL, most CDTs grew at a rate >1 dB/dB. At all intensities, the levels and phases of CDTs remained constant within 0.2 dB and 5° as long as the CBF was not manipulated. The CM at $f1$ was also stable and for $L=60$ dB SPL its amplitude was of the order of 15 μV with little $f1$ dependence.

2. Exposed ears

In the six animals exposed to loud sound, the first DPgrams after exposure revealed a decrease in CDT levels at several frequencies $f2>f_{\text{expo}}$, relative to the DPgrams before tone exposure. In the four ears presenting a V-shaped high-frequency decrease in their DPgrams, it was largest for $f2$ around half an octave above f_{expo}. The other two ears presented a sloping profile of loss. Partial recovery was observed during the first 15 min after the end of exposure, then the CDT levels stabilized within 1 dB at all frequencies. At this stage and as in the unexposed ears, the phases of CDTs remained constant within 5° provided the CBF was not manipulated. Pilot experiments in six gerbils had shown that such CDT stability was granted for at least an hour, and that later on the DPgrams tended to recover further back to reach their initial profiles within 1 dB about 4 h after the end of exposure. After 30 min, the largest CDT decrease was 15–30 dB for $f2$ ranging from 10 to 15 kHz. Smaller but significant CDT decreases were present in a frequency interval of about one octave. A typical DPgram after 30 min is represented on Fig. 1 (dashed lines).

Within the frequency range of CDT level decrease, the slopes of CDT growth functions got steeper (about 2 dB/dB). Thus, exposure-induced CDT level changes tended to be smaller when primary intensities increased (see Fig. 1, thin versus bold lines for 60 vs 70 dB SPL primaries). In order to be as sensitive as possible to the effects of overexposure, the boundaries between unaltered and fatigued ranges were defined on 60 dB SPL DPgrams as the smallest and largest frequencies, $f2_{\text{min}}$ and $f2_{\text{max}}$, with CDT level changes >-1 dB. The $f2_{\text{min}}$ boundary is marked with a thin vertical line in Fig. 1. The frequency patterns of CAP threshold shifts were

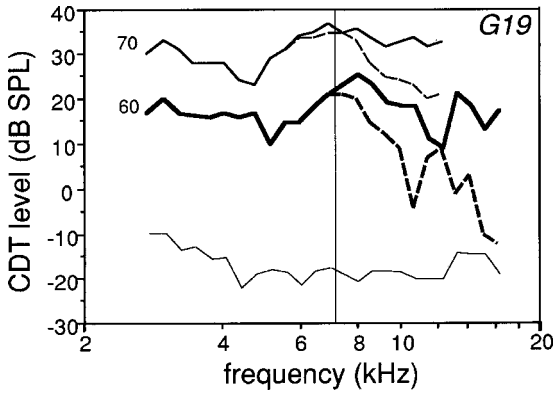


FIG. 1. Series of DPgrams representing an example of initial CDT levels against frequency $f2$ at primary intensities of 70 (continuous line) and 60 dB SPL (bold continuous line), and CDT levels 30 min after the same ear (G19) had been exposed to a loud tone at $f_{expo}=8$ kHz, at primary intensities of 70 (dashed line) and 60 dB SPL (bold dashed line). Below 7 kHz (thin vertical line), the dashed lines are hidden beneath the continuous ones as CDT levels were not changed by the tonal exposure. The thin continuous line in the bottom represents the average noise floor as a function of frequency.

very similar to those of DPgrams, the largest CAP threshold changes being observed at about half an octave above f_{expo}. None of them exceeded 20 dB.

B. Ischemia, CBF data

In all ears, the CBF remained stable as long as the cerebellopontine angle was not manipulated. When the pipette approached the porus acusticus, the LDV output could show a transient increase (around -5 and -2 s in the example of Fig. 2, top), then a sudden sharp decrease occurred as soon as the pipette was blocked in the porus. Within a few seconds, the LDV output reached a stable value. It eventually turned out to remain unchanged after the control section of the eighth-nerve bundle had been performed after the end of the experiment. This indicated that CBF interruption was quasi-immediate and complete.

C. Cochlear responses during ischemia in unexposed ears

1. Round window CM

The CM in response to the 60 dB SPL primary tone at $f1=6.6$ kHz started decreasing within seconds after CBF interruption and its average decrease after 10 s was about 6 dB. The drop in CM amplitude continued steadily during the next 50 s. An average level of -20 dB (s.d. 7) with reference to CM initial value was reached after 1 min of complete ischemia (example in Fig. 2, closed symbols). The foregoing decrease of CM was monotonic and very slow, so that the noise floor (around -30 dB below initial level) was hardly reached before 1 h.

2. CDT level at f2=8 kHz for 60 dB SPL primary intensities

The average CDT level change (Fig. 3) was less than 1 dB for the first sample collected after the beginning of ischemia (i.e., within 10 s). Ten seconds later, the next sample started to exhibit a more pronounced decrease (about 5 dB

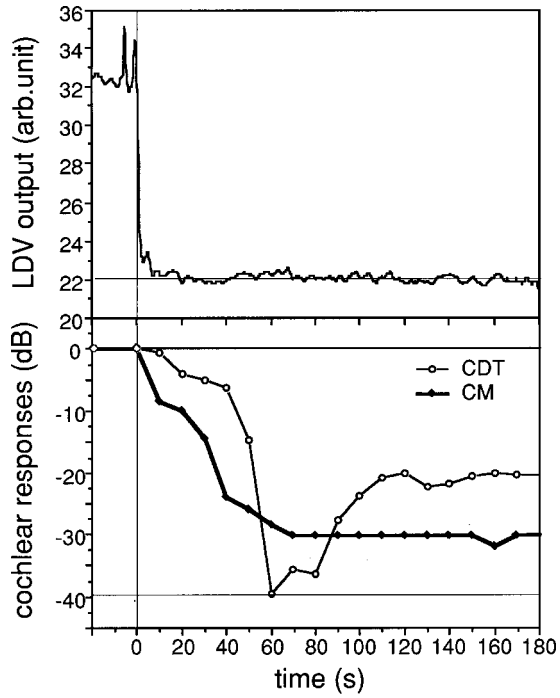


FIG. 2. (top) Example of output of the laser Doppler velocimeter, in arbitrary perfusion units (thin vertical line: application of the pipette in the porus acusticus; thin horizontal line: residual signal at the end of the experiment after the acousticofacial bundle was severed). (bottom) Resulting changes of CDT ($f2=8$ kHz, $f1=6.6$ kHz, primary intensities 60 dB SPL; open circles) and of CM at 6.6 kHz (closed circles). The two cochlear responses are expressed in dB relative to preischemic level. Thin horizontal line: noise floor for CDT. The noise floor for CM was -36 dB.

on average) and the CDT level reached an average value of -14 dB relative to its initial value after 60 s. In 7 ears out of 12, the early course of CDT level was not monotonic although CM had almost reached its plateau, so that after decreasing to a sharp notch of $-30/-35$ dB after 1–2 min (example in Fig. 2, bottom, open symbols), the CDT levels immediately rose to a plateau slightly >-20 dB below initial level, which lasted about 2 min (Fig. 2. Also see Fig. 3 around 175 s). The type of notch observed here will be termed ''early notch'' in the following. It happened at variable times from one ear to another and could not be identi-

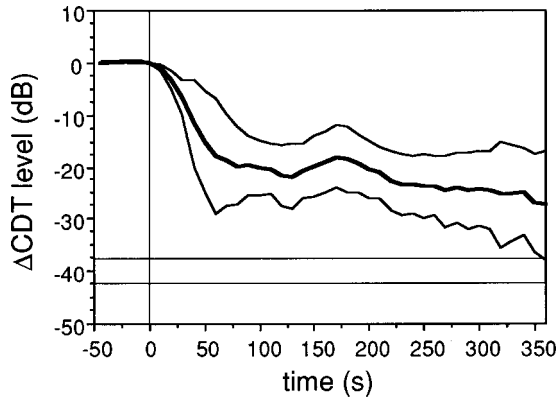


FIG. 3. Average changes of CDT ($f2=8$ kHz, $f1=6.6$ kHz, primary intensities 60 dB SPL) over the first minutes of complete ischemia starting at $t=0$ (in dB relative to preischemic level). Thin lines: average±one s.d. ($n=12$). Thinner horizontal lines: average noise floor±one s.d.

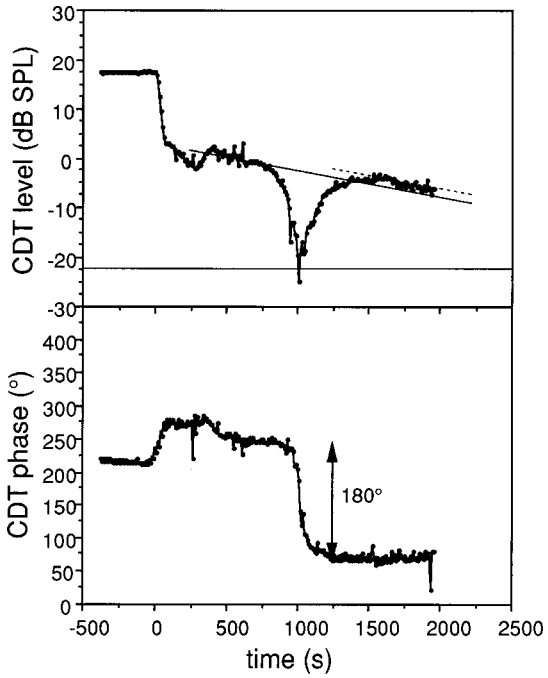


FIG. 4. Example of change of CDT level (top diagram, in dB SPL; $f2 = 8$ kHz, $f1 = 6.6$ kHz, primary intensities 60 dB SPL) and CDT phase (bottom diagram, in degrees) over prolonged complete ischemia, exhibiting a late notch around $t = 1000$ s. Thin horizontal line (top diagram): average noise floor. The thin oblique lines on the top diagram represent the linear least-square fits of CDT levels between 500 and 800 s (continuous), then 1500 and 1800 s (dashed).

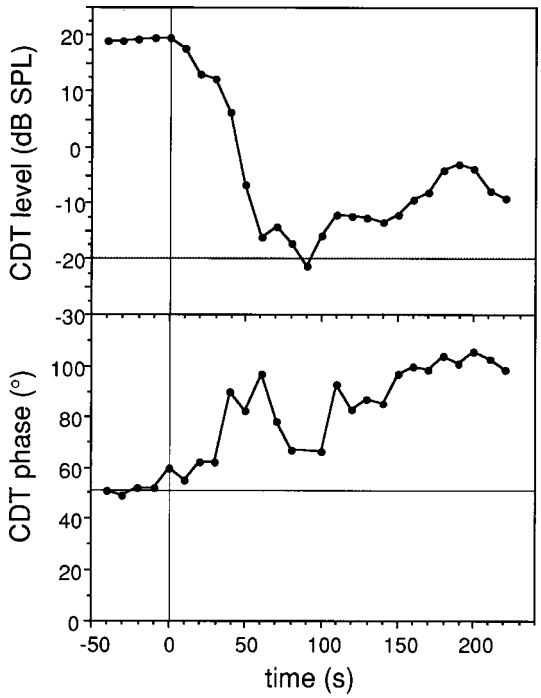


FIG. 5. Example of change in CDT level (top diagram, in dB SPL; $f2 = 8$ kHz, $f1 = 6.6$ kHz, primary intensity 60 dB SPL) and CDT phase (bottom diagram, in degrees) over the first minutes of a complete ischemia, exhibiting a sharp early notch around $t = 90$ s. Thin horizontal lines: preischemic CDT level and average noise floor (top diagram); preischemic CDT phase (bottom diagram).

fied in 5 ears (Example in Fig. 4, top diagram, between 0 and 300 s). This is why in Fig. 3 the CDT level seems to barely decrease below −20 dB in the first minute of its time course. As individual noise floors were between −37 and −42 dB with reference to CDT initial level, CDTs spent most of the first 5 min after ischemia onset at least 10 dB above noise or instrumental distortion.

The CDT decrease following the plateau was very slow during the next few minutes (Fig. 4). At a highly variable time ranging from 6 to 20 min after CBF blockage (16 min in Fig. 4), all ears exhibited a sudden drop of CDT almost down to the noise floor. It was rapidly followed by a rebound back to a level ranging from +3 to −11 dB (average: −3.0 dB) relative to the level predicted by linear extrapolation of the slow monotonic decrease preceding the notch (thin continuous line in Fig. 4; in this example, the postrebound level is almost identical to the extrapolated one). Another similar sudden notch was sometimes observed several minutes later. These notches will be termed ''late notches'' in the following.

In between and after the brief late notches, CDT levels tended to slowly decrease down to noise floor with an average slope of 0.25 dB/min (s.d. 0.06). In all ischemic ears, more than 1 h was required for CDTs to be less than 5 dB above noise floor.

3. CDT phase for f2=8 kHz and 60 dB SPL primary intensities

Just after CBF interruption, within 10 s, a significant increase in phase was observed. The phase increase was slow during the first 2–3 min of ischemia, and often slightly irregular as in the example of Fig. 5 (bottom diagram). The phase changes observed in the course of the early notch never exceeded 60° in the 7 ears in which it was identified. After about 3 min, the phase reached a plateau corresponding to an average change of +62° relative to initial phase (s.d. 25°; range, 29°–93°). The plateau remained relatively stable until a late notch appeared in the course of CDT level change: within 1 min, the CDT phase underwent a sharp shift of 180° (within ±10°) (at about 1000 s after ischemia onset in Fig. 4) while the CDT level got close to the noise floor and bounced back.

4. CDT characteristics at other frequencies and primary intensities

The repeatedly recorded DPgrams remained comparatively flat. In other words, their shapes did not change much throughout ischemia, the only effect being an overall decrease of CDT levels following the same course as the CDT with $f2 = 8$ kHz. The time course of CDT levels for alternating primary intensities (60/50 dB SPL) was reconstructed off line from the phase-gradient protocol performed on 5 gerbils at $f2 = 8$ kHz. The level change of the CDT elicited by 50 dB SPL primaries tended to parallel that at 60 dB. For instance, Fig. 6(a) shows that the early notch was present at about the same time, i.e., 3.5 min, and that a plateau also existed for 50 dB CDTs between 4 and 7 min. After 7 min, the signal-to-noise ratio at 50 dB SPL was too low for CDTs to be considered significant.

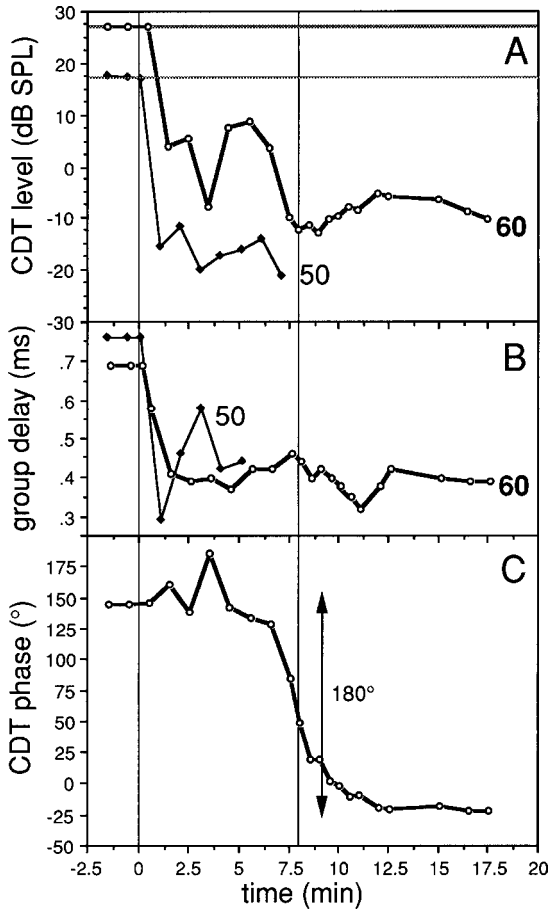


FIG. 6. Example of change of CDT characteristics with complete ischemia starting at $t=0$ (leftmost thin vertical line). (a) CDT levels in dB SPL; $f2=8$ kHz, $f1=6.6$ kHz; primary intensities 60 dB SPL—bold line—or 50 dB—thinner line. An early notch is visible around $t=3.5$ min, and a late notch around $t=7.5$ min (rightmost thin vertical line). (b) Group delays of CDTs derived from their phase gradients when $f1$ was swept around 6.6 kHz; primary intensities 60 dB SPL—bold line or 50 dB—thinner line. The group delay after mechanical destruction of the cochlea (primary intensity 80 dB SPL) was 0.03 ms. (c) phases of CDT in degrees; $f2=8$ kHz, $f1=6.6$ kHz; primary intensities 60 dB SPL.

Comparison of the two profiles of CDT levels for 50 and 60 dB SPL primaries at about the same times (within the 30 s required to switch from a primary intensity to the other) provided an estimation of the slope of the CDT growth function. Prior to ischemia, the average slope was 0.6 and CBF interruption induced a slope increase, up to 2.5 on average, as long as CDT could be detected at 50 dB SPL.

5. CDT group delay

After phase unwrapping, a regression line was adjusted to the experimental data in order to derive the group delay τ from its slope. It was always possible to find a straight line ensuring a determination coefficient $r^2>0.9$. The delay was plotted against time as shown in Fig. 6(b). Instrumental delay never exceeded 0.1 ms. Likewise, delays <0.1 ms were found in the contralateral ear using primary intensities >70 dB SPL after mechanical destruction of its cochlea. A decrease of τ was observed shortly after ischemia, then within 2 min, τ tended to plateau at about 57%–72% of its reference value before ischemia. Fot the next 30 min, τ did not decrease below this plateau value. Instead, it could even sometimes increase to about 80% of the reference value after several minutes. The group delay was insensitive to the presence of a late notch: Figures 6(a) and (c) depict a late notch with phase reversal around 7.5 min after CBF interruption, while the group delay does not vary by more than 0.05 ms between 10 and 20 min. Only after more than 1 h did the group delays get closer to the instrumental-distortion value.

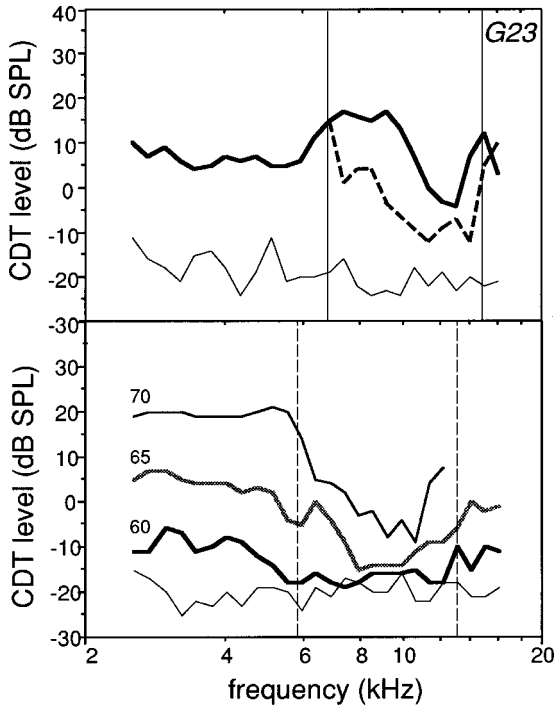


FIG. 7. Example (in ear G23) of changes in DPgrams due to exposure to a loud tone followed by complete ischemia. Top diagram: 30 min after vs before exposure to a 8 kHz/90 dB SPL tone, 60 dB SPL primaries, continuous line before exposure, dashed line after exposure; Thin broken line: noise floor. The two thin vertical lines delimit the frequency interval with significant CDT level drop due to auditory fatigue. Bottom diagram: DPgrams 10 min after onset of complete ischemia. Primary intensities: left labels; thin line: average noise floor. The dashed vertical lines delineate the frequency range where postischemic CDTs clearly differ from those coming from the fatigue-free range.

D. Ears previously exposed to a loud tone

In striking contrast with the results of the previous paragraph in unexposed ears, the behavior of CDTs in fatigued ears was highly frequency dependent. In the frequency range where CDTs had been unaffected by exposure to the loud tone, their ischemic level and phase changes were similar to those in unexposed ears, and CDT level growth ranged between 2.2 and 3.2 dB per dB increase of primary intensities between 60 and 70 dB SPL (example in Fig. 7, bottom, for $f2<6$ kHz, ear G23). Accordingly, postischemic CDT levels remained comparatively high for 70 dB SPL primaries.

The behavior of postischemic CDTs clearly departed from that in fatigue-free frequency intervals in the range marked with dashed vertical lines ($6<f2<13$ kHz in Fig. 7): These lines enclose $f2$ frequencies with a CDT level more than 5 dB below the average level computed within fatigue-free intervals. Within this range, which almost coincides with $[f2_{\min}=7,\ f2_{\max}=14.5\ \text{kHz}]$, CDTs dropped down to within

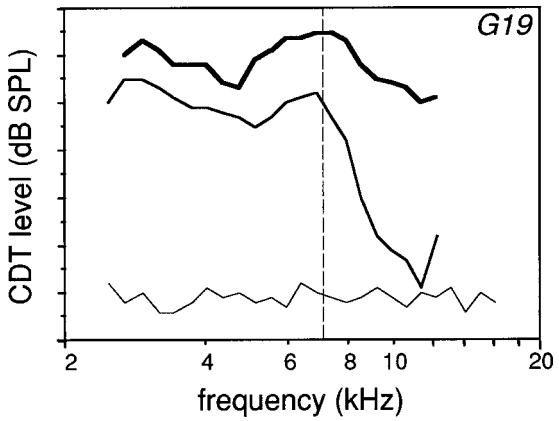


FIG. 8. Other example (in ear G19—see Fig. 1) of DPgram change due complete ischemia following pure-tone overexposure. Primary intensity 70 dB SPL. Bold line: CDTs after exposure and before ischemia onset. Thin line: CDTs 30 min after ischemia onset. Thinner broken line at the bottom: average noise floor. Dashed vertical line: boundary of the frequency range where postischemic CDTs clearly differ from those coming from the fatigue-free range. Notice that according to Fig. 1, the limit of the fatigue-free range in this ear coincides with the dashed vertical line: to the left of this limit, the CDT level is little affected by ischemia; to its right, it is dramatically reduced.

10 dB of the noise floor whatever the primary intensity up to 75 dB SPL (Fig. 7, bottom). Figure 8 depicts another example in ear G19, already presented in Fig. 1 after loud exposure and before ischemia (see frequencies to the right of the dashed vertical line in Fig. 8). In this example, the fatigue-induced preischemic decrease did not exceed 10 dB below 10 kHz: it is too small to account for the large level difference of ischemic CDTs above versus below 7 kHz. For ear G23 (Fig. 7), the cutoffs between large and small ischemic CDTs were found around 6 and 13 kHz whereas auditory fatigue had been present in the range [7, 14.5 kHz]. The ear-probe transducers were not powerful enough to produce a 70 dB SPL primary at $f2>12\,\mathrm{kHz}$, nevertheless CDTs at 60 dB SPL reappeared for $f2>13\,\mathrm{kHz}$. The lower limit $f2_{\min}$ of the frequency range of auditory fatigue and the lower cutoff in the ischemic DPgram coincided within 1 kHz in all ears, while $f2_{\max}$ and the higher cutoff coincided within 1.5 kHz. Although Fig. 7 suggests that high-intensity ischemic CDTs tended to be abolished at slightly lower $f2$'s than those affected by auditory fatigue, thus suggesting a possible 1 kHz tonotopic shift of CDT sources in ischemic cochleas, this trend was not systematic (e.g., it was absent in Fig. 8 versus Fig. 1). The dramatic drop of CDTs in fatigued intervals was clearly seen just after the early notch while the plateau was reached in fatigue-free intervals (e.g., the DPgrams in Fig. 7 were recorded 10 min after ischemia onset) and remained unchanged after the late notch was observed in fatigue-free intervals (Fig. 8, recorded 30 min after ischemia onset).

IV. DISCUSSION

Following Gold (1948), it is now widely accepted that the BM motion in response to a low-intensity tone is enhanced by 40–60 dB in a frequency-selective manner by the operation of a feedback loop (review in Patuzzi, 1996). As the OHCs are thought to be its key element, OHC damage is expected to result in a 40–60 dB loss of sensitivity, as confirmed experimentally (Liberman and Dodds, 1984). The loop-like organization of cochlear processes implies that a strong coupling exists between the main characteristics of cochlear mechanics. High sensitivity comes with sharp bandpass filtering, i.e., high Q factors in the frequency domain and long response delays in the time domain. The notable compressive behavior of the BM growth function in the range of intermediate stimulus intensities (Yates, 1990; Ruggero, 1992) results from the loop-induced enhancement of vibrations gradually losing its efficiency when the tone intensity increases up to about 80 dB SPL. Intermodulation-product otoacoustic emissions are another necessary consequence of the OHC loop because it normally includes nonlinear elements. In particular, the forward mechano-electrical transduction step relates transduction currents to stereocilia deflection through a nonlinear function (Patuzzi *et al.*, 1989b; Kros *et al.*, 1992; Jaramillo *et al.*, 1993; Withnell and Yates, 1998). Thus, the presence of CDTs cannot be separated from the cochlear tuning and its consequences. If one of them happened to disappear or exhibit abnormally poor characteristics, all the other related ones would be expected to decay accordingly. Conversely, any discrepancy, e.g., the presence of normal CDTs despite increased auditory thresholds, would strongly suggest that the OHC feedback loop is not primarily involved and would rather point to IHC or VIIIth nerve damage (e.g., Bronx waltzer mice, Horner *et al.*, 1985; auditory neuropathy, Starr *et al.*, 1996).

A proper use of this framework is essential for CDTs to be clinically useful. Evidence abounds that an ischemic cochlea loses its sensitivity, tuning, and compressive growth function within minutes. Hence the presence of CDTs after ischemia or asphyxia, first described by Schmiedt and Adams in 1981, is particularly disturbing (Norton *et al.*, 1991; Mom *et al.*, 1997) and prompted many authors to caution the clinicians not to use the combinations of stimulus parameters favoring the presence of such CDTs (especially, primary intensities above 60 dB SPL). As this caveat does not explain the origin of unexpected CDTs and provides no rationale for choosing a really conservative stimulus intensity, systematic investigations are needed in order to understand their origin. The most widely accepted explanation so far has relied upon a two-component interpretation (Mills and Rubel, 1994; Mills, 1997) assuming that residual CDTs after disruption of the fine cochlear tuning were actually due to ''passive,'' poorly tuned or macromechanical phenomena. These ''passive'' CDTs would be smaller than the ''active'' ones when the cochlea is healthy and for low-intensity primaries so that the overall CDT properties would be chiefly determined by those of the ''active'' contribution. High-intensity primaries would elicit a mixture of the two contributions, because the ''passive'' component is expected to grow fast and catch up with the ''active'' one. As the ''active'' component is vulnerable, only the high-intensity ''passive'' CDTs would persist in the case of disrupted OHC function. If the two contributions are out of phase and happen to have the same level, mutual cancellation occurs, thereby possibly explaining the notches observed in CDT growth functions and the nonmonotonic time course of CDT levels despite the probably

steady decay of the ''active'' contribution. It is thus essential that a phase reversal of CDTs be observed when a notch occurs for the two-component model to explain the notch (Mills and Rubel, 1994).

The ''active''/''passive'' model of Mills (1997) uses only one mechanism of generation for the two CDTs, assuming that the ''passive'' component comprises the less vulnerable emissions basal to the peak of the $f2$ wave. The possibility that CDTs come from basal places was also evoked by Martin *et al.* (1998) with suppression experiments and Avan *et al.* (1998) with intracochlear pressure measurements. More general two-component models may assume the existence of two different nonlinear mechanisms involving not only different places, but also different anatomical elements such as the BM or inner hair cells (IHCs). The role of IHCs in CDT generation in impaired cochleas has been ruled out only recently by Trautwein *et al.* (1996). Here we use the term ''*two-component* model'' to contrast it with the *two-source* model of Brown *et al.* (1996) and Talmadge *et al.* (1999), who showed that the place tuned to $2f1-f2$ could contribute significantly to the overall CDT as the CDT initially produced by nonlinear interaction in the $f2$ area can propagate forward to the place tuned to its frequency where it is linearly reflected backward (Shera and Guinan, 1999). The stimulus characteristics in the present work were not particularly favorable for revealing the $2f1-f2$ source, furthermore ischemia most likely affects it as much as the $f2$ ''active'' source as all frequencies are uniformly impaired, thus the two-source issue was not addressed here.

Two questions have to be answered in order to validate the two-component model in the case of ischemia: If high-intensity CDTs persist, where does the ''passive'' component come from and what is its mechanism? There is cogent evidence that high-intensity CDTs came from within the cochlea. They were vulnerable since they did decrease within a few seconds and disappeared when the cochlear ischemia was prolonged more than an hour, in keeping with Schmiedt and Adams (1981). Although shorter than in a healthy cochlea, CDT group delays were far longer than instrumental ones. Besides, histologically controlled destruction of OHCs, e.g., by prolonged exposure to noise (Hamernik *et al.*, 1996), administration of combined ototoxic drugs (Lonsbury-Martin *et al.*, 1993) or progression of a genetic disease (Le Calvez *et al.*, 1998) has been shown to lead to complete loss of CDTs including those elicited by primary intensities up to 80 dB SPL. Frolenkov *et al.* (1998) have shown that complete disruption of OHC electromotility abolishes CDTs generated by high primary intensities. All these findings suggest, as stated by Lonsbury-Martin *et al.* (1993), that the high-intensity CDTs require OHC integrity instead of being produced at a macromechanical level such as the BM.

The first issue can thus be reformulated as follows: Where in the cochlea can high-intensity CDTs be produced? In the two-component model of Mills (1997), ''passive'' CDTs come from a broad cochlear region basal to $f2$. The finding of shortened CDT group delays might support this idea. The two-component model also requires a specific phase relationship between ''active'' and ''passive'' components if a sharp CDT level notch were to represent an ''active''-to-''passive'' transition. The early notches observed here for intermediate primary intensities (60 dB SPL) were very sharp when they existed, but went with too small a phase shift for the forthcoming CDT level plateau to be consistent with a ''passive,'' out-of-phase residual contribution. Mills *et al.* (1993) hypothesized that some OHC adaptation could allow ''active'' CDTs to recover with a persisting low endocochlear potential, and the residual CDT component after an early notch might well be an ''active,'' adapted one instead of being ''passive.'' The late notches did coincide with a phase reversal, however the two hypothetical ''active'' and ''passive'' contributions were found to exhibit similar slow decays with time suggesting the same vulnerability, all the more since they had the same group delay. It is noteworthy that alternative explanations of CDT phase reversals not relying upon the existence of two interfering components have been recently set forth (Lukashkin and Russell, 1999): the asymmetry of the distortion-generating characteristic function relating current through stereocilia to deflection and the position of its operating point are assumed to account for phase reversals and level notches (also see Frank and Koessl, 1996).

The way mild auditory fatigue modified the subsequent response of CDTs to complete ischemia provides even stronger arguments against the concept of ''passive'' ischemic CDTs. The consequences on cochlear micromechanics of auditory fatigue due to mild pure-tone exposure are known to be limited and reversible. That the present exposures were mild (90 dB SPL, 15–30 min) was attested to by the small level reduction of CAP threshold shifts and CDT levels, barely exceeding 20 dB at the most affected frequency. Such an auditory fatigue is most likely due to a loss of sensitivity of mechano-electrical transduction channels after their ability has been decreased to gate potassium ions when the stereocilia are deflected (Patuzzi, 1998). It is widely held that DPgrams plotted at moderate primary intensities accurately ''map'' the changes in cochlear function (Martin *et al.*, 1990; Puel *et al.*, 1995; Le Calvez *et al.*, 1998). Accordingly, postexposure preischemic DPgrams showed that only the OHCs tuned to about half an octave above exposure frequency had been affected by auditory fatigue in keeping with the classical half-octave-shift rule (Davis, 1983; McFadden, 1983). Its widely accepted interpretation is that the peak in the envelope of BM responses to a tone at f_{expo} shifts basally when the tone gets louder so that the most fatigued places, near the peak, are tuned to a frequency $>f_{\text{expo}}$: this is where the sharply tuned responses elicited at low and intermediate intensities lose sensitivity and where the related intermodulation nonlinearities diminish.

In the present experiment, stimuli with $f2$ outside the fatigued range generated the same large high-intensity CDTs as in the absence of preliminary sound exposure. Conversely, when $f2$ fell in the range where CDT levels had been affected by sound exposure, little or no CDTs were generated even at 70 dB SPL. Postischemic high-intensity CDTs were therefore highly vulnerable to pre-existing mild auditory fatigue. A first inference is that such a vulnerability can hardly be associated to a ''passive'' property. If a ''passive'' structure were able to generate CDTs, for instance the BM when

it does not benefit from OHC activity, there is no reason why it would stop generating high-intensity CDTs just because the function of OHCs on top has temporarily diminished. The most parsimonious interpretation is that the vulnerable ischemic CDTs were produced by the same elements that suffered from auditory fatigue. As the only structures known to be that sensitive to auditory fatigue are the OHC stereocilia bundles (Patuzzi, 1998), it is tempting to equate them with the sources of postischemic high-intensity CDTs, all the more since hair cell bundles present a conspicuous nonlinearity (Jaramillo *et al.*, 1993). This nonlinearity may relate to mechanical properties since the stiffness of stereocilia is greater for excitatory than inhibitory deflection (Flock and Strelioff, 1984) and is reflected in the Boltzmann-shaped characteristics of gating processes (Patuzzi *et al.*, 1989b; Kros *et al.*, 1992). Other stages of OHC function are less likely to generate intermodulation, e.g., OHC electromotility seems rather linear (Wit *et al.*, 1996). Finally, since postischemic high-intensity CDTs were affected by auditory fatigue for $f2$ falling in the range closely obeying the regular tonotopic rules of active mechanisms, they cannot have come from places basal or apical to $f2$.

The possibility for an ischemic cochlea to emit CDTs via the same mechanisms as a normal cochlea is in fact not surprising. First, an ischemic cochlea is not a ''dead'' one in that some electrochemical gradient remains temporarily present at the apical end of OHCs. Previous reports have shown that in an anoxic cochlea, the endocochlear potential (EP) recorded in scala media very rapidly drops to about -25 to -40 mV (Konishi, 1979; Schmiedt and Adams, 1981). In the meantime, the intracellular potential of OHCs, V_{OHC}, likely remains close to its normal value for a while. Neither EP nor V_{OHC} could be measured directly here. However, $(\mathrm{EP}-V_{\mathrm{OHC}})$ could be indirectly derived from the suggestion by Patuzzi *et al.* (1989a) that the dominant contribution to the round-window CM (here, at $f1=6.6$ kHz) comes from basalmost places close to the electrode. As these places are tuned to frequencies well above $f1$, the cochlear feedback loop is too far from resonance to operate efficiently and influence CM values. This is even more exact in an ischemic cochlea where the OHC feedback loses its effectiveness at all frequencies. In the present case, CM should just be proportional to the product of K^+-channel conductance through stereocilia bundles, times the electrochemical gradient EP $-V_{\mathrm{OHC}}$. As a result, the ratio of CM before versus after ischemia should be given by

$$\mathrm{CM}_{\mathrm{normal}}/\mathrm{CM}_{\mathrm{after}}=(\mathrm{EP}-V_{\mathrm{OHC}})_{\mathrm{normal}}/(\mathrm{EP}-V_{\mathrm{OHC}})_{\mathrm{after}}.$$

The measured CM values suggest that $(\mathrm{EP}-V_{\mathrm{OHC}})_{\mathrm{after}}$ was decreased by a factor of 5–10, and assuming $(\mathrm{EP}-V_{\mathrm{OHC}})_{\mathrm{normal}}$ to be about 150 mV, this lines up with the aforementioned estimations of $\mathrm{EP}_{\mathrm{after}}\approx-40$ mV and $V_{\mathrm{OHCafter}}\approx-70$ mV.

The sequence of events in the OHC feedback loop is well acknowledged (Patuzzi *et al.*, 1989b). First, sound-induced deflection of stereocilia opens their ion channels. $(\mathrm{EP}-V_{\mathrm{OHC}})$ provides for the electrical drive and a K^+ current flows into OHCs, thereby generating an alternative receptor potential. This mechano-electrical stage is strongly nonlinear and vulnerable to stereocilia damage. In turn, the K^+ current generates an ac membrane potential to which motors in the lateral plasma membrane of OHCs respond by force generation. This electro-mechanical stage works *in vitro* and requires only an intact plasma membrane and some voltage drive (Huang and Santos-Sacchi, 1994). The persistent electrochemical gradient for K^+ was certainly sufficient here to fulfill this function. The motile response of OHCs adds up to the stimulus in a frequency-dependent manner and contributes to cochlear tuning, low-intensity sensitivity, and compressive behavior at intermediate intensities (review in Ruggero, 1992).

Assume that CDTs primarily originate from the mechano-electrical stage, although they probably also require other steps in the feedback loop to come out as otoacoustic emissions. Ischemia alone has no reason to disrupt the mechanoelectrical stage immediately and indeed Mom *et al.* (1997) have shown that CDTs can fully recover when CBF is let back to normal after more than 5 min of ischemia. Instead, the main effect of ischemia was to reduce $(\mathrm{EP}-V_{\mathrm{OHC}})$ to 10%–20% of its normal value, while some degree of mechano-electrical transduction through OHC stereocilia remained possible, as well as some degree of electromechanical function. We contend that it must have been enough to maintain the usual generation processes of CDTs, though at a decreased rate.

Nonetheless, the residual OHC feedback loop undoubtedly failed to achieve its main purpose. Loop systems are very sensitive to the amount of feedback and it is likely that a drop of $(\mathrm{EP}-V_{\mathrm{OHC}})$ to $<20\%$ of its normal value led to a dramatic gain decrease. In such situations, irrespective of their physiopathological cause, BM low-intensity sensitivity and intermediate-intensity compression are largely lost as shown, e.g., by Sellick *et al.* (1982) and Ruggero (1992) or predicted by Patuzzi *et al.* (1989b). The BM growth function becomes linear and its slope gets closer to 1 dB/dB at all intensities including the intermediate ones. Around 80–90 dB SPL, the difference between normal and pathological BM responses is small while it increases when stimulus intensity gets lower to reach 40–60 dB below 30 dB SPL (experimental examples in Sellick *et al.*, 1982, and Ruggero and Rich, 1991). These data allow CDT level changes due to ischemia to be guessed. If the alleged nonlinear element was not immediately altered in line with the idea that only the battery $(\mathrm{EP}-V_{\mathrm{OHC}})$ got flat, the distortion level only depended on the *actual* input to the nonlinear element. It is important to notice that this input is *not* the sound level in the ear canal, instead, it is proportional to the BM movement (e.g., it may be the stereocilia deflection; furthermore, because the system is a loop, the input to stereocilia is also the output from the loop). For stimulus intensities >80 dB SPL, published BM data show that this input is almost unchanged in a pathological cochlea, thus the CDT must remain almost constant (it would not hold true if the nonlinear element were affected by the pathology). For smaller stimulus intensities in a pathological cochlea, the input to the nonlinear element decreases by 1 dB per dB decrease in the ear canal. Because the CDT is a third-order combination tone, its level is then expected to decrease by 3 dB/dB decrease of the input. The 2.5 dB/dB

observed here may indicate some residual BM compression (Mills, 1997), thus some residual activity of the cochlear feedback loop. Alternatively, ischemia might have induced enough ionic disturbances or enough adaptation (Mills *et al.*, 1993) to alter the characteristics of the nonlinear element involved in CDT generation in such a way that CDTs at intermediate stimulus intensities would have been larger than those predicted by the 3 dB/dB theoretical law. This could account for the emergence of the plateau in CDT level plots following the early notch. Anyway, the much steeper growth of residual high-intensity CDTs partially made up for the impairment of the cochlear loop due to EP decrease while low-intensity CDTs were dramatically affected.

V. CONCLUSION

Far from ''passive'' and robust, residual ischemic CDTs proved to be highly vulnerable to mild auditory fatigue applied prior to ischemia. It strongly suggests that the generation of ischemic CDTs crucially depends on the integrity of mechano-electrical transduction channels. Since the pattern of absence of high-intensity CDTs closely resembled that of auditory fatigue, it is likely that when they existed, high-intensity CDTs were locally produced following the usual tonotopy. A one-component model of CDT generation (Lukashkin and Russell, 1999) seems more parsimonious that a two-component one (Mills, 1997) to account for these results.

Early stages of cochlear ischemia and presumably euthanasia provide deceptive models of a ''passive'' cochlea as far as distortion is concerned. As some events leading to CDT generation resist some cochlear insults, it would not be conservative in a screening procedure to use stimuli favoring the generation of CDTs from impaired cochleas. Diagnostic purposes are different. That high-intensity CDTs provide a reliable means to map functioning mechano-electrical transduction channels remains to be ascertained in pathological models other than selective cochlear ischemia but if it were, the protocols using comparatively high primary intensities (Kemp, 1992) would be particularly useful.

ACKNOWLEDGMENTS

The authors gratefully acknowledge D.M. Mills for his suggestion to use systematic CDT phase measurements in order to search for possible contributions of out-of-phase components over the course of acute experimental pathology.

Allen, J. B. (**1990**). ''User manual for the Cubdis distortion product measurement system,'' AT&T Bell Labs.

Avan, P., Magnan, P., Smurzynski, J., Probst, R., and Dancer, A. (**1998**). ''Direct evidence of cubic difference tone propagation by intracochlear acoustic pressure measurements in the guinea-pig,'' Eur. J. Neurosci. **10**, 1764–1770.

Billett, T. E., Thorne, P. R., and Gavin, J. B. (**1989**). ''The nature and progression of injury in the organ of Corti during ischemia,'' Hear. Res. **41**, 189–197.

Brown, A. M., Harris, F. P., and Beveridge, H. A. (**1996**). ''Two sources of acoustic distortion products from the human cochlea,'' J. Acoust. Soc. Am. **100**, 3260–3267.

Brown, A. M., and Kemp, D. T. (**1984**). ''Suppressibility of the $2f1-f2$ stimulated otoacoustic emissions in gerbil and man,'' Hear. Res. **13**, 29–37.

Brownell, W. E., Bader, C. R., Bertrand, D., and de Ribaupierre, Y. (**1985**). ''Evoked mechanical responses of isolated cochlear outer hair cells,'' Science **227**, 194–196.

Cody, A. R., and Johnstone, B. M. (**1981**). ''Acoustic trauma: Single neuron basis for the 'half octave shift,'" J. Acoust. Soc. Am. **70**, 707–711.

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,'' Hear. Res. **9**, 79–90.

Fahey, P. F., and Allen, J. B. (**1997**). ''Measurement of distortion product phase in the ear canal of the cat,'' J. Acoust. Soc. Am. **102**, 2880–2891.

Flock, A., and Strelioff, D. (**1984**). ''Graded and nonlinear mechanical properties of sensory hairs in the mammalian hearing organ,'' Nature (London) **310**, 597–599.

Frank, G., and Koessl, M. (**1996**). ''The acoustic two-tone distortions $2f1-f2$ and $f2-f1$ and their possible relation to changes in the operating point of the cochlear amplifier,'' Hear. Res. **98**, 104–115.

Frolenkov, G. I., Belyantseva, I. A., Kurc, M., Mastroianni, M. A., and Kachar, B. (**1998**). ''Cochlear outer hair cell electromotility can provide force for both low and high intensity distortion product otoacoustic emissions,'' Hear. Res. **126**, 67–74.

Gold, T. (**1948**). ''Hearing II. The physical basis for the action of the cochlea,'' Proc. R. Soc. London, Ser. B **135**, 492–498.

Hamernik, R. P., Ahroon, W. A., and Lei, S. F. (**1996**). ''The cubic distortion product otoacoustic emissions from the normal and noise-damaged chinchilla cochlea,'' J. Acoust. Soc. Am. **100**, 1003–1012.

Horner, K. C., Lenoir, M., and Bock, G. R. (**1985**). ''Distortion product otoacoustic emissions in hearing-impaired mutant mice,'' J. Acoust. Soc. Am. **78**, 1603–1611.

Huang, G., and Santos-Sacchi, J. (**1994**). ''Motility voltage sensor of the outer hair cell resides within the lateral plasma membrane,'' Proc. Natl. Acad. Sci. U.S.A. **91**, 12268–12272.

Jaramillo, F., Markin, V. S., and Hudspeth, A. J. (**1993**). ''Auditory illusions and the single hair cell,'' Nature (London) **364**, 527–529.

Kemp, D. T. (**1992**). *User Manual for the ILO92 Otodynamics Equipment* (Otodynamics, London).

Kemp, D. T., Ryan, S., and Bray, P. (**1990**). ''A guide to the effective use of otoacoustic emissions,'' Ear Hear. **11**, 93–105.

Khanna, S. M., and Leonard, D. G. (**1982**). ''Basilar membrane tuning in the cat cochlea,'' Science **215**, 305–306.

Kim, D. O., Molnar, C. E., and Matthews, J. W. (**1980**). ''Cochlear mechanics: Nonlinear behavior in two-tone responses as reflected in cochlear nerve-fiber responses and in ear-canal sound pressure,'' J. Acoust. Soc. Am. **67**, 1704–1721.

Kimberley, B. P., Brown, D. K., and Eggermont, J. J. (**1993**). ''Measuring human cochlear traveling wave delay using distortion product emission phase responses,'' J. Acoust. Soc. Am. **94**, 1343–1350.

Konishi, T. (**1979**). ''Some observations on negative endocochlear potential during anoxia,'' Acta Oto-Laryngol. **87**, 506–516.

Kros, C. J., Rusch, A., and Richardson, G. P. (**1992**). ''Mechano-electrical transducer currents in hair cells of the cultured neonatal mouse cochlea,'' Proc. R. Soc. London, Ser. B **249**, 185–193.

Le Calvez, S., Avan, P., Gilain, L., and Romand, R. (**1998**). ''CD1 hearing-impaired mice. I. Distortion product otoacoustic emission levels, cochlear function and morphology,'' Hear. Res. **120**, 37–50.

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., Whitehead, M. L., and Martin, G. K. (**1993**). ''Distortion-product otoacoustic emissions in normal and impaired ears: Insight into generation processes,'' Prog. Brain Res. **97**, 77–90.

Lukashkin, A. N., and Russell, I. J. (**1999**). ''Analysis of the $f2-f1$ and $2f1-f2$ distortion components generated by the hair cell mechanoelectrical transducer: Dependence on the amplitudes of the primaries and feedback gain,'' J. Acoust. Soc. Am. **106**, 2661–2668.

Martin, G. K., Jassir, D., Stagner, B. B., Whitehead, M. L., and Lonsbury-Martin, B. L. (**1998**). ''Locus of generation for the $2f1-f2$ vs $2f2-f1$ distortion product otoacoustic emissions in normal-hearing humans revealed by suppression tuning, onset latencies and amplitude correlations,'' J. Acoust. Soc. Am. **103**, 1957–1971.

Martin, G. K., Lonsbury-Martin, B. L., Probst, R., Scheinin, S. A., and Coats, A. C. (**1987**). ''Acoustic distortion products in rabbit ear canal. II. Sites of origin revealed by suppression contours and pure-tone exposures,'' Hear. Res. **28**, 191–208.

Martin, G. K., Ohlms, L. A., Franklin, D. J., Harris, F. P., and Lonsbury-Martin, B. L. (**1990**). ''Distortion product emissions in humans. III. Influence of sensorineural hearing loss,'' Ann. Otol. Rhinol. Laryngol. **99**, 30–42.

McFadden, D. (**1983**). ''Intense sounds may alter the mechanical properties of the cochlear partition,'' J. Acoust. Soc. Am. **74**, 447–455.

Mills, D. M. (**1997**). ''Interpretation of distortion product otoacoustic emissions measurements. I. Two stimulus tones,'' J. Acoust. Soc. Am. **102**, 413–429.

Mills, D. M., Norton, S. J., and Rubel, E. W. (**1993**). ''Vulnerability and adaptation of distortion product otoacoustic emissions to endocochlear potential variation,'' J. Acoust. Soc. Am. **94**, 2108–2122.

Mills, D. M., and Rubel, E. W. (**1994**). ''Variation of distortion product otoacoustic emissions with furosemide injection,'' Hear. Res. **77**, 183–199.

Mom, T., Avan, P., Bonfils, P., and Gilain, L. (**1999**). ''A model of cochlear function assessment during reversible ischemia in the Mongolian gerbil,'' Brain Res. Prot. **4**, 249–257.

Mom, T., Avan, P., Romand, R., and Gilain, L. (**1997**). ''Monitoring of functional changes after transient ischemia in gerbil cochlea,'' Brain Res. **751**, 20–30.

Norton, S. J., Bargones, J. Y., and Rubel, E. W. (**1991**). ''Development of otoacoustic emissions in gerbil: Evidence for micromechanical changes underlying development of the place code,'' Hear. Res. **51**, 73–91.

O Mahoney, C. F., and Kemp, D. T. (**1995**). ''Distortion product otoacoustic emission delay measurement in human ears,'' J. Acoust. Soc. Am. **97**, 3721–3735.

Patuzzi, R. B. (**1996**). ''Cochlear micromechanics and macromechanics,'' in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer, New York), pp. 186–257.

Patuzzi, R. B. (**1998**). ''A four state kinetic model of the temporary threshold shift after loud sound based on inactivation of hair cell transduction channels,'' Hear. Res. **125**, 39–70.

Patuzzi, R. B., Johnstone, B. M., and Sellick, P. M. (**1984**). ''The alteration of the vibration of the basilar membrane produced by loud sound,'' Hear. Res. **13**, 99–100.

Patuzzi, R. B., Yates, G. K., and Johnstone, B. M. (**1989a**). ''The origin of the low-frequency microphonic in the first cochlear turn of guinea pig,'' Hear. Res. **39**, 177–188.

Patuzzi, R. B., Yates, G. K., and Johnstone, B. M. (**1989b**). ''Outer hair cell receptor and sensorineural hearing loss,'' Hear. Res. **42**, 47–72.

Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). ''A review of otoacoustic emissions,'' J. Acoust. Soc. Am. **89**, 2027–2067.

Puel, J. L., Durrieu, J. Ph., Rebillard, G., Vidal, D., Assié, R., and Uziel, A. (**1995**). ''Comparison between auditory brainstem responses and distortion products otoacoustic emissions after temporary threshold shift in guinea pig,'' Acta Acust. (Beijing) **3**, 75–82.

Robles, L., Ruggero, M. A., and Rich, N. C. (**1991**). ''Two-tone distortion in the basilar membrane of the cochlea,'' Nature (London) **349**, 413–414.

Ruggero, M. A. (**1992**). ''Responses to sound of the basilar membrane of the mammalian cochlea,'' Curr. Opin. Neurobiol. **2**, 449–456.

Ruggero, M. A., and Rich, N. C. (**1991**). ''Furosemide alters organ of Corti mechanics: evidence for feedback of outer hair cells upon the basilar membrane,'' J. Neurosci. **11**, 1057–1067.

Schmiedt, R. A., and Adams, J. C. (**1981**). ''Stimulated acoustic emissions in the ear canal of the gerbil,'' Hear. Res. **5**, 295–305.

Sellick, P. M., Patuzzi, R. B., and Johnstone, B. M. (**1982**). ''Measurement of basilar membrane motion in the guinea pig using the Mossbauer technique,'' J. Acoust. Soc. Am. **72**, 131–141.

Shera, C., and Guinan, Jr., J. J. (**1999**). ''Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs,'' J. Acoust. Soc. Am. **105**, 782–798.

Starr, A., Picton, T. W., Sininger, Y., Hood, L. J., and Berlin, C. I. (**1996**). ''Auditory neuropathy,'' Brain **119**, 741–753.

Talmadge, C. L., Long, G. R., Tubis, A., and Dhar, S. (**1999**). ''Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **105**, 275–292.

Trautwein, P., Hofstetter, P., Wang, J., Salvi, R., and Nostrant, A. (**1996**). ''Selective inner hair cell loss does not alter distortion product otoacoustic emissions,'' Hear. Res. **96**, 71–82.

Whitehead, M. L., Lonsbury-Martin, B. L., and Martin, G. K. (**1992a**). ''Evidence for two discrete sources of $2f1-f2$ distortion-product otoacoustic emission in rabbit. I. Differential dependence on stimulus parameters,'' J. Acoust. Soc. Am. **91**, 1587–1607.

Whitehead, M. L., Lonsbury-Martin, B. L., and Martin, G. K. (**1992b**). ''Evidence for two discrete sources of $2f1-f2$ distortion-product otoacoustic emission in rabbit. II. Differential physiological vulnerability,'' J. Acoust. Soc. Am. **92**, 2662–2682.

Wit, H. P., van Dijk, P., and Segenhout, H. M. (**1996**). ''Frequency response for electromotility of isolated outer hair cells of the guinea pig,'' Hear. Res. **98**, 165–168.

Withnell, R. H., and Yates, G. K. (**1998**). ''Onset of basilar membrane nonlinearity reflected in cubic distortion tone input-output functions,'' Hear. Res. **123**, 87–96.

Yates, G. K. (**1990**). ''Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions,'' Hear. Res. **50**, 145–162.

A level of stimulus representation model for auditory detection and attention

Ervin R. Hafter[a] and Kourosh Saberi[b]
Department of Psychology, University of California, 3210 Tolman Hall, Berkeley, California 94720

(Received 21 May 1998; revised 29 May 2001; accepted 1 June 2001)

A model is offered here to address an asymmetry of cueing in signal detection [Hafter *et al.* (1992)] where the effect of frequency uncertainty on the detection of a randomly chosen tone was ameliorated by cueing with a sequence of its harmonics, but detection of a randomly chosen sequence of harmonics was not improved by cueing with their fundamental. The model proposes that signal detection can be based on various levels of neural representation that, for the case at hand, refer to levels organized either by frequency or by complex pitch. Experiments offered to test the model used three-tone complexes for both cues and signals. These stimuli consisted of either three randomly chosen frequencies or three randomly chosen harmonics (from the set $2f_1$ to $7f_1$) of a randomly chosen fundamental. Support for the idea of cueing and detection at different levels of representation was found in higher performance with uncued detection of harmonic complexes relative to that found with complexes of unrelated tones and by successful cueing of each type of information with cues created to remove uncertainty about the relevant information. A final comparison suggests independence of performance (presumably of the limiting noise) at each of the putative levels of representation.
[DOI: 10.1121/1.1394220]

PACS numbers: 43.66.Hg, 43.66.Dc, 43.66.Lj, 43.66.Fe [JWH]

I. INTRODUCTION

A. Uncertainty reduces performance

The peripheral auditory system is generally characterized as a bank of bandpass filters called critical bands, or auditory filters, whose widths are roughly proportional to their center frequencies. Because they derive from neuromechanical processes in the cochlea, these filters are commonly thought to be immutable in shape and bandwidth. However, the possibility of top-down control over their *effective* bandwidths has been discussed in speculations on the role of efferent neural connections to the cochlea (Scharf *et al.*, 1994, 1997), proposed as the basis for interactions between motivational instructions and the effects of frequency uncertainty in signal detection (Hafter and Kaplan, 1976), and observed directly in a study of uncertainty using a probe-signal method (Schlauch and Hafter, 1991).

Traditional psychoacoustic descriptions of the auditory filters have relied upon studies of the detection of pure-tone signals in the presence of noise. An indirect approach infers the bandwidths from so-called ''critical ratios'' defined as the signal level at threshold divided by the spectrum level (level/Hz) of a wideband masker, while more direct methods describe ''critical bandwidths'' in terms of the relation between performance and the bandwidths of either band-limited maskers or spectral notches in a wideband masker. (For review, see Scharf, 1970; Patterson and Moore, 1986; Moore, 1997.) In all such measures, it is tacitly assumed that the subject monitors and responds only to frequencies falling within the appropriate auditory filter which, in turn, means that he or she *knows* the signal's frequency. If it is unknown, performance must decline, even for the ''ideal observer'' of signal detection theory (SDT) (Green and Swets, 1966) due to the increased probability of large peaks in the noise appearing in wrong, i.e., nonsignal filters. Studies where the signal's frequency has been drawn at random on each trial from a list of $M>1$ possibilities have found that the masked threshold relative, to the case of $M=1$, rises to an asymptotic value of 3 to 5 dB for large values of M (e.g., Green, 1961; Schlauch and Hafter, 1991).

B. Reducing uncertainty with cues

The deleterious effects of frequency uncertainty on detection can be reduced or even eliminated by presenting pretrial cues that tell the subject what to listen for. The most effective such cue is a tone matched in frequency to the signal (e.g., Swets and Sewall, 1961; Hafter and Kaplan, 1976; Johnson and Hafter, 1980; Schlauch and Hafter, 1991; Dai *et al.*, 1991). While this implies a crucial role for shared phenomenology whereby a cue works because it ''sounds like'' the signal, successful reductions of uncertainty have also been found with a variety of cues that are not identical to the signal. These include a tone whose frequency relates to the signal by a small integer ratio such as 5/4 (the musical third) (Hafter and Kaplan, 1976) or 3/2 (a musical fifth) (Hafter *et al.*, 1993), a melodic sequence of tones for which the signal is a musically acceptable extension of the melody (Howard *et al.*, 1984, 1986), a chord made up of harmonic frequencies whose missing fundamental is the same frequency as the signal (Hafter *et al.*, 1992), a multi-tonal complex made up of randomly chosen frequencies, one of which matches the signal (Schlauch and Hafter, 1991), and even a

[a]Author to whom correspondence should be addressed. Electronic mail: hafter@socrates.berkeley.edu
[b]Now at: Department of Cognitive Sciences, University of California, Irvine, CA 94720-1650. Electronic mail: kourosh@uci.edu

visual cue that describes the frequency of the signal in musical notation for subjects who have absolute pitch (Plamondon and Hafter, 1990). Thus, the ''sounds-like'' hypothesis must be extended to include comparisons made in the *mind's ear* between similar percepts that arise through different auditory mechanisms.

C. Is "sounds alike" sufficient for cueing?

The present experiment was begun in part because of a result that did not seem readily explained by a sounds-like hypothesis. In this condition, where the stimuli used for cues and signals shared a common percept, their pitch, reduction of the effects of uncertainty was asymmetric, depending on their order of presentation. One stimulus was a single, randomly chosen tone whose frequency was dubbed f_1; the other was a harmonic sequence of either $2f_1 \rightarrow 6f_1$ or $3f_1 \rightarrow 7f_1$. While a pure tone and a set of its harmonics are quite different from one another in timbre, a subject asked to adjust the pitch of a pure tone until it matches the pitch of a set of harmonics typically picks the fundamental frequency (f_1) of the sequence, even when f_1 itself is absent from the complex. The common pitch of tones and complexes is thought to arise through separate auditory mechanisms. ''The former has traditionally been called a *place* pitch in reference to a place of maximum displacement in the cochlear excitation patten, though analysis of periodic activity in auditory neurons shows that information about frequency also exists in the time-domain, even for tones.'' Conversely, the pitch of the latter relies on further analysis of the complex stimulus based on interactions between its components. Variously called a ''residue,'' ''periodicity,'' or ''virtual-pitch'' (e.g., Schouten, 1940; Licklider, 1956; Terhardt, 1974; Terhardt *et al.*, 1982), it has been the object of a long-standing debate about its origin. Whether it is derived in the time or frequency domain is irrelevant to the present discussion and so we will use the more neutral term, *complex* pitch. (Excellent discussions of the mechanisms of pitch can be found in De Boer, 1976; Houtsma, 1995).

As described earlier, Hafter *et al.* (1992) found that a set of harmonics used as a cue improved detection of a randomly chosen tone set to its missing fundamental.[1] However, when those roles were reversed, that is, when randomly chosen f_1's were used to cue signals that were a sequence of their harmonics, there was no improvement due to cueing. We will refer to this dependency on order of presentation as an asymmetry in cueing.

II. A MODEL OF CUEING BASED ON LEVELS OF STIMULUS REPRESENTATION

A. The level of stimulus representation used for detection or LSRD

In seeking to explain the asymmetry of cueing described earlier, this model concentrates less on the shared phenomenology of a ''sounds-like'' hypothesis than on shared elements at the level of stimulus representation or LSR whose neural activity serves as the basis for signal detection. We call this the level of stimulus representation used for detection or LSRD. It is well known that various features of an acoustic stimulus are represented at multiple LSRs throughout the auditory nervous system (Pickles, 1988); for example, tonotopic maps of acoustic frequency have been identified in regions ranging from cochlea to cortex. While band limitations found in detection experiments are often said to result form ''cochlear filters,'' what is generally meant by this is that the filtering began in the cochlea and not that the subject's judgments were based directly on neural activity at the interface of cochlea and auditory nerve. Because it is unlikely that a neural site simply relays information, one assumes that additional processing at each LSR provides a unique version of the original stimulus. By further assuming that the decision process in a signal-detection task has access to many, if not all, of the LSRs, it follows that best performance requires that the LSRD be the level whose data present the highest signal-to-noise ratio (S/N). From this perspective, the optimal subject must use knowledge of the signal's parameters to select both the optimal LSRD and the appropriate elements within it. This is especially important for multidimensional signals whose potential LSRDs represent different stimulus dimensions.

B. Formal assumptions of the LSRD model

(1) Neurons in LSRs are organized topologically in accord with their sensitivity to values along a stimulus dimension. The dimension may be based on a primitive feature of the stimulus such as acoustic frequency or on a more complex feature that derives from interactions between primitives. Examples of the latter include complex pitch and locations in auditory space.

(2) Each neuron in an LSRD is best tuned to a specific value of the represented dimension but responds, to some extent, to nearby values of the dimension falling into its ''receptive field.'' Pooled responses from adjacent receptive fields then determine the effective bandwidth of a masker, providing a kind of filter in the represented dimension.

(3) For signals represented at multiple LSRs, optimal performance requires that the LSRD be the level providing the highest S/N. When signals are represented both by primitive dimensions and by complex interactions between those primitives, the larger S/N will generally be found at the higher-order representation (for a discussion, see Sec. II C).

(4) For maximum effectiveness in reducing signal uncertainty, a cue should specify, unambiguously, both the level being used as the LSRD and the correct filter within that representation.

C. Application of the model to tones and multi-tonal complexes without uncertainty

In the framework of signal detection theory (SDT) (Green and Swets, 1966), each observation is judged in accord with the likelihood (λ) that it arose from signal-plus-noise rather than from noise alone. For detection of a single tone of *known* frequency in a background wideband noise, the optimal decision rule is to compute λ on the basis of

energy within the single auditory filter centered on the frequency of the signal. Thus, in terms of the model, the LSRD used to detect those signals is organized according to frequency. If the frequency of the signal is *unknown*, the optimal subject must calculate a likelihood for every filter in the LSRD that might contain the signal, and performance must decline due to the increased probability of high likelihoods produced by noise alone in nonsignal filters.

For a multi-tonal signal made up of *J known but unrelated* frequencies, the optimal rule is to decide on the basis of the product of the *J* individual likelihood ratios,

$$\lambda = \prod_{i=1}^{J} \lambda_i, \tag{1}$$

based on energies in the appropriate *J* auditory filters. While this rule could also be used for a signal made up of *J* harmonically related frequencies with a known fundamental, an alternative decision rule might be to listen for the signal's complex pitch. In that case, detection would be limited by a pitch-masker which both reduced the effectiveness of signals by adding noise to their individual components and produced false pitches through the accidental occurrence of peaks in the acoustic noise at harmonics of the signal. In terms of the LSRD model, detections would be based on neural activity at an LSRD that is topologically arrayed according to complex pitch. Similarly, the extent to which adjacent elements are pooled at this level would define the effective bandwidth of the pitch masker. Unfortunately, without knowing the statistics of the putative pitch masker, one cannot say whether a *J*-tone harmonic complex with *known* frequencies would be better detected on the basis of the *J* individual frequencies or on its complex pitch.

D. Application of the model to multi-tonal complexes with uncertainty

The situation with *J*-tone signals is quite different with frequency uncertainty. If the individual tones are unrelated, the best LSRD is still one organized by frequency, only here, the optimal statistic, $\Pi\lambda_i$ [Eq. (1)], is based on the *J* highest individual likelihoods found across all auditory filters in the range of possible frequencies. However, when the unknown frequencies in the signal are known to be harmonically related, albeit with no knowledge of their f_1, the superior strategy is to listen for and respond to the emergence of a complex pitch. That is because noise alone can produce a false positive in the pitch domain only if the highest peaks in the acoustic noise happen to fall into filters related by a common fundamental frequency. Thus, signal uncertainty should have less of an effect on performance using a LSRD organized by complex pitch than one organized by frequency.

The fourth assumption of the model was proposed to address the asymmetry of cueing described in Sec. I C. According to the model, a missing-fundamental harmonic was a useful cue for detection of f_1 because it specified the single, appropriate filter for detection at a LSRD organized by frequency. Conversely, cueing was ineffective in the reverse condition because a single tone does not specify, unambiguously, a single complex pitch. Consider, for example, a tonal cue of 600 Hz. If treated as part of a complex pitch, it could thought of as the second harmonic of 300 Hz (an octave), the third harmonic of 200 Hz (an octave plus a just-fifth), or the fifth harmonic of 120 Hz (two octaves plus a just-major third).

E. A test of the LSRD model

While the model offered above would seem to explain the asymmetry of cueing with sets of harmonics and their fundamentals, it arose as a *post hoc* analysis of that result. The present study was planned as a more rigorous test of the model, restricting comparisons to stimuli more similar to one another on features not directly addressed in predictions of the model. Here, all stimuli would be three-tone complexes whose frequencies would be either unrelated to one another, thus offering no complex pitch for detection, or related by a common fundamental. As discussed in Sec. II D, under conditions of signal uncertainty, the model predicts better performance with harmonic complexes than with the randomly chosen tones. In line with assumption 4, the cues would also all be three-tone complexes whose frequencies would be chosen to ameliorate uncertainty about either the individual frequencies in the signal, its complex pitch, or both.

III. PROCEDURE

A. Stimulus generation

The experiment compared performance across five conditions, each of which measured the detectability of three-tone complexes. Frequencies in the signals were different on every trial, but the levels of the three tones were set to be equally detectable through reference to an empirically derived, equal-detectability function (EDF). This function was found by measuring thresholds for the three subjects at five frequencies covering the range from 400 to 4725 Hz. A straight line was fitted to these data in dB/Hz to provide an approximation to an EDF (Green *et al.*, 1959; Schlauch and Hafter, 1991). When the individual differences between the subjects proved to be insignificant, a single, averaged EDF was constructed and used throughout the experiment for all subjects. Signals were generated digitally, with a sampling rate of 50 kHz, and played through a locally constructed 16-bit D/A converter and a low-pass filter with a cutoff frequency of 20 kHz and slope of 48 dB/oct (Frequency Devices Model 901).

The continuous wideband masker was produced by an analog white-noise generator and filtered only by the frequency response of the Stax (SR5) electrostatic headphones. Its spectrum was essentially flat across the frequency range of interest. The spectrum level of the noise was 20 dB SPL as determined with a Hewlett-Packard (3582A) spectrum analyzer.

B. Psychophysical measurement

Performance was measured in a two-interval, forced-choice psychophysical (2IFC) task which presented signals with equal probability in one of two 300-ms intervals. Times between those two intervals as well as between the cue

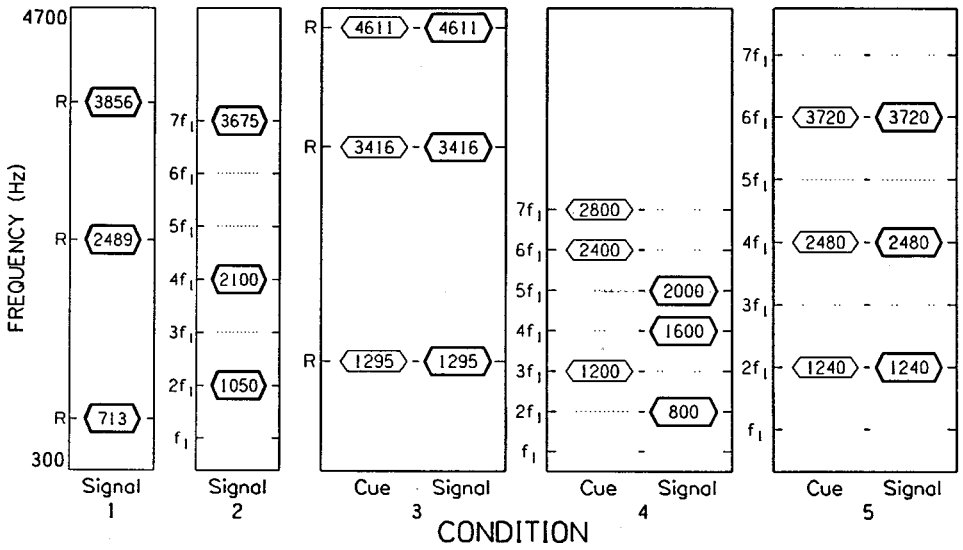


FIG. 1. Time-lines descriptive of five kinds of trials from the present experiment. In condition 1, uncued signals were three randomly chosen tones (see text). In condition 2, uncued signals were three harmonics chosen at random from the set $2f_1 \rightarrow 7f_1$ of a randomly chosen f_1. In condition 3, a random signal as in condition 1 was preceded by a cue matched to it in frequencies. In condition 4, a harmonic signal as in condition 2 was preceded by a cue matched to it in complex pitch. In condition 5, a harmonic signal as in condition 2 was preceded by a cue matched to it in both frequencies and pitch.

(when present) and the first interval were 250 ms. Durations of the cues and signals were 300 ms, including 10-ms linear onset and offset ramps. Each trial was terminated by the subject's response and followed by visual feedback that identified the correct response interval. Proportions of correct responses, $P(C)$'s, were obtained over blocks of 50 trials. Each subject ran at least seven blocks of each condition. Experimental sessions generally consisted of ten blocks, with only a single stimulus condition presented within a block. Before each block, a subject was allowed as many practice trials as he or she wished, although such practice trials never exceed ten. All conditions were practiced until performance seemed stable before the actual experiment began; at that point, the order of conditions was randomized and each subject was tested in a different order. Subjects were students at the University of California, including one of the authors, KS. All reported normal hearing. Testing was done in a double-walled, audiometric listening booth.

C. Signals and cues

A single EDF was used to set levels of cues and signals throughout all five conditions of the experiment. Thus, conditions are compared in units of performance. The five stimulus conditions are described below. In addition, a spectrogramlike depiction of a representative trial from each condition is shown in Fig. 1. In order to save space, the figure is not drawn to scale and interstimulus intervals are omitted. Cues, when used, were 6.3 dB higher than the signals, leaving them weak but clearly audible. To be more specific, each component in a cue was 6.3 dB higher than the level of that same frequency if drawn from the EDF used to generate signals.

1. Condition 1: Random-complex signals: No cues

The idea here was to present useful information only at an LSRD organized by frequency. To this end, each signal consisted of three unrelated tones selected at random from a uniform distribution of frequencies that ranged from 400 to 4725 Hz, with the sole restriction being that the ratio between adjacent frequencies must exceed 1.10. In the example shown in Fig. 1, the three unrelated frequencies are 713, 2489, and 3856 Hz. We assumed that this condition would provide the poorest performance, thus allowing room for improvement in conditions with less uncertainty. For this reason, pretesting was used to pick a signal level that would produce especially weak scores of $P(C) \sim 0.60$, a value well below the $P(C) = 0.75$ conventionally used to define threshold. The EDF so-chosen was anchored at 500 Hz to 33 dB SPL, corresponding to a signal-energy-to-noise-power ratio (E/N_o) of 7.7 dB. To reiterate, this EDF was then used for all five conditions.

A subject in condition 1 could, of course, adopt a nonoptimal strategy that ignored the fact that there were three tones in the signal and respond to magnitudes of the two or even single largest values of λ across the range of frequencies. In order to see if subjects were doing that, informal tests were run during the preexperimental period with signals consisting of either one or two randomly chosen tone(s). Performance was lower with only two tones and still lower with one and, leading us to conclude that subjects in the experiment proper would listen for (at least) three tones.

2. Condition 2: Harmonic-complex signals: No cues

Unlike the case in condition 1, the three tones in these signals bore a simple harmonic relation to one another. At the beginning of each trial, a fundamental frequency (f_1) was selected at random from the range 200 to 675 Hz. The next six harmonics of that fundamental $(2f_1 \rightarrow 7f_1)$ were computed and, from these, three were selected at random to be the signal. Thus, no signal contained f_1, and the harmonic number of components in the signal differed from trial to trial. This procedure ensured a minimum ratio between adjacent components in the signal of 1.17 (7/6). In the example in Fig. 1, the randomly chosen f_1 is 525 Hz and the randomly chosen tones in the signal are $2f_1$ (1050 Hz), $4f1$ (2100 Hz), and $7f_1$ (3675 Hz). While the individual components were represented by increased energy in the three auditory filters, just as in condition 1, this signal also presented information potentially useful for detection on the basis of its complex pitch (525 Hz). The model predicts higher perfor-

mance based on the complex pitches of these signals than on the three frequencies drawn purely at random in condition 1.

3. Condition 3: Random-complex signals: Cues matched to the signal's frequencies

Here, the three frequencies were selected purely at random, as in condition 1, but now each trial began with a three-tone cue made up of the same three frequencies. The example in Fig. 1 shows both cues and signals at 1295, 3416, and 4611 Hz. We call these frequency or *F* cues because, in terms of the model, their effectiveness should indicate specification of the appropriate auditory filters at an LSRD organized by frequency. The prediction here is of performance better than that found in condition 1.

4. Condition 4: Harmonic-complex signals: Cues matched to the signal's pitch

Harmonically related signals here were chosen in the same way as in condition 2, but each trial now began with a cue intended to remove uncertainty about the signal's complex pitch without sharing its elements at a LSR organized by frequency. For this, the cue was a three-tone harmonically related complex with the same fundamental as the signal but composed of different harmonic numbers. Thus, after three of the components $2f_1 \rightarrow 7f_1$ had been designated as the signal, the remaining three made up the cue. In the example in Fig. 1, the randomly chosen f_1 was 400 Hz and the randomly selected components for the signal were $2f_1$ (800 Hz), $4f_1$ (1600 Hz), and $5f_1$ (2000 Hz). The cue was made up of the remaining components, $3f_1$ (1200 Hz), $6f_1$ (2400 Hz), and $7f_1$ (2800 Hz). Again, cues were 6.3 dB higher than signals. We call these pitch or *P* cues because, in terms of the model, their effectiveness would indicate specification of the appropriate pitch filters at a LSRD organized by complex pitch. Thus, the prediction is of higher performance than in condition 2.

5. Condition 5: Harmonic complexes: Cues matched to the signal's frequencies and pitch

Signals here were again harmonic complexes chosen as in conditions 2 and 4. However, the cues were made up of the same three harmonics as the signals. Thus, in the example in Fig. 1, both signals and cues were $2f_1$ (1240 Hz), $4f_1$ (2480 Hz), and $6f_1$ (3720 Hz) of the randomly chosen f_1 (620 Hz). We call these ''bi-dimensional'' cues FP because they presented information about both the individual frequencies in the signal and its complex pitch. The idea was to see if these cues could enhance performance by cueing both the individual frequencies as in condition 3 *and* the complex pitch as in condition 4. If signals at the putative LSRDs were limited by independent noise, informational enhancement produced by the FP cues should be additive. Using $(d')^2$ as the SDT measure of transmitted information (Green and Swets, 1966), the prediction would be that the d' for condition 5 should equal the root-mean-square (rms) value of the d'-values found in conditions 3 and 4.

TABLE I. Summaries of the individual scores as well as averages across the three subjects for the five experimental conditions depicted in Fig. 1. Individual performance in proportion correct, $P(C)$, as well as means across subjects.

Condition	Cues	Signals	Subject 1	Subject 2	Subject 3	$\bar{x}$
C1		Random	0.630	0.597	0.571	0.599
C2		Harmonic	0.702	0.732	0.678	0.704
C3	F	Random	0.764	0.788	0.751	0.763
C4	P	Harmonic	0.827	0.780	0.765	0.791
C5	FP	Harmonic	0.893	0.917	0.913	0.908

IV. RESULTS

Results from the individual subjects as well as the averaged means are shown in Table I. The averaged data are also plotted in Fig. 2 for visualization. The significance of differences between conditions predicted by the model were tested through use of individual *z*-score tests as described in the Appendix.

A. Detection based on complex pitch

Based on the LSRD, we predicted in Sec. II D that subjects should be better at detecting the harmonically related complexes in condition 2 than the unrelated complexes of condition 1. This prediction, C2>C1, proved to be true (see the Appendix).

B. Cueing at the level of frequency

Because cues matched in frequency are highly effective for one-tone signals, one would expect a similar improvement using matched-frequency (*F*) cues with the unrelated three-tone signals. This predicted amelioration of frequency uncertainty, C3>C1, was confirmed (see the Appendix).

C. Cueing at the level of complex pitch

Results discussed in Sec. IV A suggest that signals in condition 2 were detected on the basis of an emergent property of the relation between their components, a complex

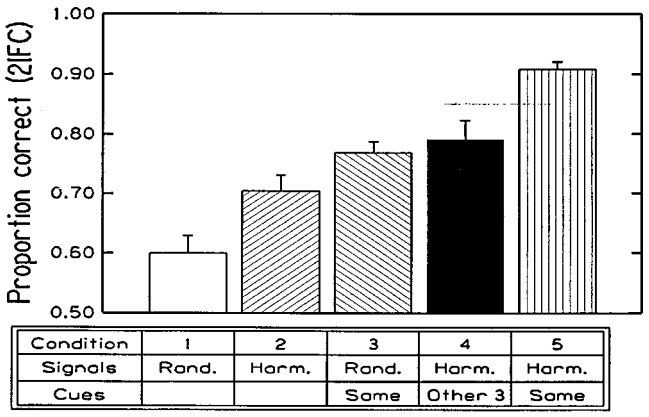


FIG. 2. Mean performance in the five conditions described by Table I and Fig. 1. Data are shown as the proportion of correct responses, averaged across three subjects. Error bars depict the standard error of the means.

TABLE II. Data obtained with bi-level cues as well as predicted results for condition 5 based on summation of the information conveyed in conditions 2 and 4 [see Eq. (1)].

	Subject 1	Subject 2	Subject 3
Obtained $P(C)$	0.89	0.92	0.91
Predicted $P(C)$	0.82	0.86	0.84

pitch. Given this, the model suggests that amelioration of pitch-uncertainty requires use of cues matched in pitch (P). This prediction, C4>C2, was also confirmed (see the Appendix).

D. The effect of bi-dimensional cueing at both LSRDs

Although successful cueing in condition 4 would seem to indicate that P cues worked by alleviating uncertainty about the complex pitch of the signals, the $P(C)$'s were not much different from those found with frequency-matched cues in condition 3. Thus, one could argue that subjects in condition 4 might have used a knowledge of harmonicity to determine the frequency of f_1 from the cue and then calculated the tones in the signal for detection at an LSRD organized by frequency. However, this idea was dispelled by results with the FP cues in condition 5. If one assumes independence of the limiting noise at LSRDs organized by frequency and by complex pitch, performance in condition 5 measured in d' should reflect the rms sum of the d's found with the one-dimensional cues in conditions 3 and 4 (Green and Swets, 1966). In order to test this hypothesis, $P(C)$ values from C3 and C4 were converted to d's using Elliot's tables (in Swets, 1964) and used to predict performance with bi-dimensional cues (FP):

$$d'_{\mathrm{FP}} = \sqrt{[d'_F]^2 + [d'_P]^2}. \tag{2}$$

Predicted values of d'_{FP}, converted back into $P(C)$ for comparisons to the obtained data, are presented in Table II. It shows that the obtained values were actually slightly higher than those predicted from the combination of information in Eq. (2). While hyper-additivity of this type is not predicted by classic SDT, one might speculate that the two sources of information in the FP cues somehow enhanced each other's effectiveness. For example, knowing the pitch might have helped the listener to focus more precisely on the appropriate elements in an LSRD organized by frequency and *vice versa*. Regardless, by showing that performance was as least as good as that predicted by additivity of cued performance at the two putative LSRDs, this lends further support to the hypothesis of independent accessibility to information at separate LOPs organized by frequency and complex pitch. Seeking statistical support for the independence-of-cueing hypothesis, we compared the case where cues carried both the frequencies and the pitch of the signals to the one in which the subjects were cued with pitch alone. In support of independence, C5>C4 also proved to be highly significant (see the Appendix).

V. DISCUSSION

A. Detection based on a complex feature of the signal

We have argued that the special performance found in conditions 2, 4, and 5 support the notion that the subjective decision maker had access to neural data at a neural level that specifically represents complex pitch. While it is difficult to point to direct evidence of LSR topologically organized by complex pitch, the musical perception of a sequential relation between successive notes, e.g., C, C♯, D, D♯, etc., even with missing-fundamental, harmonic complexes, would seem to suggest one. Furthermore, while direct physiological results of such an organization have been scarce, studies of the neural code for amplitude modulation and complex pitch (e.g., Schreiner and Langner, 1988; Langner *et al.*, 1997) lend credence to its existence, and one expects that the advent of new brain-imaging techniques will clarify this important issue in the near future. Interestingly, our prediction of improved detectability of a complex based on the relation between its tones reflects a more general principle, namely that detection of any complex signal under stimulus uncertainty should be better if the judgments are based on the relationship between its primitive components rather than on an independent analysis of its primitives alone.

B. Assumptions of the model

Formal assumptions of the LSRD model were made purposefully strong to simplify predictions for the experiments. While assumptions 1–3 seem well justified by the apparent usefulness of complex pitch as a dimension for detection, the assertion in assumption 4 that the shared representation between cue and signal at the LSRD must be unambiguous was probably overstated. A softer proposal might say that while such representation is necessary for maximally effective cueing, a partial relation between the cue and signal could partially reduce the effects of uncertainty. In support of the softer view, we point to the condition described in footnote 1, where a five-component, missing-fundamental harmonic sequence improved the detectability of a signal set to its fundamental frequency, albeit by not as much as a tone of the same frequency as the signal. While performance with the missing-fundamental cue was significantly higher than with no cue at all [$P(C)=0.835$ *vs* 0.68], it was less than with a single-tone cue set to the same frequency as the signal [$P(C)=0.92$] (Hafter and Schlauch, 1989). Whether this was because the multi-tonal cue pointed to additional filters such at subtharmonics of the fundamental or because it produced a correct but poorly defined representation at the LSRD organized by frequency, we simply do not know.

C. An alternative explanation for cueing at the level of pitch

A reviewer of an earlier submission of this article (Darwin, 1995) pointed out a potential confound in our methodology that might also have produced an improvement from condition 1 to 2. He noted that while purely random frequencies in condition 1 were drawn from a distribution whose

upper boundary was 4725, the harmonically related frequencies in condition 2 were drawn from a distribution whose trial-by-trial upper limit was f_7 of a randomly selected f_1, ranging from 1400 to 4725 Hz. Thus, better performance in condition 2 might simply have reflected less frequency uncertainty. Because of this, the second author recruited a new crew of subjects at the University of Florida to retest the comparison between conditions 1 and 2 when tones in each condition were drawn from the same distributions. To this end, each trial in the revised condition 1 began with the random choice of a separate, range-setting frequency (f_R) from the same 200 to 675-Hz range used to select f_1 in condition 2. Thus, the three unrelated tones for that trial were drawn at random from the range $2f_R$ to $7f_R$, the same as for the related harmonically related tones in condition 2. Concerns that the reduced uncertainty was fully responsible for the results in Fig. 3 were dispelled when thresholds (from a tracking procedure) for the new condition 2 were significantly lower than those for new condition 1 ($p<0.05$)[2].

D. Failure to find successful cueing of a fundamental frequency by a matched complex pitch in "informational masking"

McFadden (1988) did not find what he called ''periodicity cueing'' for the amelioration of ''uncertainty'' in a case where the signal was a pure tone and the cue a four-tone sequence of its harmonics. However, we do not find this incompatible with the cueing reported here. In McFadden's study, the masker was a set of six other tones whose frequencies did not relate in a harmonic fashion to the signal. They were played in a temporal sequence, three before the signal and three after it. Because the masker tones were chosen so as to not affect the auditory filter centered on the signal, his paradigm falls into a class often referred to as informational masking, with uncertainty referring to the order of presentation of the individual tones in the masker. Because Watson and Kelly (1981) had shown that masker uncertainty of this kind depresses performance, McFadden (1988) thought that presenting the harmonically related cue during the signal interval would reduce the effects of uncertainty and thus increase detection. When this did not turn out to be the case, he reasoned that ''...one might conclude that periodicity cueing does exist for sensory masking but not for informational masking.'' We agree. From our perspective, a cue chosen to reduce uncertainty about a signal works by reducing the number of potential signals to be listened for and hence the number of filters that must be monitored. Thus, when the masker changes but the signal is always the same, one should not expect that cueing the signal would have an effect.

E. The effects of differences in timbre between cues and signals

The purpose of condition 4 was to see if signals made up of harmonic complexes would be successfully cued by their complex pitch if there were no actual frequencies in common. On first listening to these stimuli when setting up the experiment, we worried that cueing at the level of complex pitch might be overshadowed by the large differences between the timbres of cues and signals. Less problematic, but still of concern, might be differences between them in pitch or pitch strength, the worst case being the rare occasion when a cue was made up of only even harmonics and the signal only odd. However, these fears proved to groundless when the three-tone complexes were presented in noise because at low S/N, the cues and signals sounded remarkably alike in pitch and timbre, despite the differences in harmonic numbers. This is reminiscent of the one-tone residue reported by Houtgast (1976). He presented three-tone harmonic complexes (not including f_1) as standards and asked subjects to detect small changes in the f_1 of a three-, two- or one-tone test signal whose harmonic numbers differed from those of the standard. He found that while the task could, to a small extent, be done when the tones were presented in quiet, it was much easier if they were heard in a background noise. Indeed, with noise, subjects reported hearing the pitch of the fundamental in the test signal, even when it had only a single harmonic. While pattern matching models of pitch (see Moore, 1997) easily address the commonality of pitch between cues and signals in our condition 4, they do not speak to the similarity of timbres at low S/N. However, this seems less puzzling if one considers that while only three of the harmonically related filters from $2f_1$ to $7f_1$ received weak tones, all of the important harmonically related filters, including the one at f_1, were filled with noise. As such, the weak tones may be thought of as having acting as seeds, highlighting a specific complex pitch which then recruited noise-based energy in all of its first 7 harmonics to produce essentially the same noisy pitch and timbre, regardless of the seeded frequencies.

F. Phenomenology and cueing?

An important factor not directly addressed by the model is the role of phenomenology shared by the cue and signal. One could postulate that trial-by-trial feedback led subjects to attend to the appropriate elements in the LSRD without insisting that the cue and signal sounded alike, but it seems more plausible that perceived qualities of the cue should guide the listener both in selection of the appropriate LSRD and of the best filter in it. Thus, while the simple sounds-like hypothesis of cueing with its emphasis on conscious awareness of stimulus features seems insufficient for the asymmetry of cueing found with a fundamental and its harmonics, it seems equally clear that phenomenology must play an important part in focusing the attention on the stimulus dimension to be monitored during search and detection.

VI. SUMMARY

An important factor in signal detection is the extent to which the subject knows what the signal will be. When there is uncertainty about some property of the signal, it is necessary for the subject to monitor more potential filters in the stimulus domain, raising the probability of more false positives due to peaks in the masker. Typically, uncertainty is reduced in the laboratory by offering practice trials, postresponse feedback, and cueing. The present study concentrates on the latter, that is, on the effects of presenting sensory cues

that inform the subject what to listen for. In order to stress the importance of cueing, there was a high degree of signal uncertainty, with signals differing from one another on every trial. While it seems obvious to say that successful cueing must, somehow, elicit an internal representation of the signal to be detected, this study began in response to an observation that a complex cue improved the detection of one of its primitives, but that use of the primitive as a cue had little or no effect on detection of the complex. The model proposed to explain this asymmetry notes that complex signals are apt to be represented at multiple neural sites, some organized according to such primitive features as acoustic frequency and others on emergent features, such as complex pitch, derived from interactions between primitives. Based on the argument that ideal performance requires that the level of stimulus representation used for detection, or LSRD, should be the one with the highest S/N, the model predicts that best performance with a complex signal will occur when it is detected on the basis of its complex feature. Finally, the model assumes that successful cueing requires that the cue and signal share unique neural elements at the LSRD.

Tests of the LSRD model offered here used signals that were three-tone complexes chosen at random for each trial. In support of the assumption that complex features would provide a higher S/N than primitive ones, performance was weakest in the condition where the three tones bore no relation to one another, making it necessary to detect them on the basis of the individual frequencies. Performance improved when the three tones were related harmonically, providing a complex pitch as the basis for detection. Cues chosen to test the final assumption of the model were also three-tone complexes. In support of the argument that they must share representation with the signal at the appropriate LSRD, uncertainty about a signal made up of unrelated tones was ameliorated by a cue that shared its individual frequencies, while uncertainty about a signal made up of harmonically related tones was ameliorated by a cue that shared its complex pitch. Finally, in support of the idea of separate and independent levels of representation for the two types of detection, randomly chosen harmonic complexes were detected still better if preceded by cues that matched them both in frequency and in complex pitch.

ACKNOWLEDGMENTS

This work was supported by a grant from the USPHS, Institute on Deafness and Other Communication Disorders, No. NIDCD 00087. We wish to thank the many friends and colleagues who offered invaluable comments on earlier versions of the manuscript. These include Poppy Crum, Erick Gallun, Eric Jensen, Gerald Kidd, Christine Mason, Dennis McFadden, and Miriam Valenzuela.

APPENDIX

The LSRD model suggests three planned comparisons based on the argument that complex signals would be more detectable primitives, C2>C1, and the efficacy of the two types of cues, C3>C1 and C4>C2. Separate paired comparisons were made for the three subjects based on the minimum number of trials (350) collected for each condition. That is, for the mth subject, a z-score of the difference between conditions Y and X was calculated as

$$z(m,y,x)=\frac{P(C)_{(m,y)}-P(C)_{(m,x)}}{\sqrt{[P(C)_{(m,y)}][1-P(C)_{(m,y)}]+[P(C)_{(m,x)}][1-P(C)_{(m,x)}]}}. \quad \text{(A1)}$$

Results of these analyses are shown in Table AI. Use of a one-tailed test for the planned comparisons based on prior theory (Keppel and Zedeck, 1989) showed that eight of the nine $P(C)$-differences were significant ($p<0.05$), with the worst case being for subject 2 in C4>C2, where the difference was marginally insignificant ($p<0.07$). Indeed, seven of the nine cases were significant at $p<0.01$, now exempting subject 1 in C2>C1 ($p<0.02$).

For condition 5, where both types of cues were present, the argument that bi-dimensional cues aided independent processing of stimuli at the two putative LSRDs is strongly supported by results in Table II, which show that the amount of transmitted information, $(d')^2$ (Green and Swets, 1966), in condition 5 actually exceeded the sum of the $(d')^2$ values from the two singly cued conditions. Signals were the same in conditions 2, 4, and 5, each offering the possibility of detection on the basis of either frequency or complex pitch. The ability to use pitch cues with these signals is clearly shown by the comparison C4>C2. If frequency cueing provided additional improvement by ameliorating uncertainty about the individual frequencies, one would expect still higher performance in condition 5. The comparisons of C5>C4 shown in the fourth row of Table AI were significant for all subjects ($p<0.01$).

TABLE AI. Subject-by-subject planned comparisons between conditions (see text of the Appendix).

Conditions compared	Subject 1	Subject 2	Subject 3
C2>C1	2.0254	3.8216	2.9410
C3>C1	3.8990	5.5966	5.1237
C4>C2	3.9401	1.4808	2.5797
C5>C4	2.5277	5.1497	5.4384

[1]Hafter *et al.* (1992) reported a case in which detections of randomly chosen, pure tone signals of frequency f_1 were tested alone and when preceded either by a single-tone cue of the same frequency or by sets cues made up

of five harmonics $2f_1 \rightarrow 6f_1$ or $3f_1 \rightarrow 7f_1$ of the signal. The proportion of correctly identified signals in a two-alternative, forced choice task was 0.68 with no cueing, 0.92 with a same-frequency cue, and an in-between value of 0.835 with the complex cue.

[2]As noted in Sec. V C, conditions 1–4 were repeated using the same limited range of frequencies for randomly drawn frequencies harmonically related signals. Data from three subjects were collected using a two-down, one-up tracking procedure (16 reversals) to produce four to eight tracks per subject. Thresholds obtained in that way are not directly comparable to $P(C)$ values in the main experiment, but the same basic effect accrued. Relative to the case with purely random tones (condition 1), the average threshold for harmonically related tones (condition 2) was -0.66 dB; for random tones cued by the same frequencies (condition 3) it was -2.28 dB; for harmonically related signals cued by the other three harmonics (condition 4) it was 4.05 dB.

Dai, H. P., Scharf, B., and Buus, S. (**1991**). ''Effective attenuation of signals in noise under focused attention,'' J. Acoust. Soc. Am. **89**, 2837–2842.

Darwin, C. (**1995**). Review of a manuscript submitted to J. Acoust. Soc. Am.

de Boer, E. (**1976**). ''On the 'residue' and auditory pitch perception,'' in *Handbook of Sensory Physiology*, edited by W. D. Keidel and W. D. Neff, Vol. V, Part 3, pp. 479–583.

Green, D. M. (**1961**). ''Detection of auditory sinusoids of uncertain frequency,'' J. Acoust. Soc. Am. **33**, 904–911.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Krieger, Huntington, New York).

Green, D. M., McKey, M. J., and Licklider, J. C. R. (**1959**). ''Detection of a pulsed sinusoid in noise as a function of frequency,'' J. Acoust. Soc. Am. **31**, 1446–1452.

Hafter, E. R., and Kaplan, R. (**1976**). NASA Ames Report, ''A role for attention in auditory detection under conditions of frequency uncertainty.''

Hafter, E. R., and Schlauch, R. S. (**1989**). ''Factors in detection under uncertainty,'' J. Acoust. Soc. Am. **86**, S112.

Hafter, E. R., Schlauch, R. S., and Tang, J. (**1993**). ''Attending to auditory filters that were not stimulated directly,'' J. Acoust. Soc. Am. **94**, 743–747.

Hafter, E. R., Saberi, K., Schlauch, R. S., Plamondon, L., and Tang, J. (**1992**). ''Use of Stimulus Pre-cues to Study Processing at Various Levels of Auditory and Musical Processing,'' Second International Conference on Music Perception and Cognition, Univ. of California, Los Angeles, CA, 22–26 Februrary.

Houtgast, T. (**1976**). ''Subharmonic pitches of a pure tone at low S/N ratio,'' J. Acoust. Soc. Am. **60**, 405–409.

Houtsma, A. J. M. (**1984**). ''Pitch Salience of Complex Sounds,'' Music Percept. **1**, 296–307.

Houtsma, A. J. M. (**1995**). ''Pitch Perception,'' in *Handbook of Perception and Cognition*, edited by B. C. J. Moore (Academic, San Diego).

Houtsma, A. J. M., Rossing, T. D., and Wagenaars, W. M. (**1987**). ''Auditory Demonstrations, Demonstration 23,'' Institute for Perception Research, Eindhoven, The Netherlands.

Howard, Jr., J. H., O'Toole, A. J., Parasuraman, R., and Bennett, K. B. (**1984**). ''Pattern-directed attention in uncertain-frequency detection,'' Percept. Psychophys. **35**, 256–264.

Howard, J. H., O'Toole, A. J., and Rice, S. E. (**1986**). ''The role of frequency versus informational cues in uncertain frequency detection,'' J. Acoust. Soc. Am. **79**, 788–791.

Johnson, D. M., and Hafter, E. R. (**1980**). ''Uncertain-frequency detection: Cuing and condition of observation,'' Percept. Psychophys. **28**, 143–149.

Keppel, G., and Zedeck, S. (**1989**). *Data Analysis for Research Designs* (Freeman, New York).

Langner, G., Sams, M., Heil, P., and Schulze, H. (**1997**). ''Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: evidence from magnetoencephalography,'' J. Comp. Physiol., A **181**, 665–676.

Licklider, J. C. R. (**1956**). ''Auditory frequency analysis,'' in *Information Theory*, edited by C. Cherry (Academic, New York).

McFadden, D. (**1988**). ''Failure of a missing-fundamental complex to interact with masked and unmasked pure tones at its fundamental frequency,'' Hear. Res. **32**, 23–40.

Moore, B. C. J. (**1997**). *Psychology of Hearing* (Academic, London).

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1986**). ''Thresholds for hearing mistuned partials as separate tones in harmonic complexes,'' J. Acoust. Soc. Am. **81**, 1873–1880.

Patterson, R., 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, New York).

Pickles, J. O. (**1988**). *An Introduction to the Physiology of Hearing* (Academic, London).

Plamondon, L., and Hafter, E. (**1990**). ''Selective attention in absolute pitch listeners,'' J. Acoust. Soc. Am. Suppl. 1 **88**, S49.

Plomp, R. (**1964**). ''The ear as a frequency analyzer,'' J. Acoust. Soc. Am. **36**, 1628–1636.

Plomp, R., and Mimpen, A. M. (**1968**). ''The ear as a frequency analyzer II,'' J. Acoust. Soc. Am. **43**, 764–767.

Scharf, B. (**1970**). ''Critical Bands,'' in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol 1, pp. 157–202.

Scharf, B., Chays, A., and Magnan, J. (**1997**). ''The role of the olivocochlear bundle in hearing: Sixteen case studies,'' Hear. Res. **103**, 101–122.

Scharf, B., Magnan, J., Collet, L., Ulmer, E., and Chays, A. (**1994**). ''On the role of the olivocochlear bundle in hearing: A case study,'' Hear. Res. **75**, 11–26.

Schlauch, R. S., and Hafter, E. R. (**1991**). ''Listening bandwidths and frequency uncertainty in puretone signal detection,'' J. Acoust. Soc. Am. **90**, 1332–1339.

Schouten, J. F. (**1940**). ''The residue and the mechanism of hearing,'' Proc. K. Ned. Akad. Wet. **43**, 991–999.

Schreiner, C. E., and Langner, G. (**1988**). ''Periodicity coding in the inferior colliculus of the cat. II Topographical organization,'' J. Neurophysiol. **60**, 1823–1840.

Swets, J. A. (**1964**). *Signal Detection and Recognition by Human Observers* (Wiley, New York).

Swets, J. A., and Sewall, S. T. (**1961**). ''Stimulus vs. Response uncertainty in recognition,'' J. Acoust. Soc. Am. **33**, 1586–1592.

Terhardt, E. (**1974**). ''Pitch, consonance and harmony,'' J. Acoust. Soc. Am. **55**, 1061–1069.

Terhardt, E., Stoll, G., and Seewann, M. (**1982**). ''Pitch of complex according to virtual-pitch theory: Tests, examples, and predictions,'' J. Acoust. Soc. Am. **71**, 671–678.

Watson, C. S., and Kelly, W. J. (**1981**). ''The role of stimulus uncertainty in the discrimination of auditory patterns,'' in *Auditory and Visual Pattern Recognition*, edited by D. H. Getty and J. H. Howard (Erlbaum, Hillsdale, NJ).

Melody recognition using three types of dichotic-pitch stimulus

Michael A. Akeroyd[a)]
Department of Neuroscience, University of Connecticut Health Center, Farmington, Connecticut 06030 and Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, United Kingdom

Brian C. J. Moore and Geoffrey A. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, United Kingdom

(Received 5 January 2001; revised 21 May 2001; accepted 8 June 2001)

The recognition of 10 different 16-note melodies, constructed using either dichotic-pitch stimuli or diotic pure-tone stimuli, was measured. The dichotic pitches were created by placing a frequency-dependent transition in the interaural phase of a noise burst. Three different configurations for the transition were used in order to give Huggins pitch, binaural-edge pitch, and binaural-coherence-edge pitch. Forty-nine inexperienced listeners participated. The melodies evoked by the dichotic stimuli were consistently identified well in the first block of trials, indicating that the sensation of dichotic pitch was relatively immediate and did not require prolonged listening experience. There were only small improvements across blocks of trials. The mean scores were 97% (pure tones), 93% (Huggins pitch), 89% (binaural-edge pitch), and 77% (binaural-coherence-edge pitch). All pairwise differences were statistically significant, indicating that Huggins pitch was the most salient of the dichotic pitches and binaural-coherence-edge pitch was weakest. To account for these differences in salience, a simulation of lateral inhibition was applied to the recovered spectrum generated by the modified equalization cancellation model [J. F. Culling, A. Q. Summerfield, and D. H. Marshall, J. Acoust. Soc. Am. **103**, 3509–3526 (1998)]. The height of the peak in the resulting ''edge-enhanced'' recovered spectrum reflected the relative strength of the different dichotic pitches.
 [DOI: 10.1121/1.1390336]

PACS numbers: 43.66.Pn, 43.66.Hg, 43.66.Ba, 43.75.Cd [SPB]

I. INTRODUCTION

Since Cramer and Huggins' (1958) pioneering research, it has been known that pitch sensations can be created by the binaural interaction of noise stimuli. These ''dichotic pitches'' are analogous to the visual objects that can be seen in random-dot stereograms (e.g., Julesz, 1971); the stimulus at each ear gives no pitch sensation, but, when presented binaurally, disparities between the two ears lead to the perception of pitch. We report below a study of the recognition of melodies produced using three types of dichotic-pitch stimulus and also using pure tones.

The types of dichotic-pitch stimulus used were the ''Huggins pitch'' (Cramer and Huggins, 1958), the ''binaural-edge pitch'' (Klein and Hartmann, 1981), and the ''binaural-coherence-edge pitch'' (Hartmann, 1984; Hartmann and McMillon, 2001). For all of the stimuli, the presence of a single frequency-dependent transition in the interaural phase of a broadband noise gives rise to the perception of pitch. This percept is similar to that of a pure, although faint, tone. These transitions are illustrated schematically in Fig. 1. In a Huggins-pitch stimulus (left panel) the interaural phase changes progressively from 0 to 2π (equivalent to 0) radians across a narrow frequency region. The width of this region has sometimes been varied in previous studies; here, we use a fixed value of 16% of the center frequency of the transition.[1] In a binaural-edge-pitch stimulus (middle panel) the interaural phase changes abruptly from 0 to π radians. In a binaural-coherence-edge-pitch stimulus (right panel) the interaural phase changes abruptly from 0 radians to a random value. Although variations of these dichotic-pitch stimuli have been reported, differing in the interaural phase of the carrier noise in the case of Huggins pitch or the spectral direction of the transition in interaural phase in the case of the two edge pitches, we deal here only with the ''prototypical'' variations shown in Fig. 1.

The value of the perceived pitch has been measured by asking listeners to match the pitch using a pure tone of adjustable frequency. For Huggins pitch, the matching frequency is commonly found to be equal to the center frequency of the transition (e.g., Culling *et al.*, 1998). For binaural-edge pitch, Klein and Hartmann (1981) found that the distribution of matching frequencies was bimodal, with one peak slightly above and the other peak slightly below the transition frequency. Subsequent measurements conflict with this result, however, indicating a unimodal distribution with a peak centered on the transition frequency (Frijns *et al.*, 1986; Culling *et al.*, 1998); at present there is no agreement as to whether the distribution is bimodal or unimodal. For the particular variation of binaural-coherence-edge pitch used here, Hartmann and McMillon (2001) found a unimodal distribution, but with a peak placed approximately 5%–10% above the transition frequency.

Theories differ as to the nature of the binaural processing that creates the sensation of pitch from the transition in interaural phase (e.g., Licklider, 1959; Durlach, 1962; Klein

a)Author to whom correspondence should be addressed; electronic mail: maa@biols.susx.ac.uk

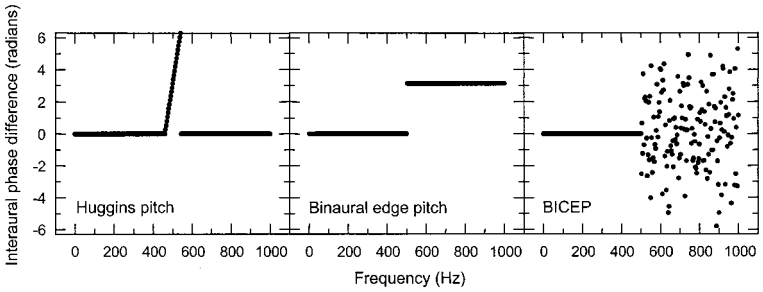


FIG. 1. Interaural phase difference plotted as a function of frequency for the three types of dichotic-pitch stimulus. The transition frequency was 500 Hz in each case. Each dot represents the interaural phase difference of one frequency component (spacing=3.3 Hz). Note that the interaural phase difference for the binaural-coherence-edge pitch (abbreviated to ''BICEP'') stimulus was random above the transition frequency and thus would have differed in each of the experimental stimuli from that plotted. The interaural amplitude difference of all the frequency components was zero, apart from the frequency components higher than the transition frequency in the BICEP stimulus, for which the interaural amplitude difference was random.

and Hartmann, 1981; Raatgever and Bilsen, 1986; Culling *et al.*, 1998; Akeroyd and Summerfield, 1999). The two most developed models of dichotic pitch are the central-spectrum model (e.g., Raatgever and Bilsen, 1986; Frijns *et al.*, 1986) and the modified-equalization-cancellation model (e.g., Culling *et al.*, 1998). In both of these models, the sensation of pitch is assumed to be produced by a peak in a spectrum calculated on the basis of binaural cues, but the method of determining this ''binaural spectrum'' differs across models. The sensation of a dichotic pitch is assumed to be due to the presence of a peak in the binaural spectrum, by analogy with the peak in a monaural excitation pattern created by a pure-tone stimulus. As the purpose of this study was not to test experimentally the different models of dichotic-pitch sensation, for simplicity we limit the present discussion to the modified-equalization-cancellation model.

Two operations are fundamental to this model. First, the signals at each ear are passed through an array of auditory filters. Second, processes of equalization (of overall differences in interaural time and intensity) and then cancellation are applied to each filter output, with the goal of minimizing the power of the remainder after cancellation (Durlach, 1972). The minimum value of the power is termed the ''residual activation.'' The model allows *independent* time delays to be used in each filter channel when applying the time equalization. The result is a spectrum of residual activation versus frequency and is termed the ''recovered spectrum.'' Figure 2 shows example recovered spectra for the three types of dichotic-pitch stimuli used here, each with a transition frequency of 500 Hz (the computational details of the calculation of these spectra are reported in Sec. IV below). For both the Huggins-pitch stimulus (solid line) and the binaural-edge-pitch stimulus (dotted line), there is a single peak at the transition frequency. For the binaural-coherence-edge-pitch stimulus (dashed line), there is instead a high-pass recovered spectrum with a sloping edge. It is presumed that a process of lateral inhibition is applied to the edge in the recovered spectrum, so creating a single peak placed slightly above the transition frequency.

It seems reasonable to assume that the strength of the pitch sensation is related to the height of the peak in the recovered spectrum. If so, binaural-edge pitch should be slightly weaker than the version of Huggins pitch used here. (Culling *et al.*, 1998, Fig. 4, showed that the height of the peak in the recovered spectrum of Huggins pitch is dependent upon the bandwidth of the transition in interaural phase. The present simulations were based on the same bandwidth, 16%, as used in our experiment.) A quantitative prediction of the strength of binaural-coherence-edge pitch cannot be made without a function representing the amount of lateral inhibition. It would seem reasonable, nevertheless, to expect that a peak introduced by this extra stage of processing would not be as large as a peak directly present in the recovered spectrum for the Huggins-pitch and binaural-edge-pitch stimuli. Hence, binaural-coherence-edge pitch should be weaker than the other two pitches.

One way of estimating the strength of a dichotic pitch is by adjusting the signal-to-noise ratio of a pure tone presented

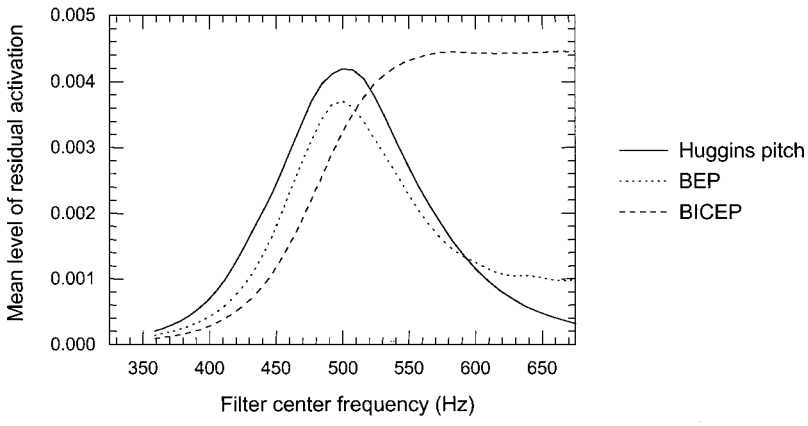


FIG. 2. Recovered spectra for a Huggins-pitch stimulus (solid line), binaural-edge-pitch stimulus (dotted line) and binaural-coherence-edge-pitch stimulus (dashed line). The calculations underlying the spectra are summarized in Sec. IV.

TABLE I. The frequencies of each of the 16 notes forming the 10 melodies used in the experiment.

Name of melody	Frequency of Note															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Au Clair de la Lune	C_5	C_5	C_5	D_5	E_5	E_5	D_5	D_5	C_5	E_5	D_5	D_5	C_5	C_5	C_5	C_5
Frere Jacques	C_5	D_5	E_5	C_5	C_5	D_5	E_5	C_5	E_5	F_5	G_5	G_5	E_5	F_5	G_5	G_5
Twinkle, Twinkle Little Star	C_5	C_5	G_5	G_5	A_5	A_5	G_5	G_5	F_5	F_5	E_5	E_5	D_5	D_5	C_5	C_5
This Old Man	F_5	D_5	F_5	F_5	F_5	D_5	F_5	F_5	G_5	F_5	$D_5^{\#}$	D_5	C_5	D_5	$D_5^{\#}$	$D_5^{\#}$
God Rest Ye Merry Gentlemen	D_5	D_5	A_5	A_5	G_5	F_5	E_5	D_5	C_5	D_5	E_5	F_5	G_5	A_5	A_5	A_5
Yankee Doodle	F_5	F_5	G_5	A_5	F_5	A_5	G_5	C_5	F_5	F_5	G_5	A_5	F_5	F_5	E_5	E_5
Good King Wenceslas	F_5	F_5	F_5	G_5	F_5	F_5	C_5	C_5	D_5	C_5	D_5	E_5	F_5	F_5	F_5	F_5
Chimes of Big Ben	A_5	F_5	G_5	C_5	C_5	G_5	A_5	F_5	A_5	F_5	G_5	C_5	C_5	G_5	A_5	F_5
Lead Us Heavenly Father Lead Us	C_5	E_5	G_5	G_5	A_5	G_5	F_5	E_5	E_5	F_5	G_5	C_5	E_5	D_5	C_5	C_5
Bobby Shaftoe	F_5	F_5	F_5	$A_5^{\#}$	A_5	C_6	A_5	F_5	C_5	C_5	C_5	F_5	E_5	G_5	E_5	C_5

in noise so as to match the pitch strength. Klein and Hartmann (1981) reported that the strength of binaural-edge pitch was slightly greater than that of Huggins pitch, although the differences between the two pitches were small and somewhat variable across frequency. We know of no equivalent measurements of the strength of binaural-coherence-edge pitch, although Hartmann and McMillon (2001) asked their listeners to make informal comparisons of the two pitches. They reported that a subset of their listeners found binaural-edge pitch to be the stronger, although none reported a ''striking'' (p. 303) difference in strength. An indirect estimate can be obtained from the variability in pitch matches. Hartmann and McMillon (2001) reported a standard deviation of approximately 3% of the transition frequency for binaural-coherence-edge pitch. This value is only slightly greater than the value of approximately 2% reported by Frijns *et al.* (1986) for binaural-edge pitch, suggesting that, within the margins of error, the two edge pitches may have about the same pitch strength. However, the possibility of individual differences in matching accuracy makes it difficult to draw firm conclusions from comparisons across studies.

The first question addressed by this study was: do the three types of dichotic pitch differ in salience, as might be expected on the basis of the above arguments? To answer this question, we used a melody recognition task. We were also interested in the salience of dichotic pitches relative to the clear pitches produced by pure-tone stimuli. Therefore we included a set of melodies produced by sequences of pure tones.

The second question of interest was: are dichotic pitches perceived without extensive training or does their perception require prolonged experience? To answer this question, we measured the degree to which melody recognition improved across blocks of trials. The experiment was conducted as part of a practical class, hence allowing a relatively large group (49) of inexperienced listeners to participate.

II. METHOD

A. Stimuli

The stimuli were created using MATLAB. Each stimulus consisted of a melody defined by a train of 16 notes of 300-ms duration, with each note separated by a silence of 100-ms duration. Ten melodies were used; they were the same as those used by Moore and Rosen (1979). They were based on well-known melodies, but were modified by slight distortions of the rhythm so as to consist of 16 equal-duration notes. These modifications required long-duration notes to be split into two or more equal-duration shorter notes. Moore and Rosen reported that their listeners did not find the modifications to be particularly disturbing. The frequencies corresponding to the notes in each melody are listed in Table I. The minimum, mean, and maximum frequencies were, respectively, 524 Hz ($=C_5$), 707 Hz ($\approx F_5$), and 1046 Hz ($=C_6$). Pilot experiments indicated that the chosen frequency range gave clear sensations of dichotic pitch, at least for experienced listeners.

In the pure-tone condition, each individual note in a melody was a tone burst with a frequency equal to that of the note. The pure-tone stimuli were presented diotically. In the three dichotic-pitch conditions, each individual note in a melody was a burst of dichotic bandpass noise which contained an interaural phase transition centered on the frequency of the note. Each dichotic-pitch stimulus was created in the spectral domain by rectangular filtering (0–4000 Hz passband) two matched 6000-point buffers representing the left and right channels of a diotic Gaussian noise sampled at 20 000 Hz. The phases of frequency components in the spectral buffer representing one channel were then modified. For each Huggins-pitch stimulus, a linear shift of 0 to 2π radians was added to the phases for frequency components from 8% below to 8% above the frequency of the note. For each binaural-edge-pitch stimulus, a shift of π radians was added to the phase of frequency components above the frequency of the note. For each binaural-coherence-edge-pitch stimulus, the amplitude and phase of frequency components above the frequency of the note in one spectral buffer were generated independently of those in the other spectral buffer. Subsequently the signal waveforms for the left and right channels were created by applying an inverse discrete Fourier transform to the two spectral buffers. A 30-ms raised-cosine ramp was applied to the onset and offset of each note. The melodies were recorded onto CD(R) for presentation to listeners using Sennheiser HD-414 earphones. All the earphones were driven in parallel from the same amplifier. The overall level of the stimuli was about 70 dB SPL at each ear.

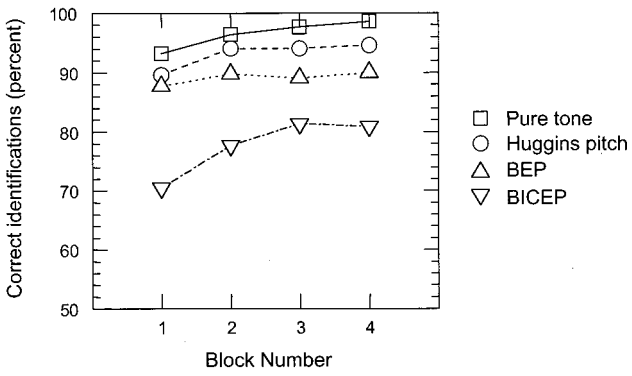


FIG. 3. Mean percentage of correct identifications in each of the four blocks of the test phase of the experiment for each type of pitch. The chance level is 10%. ''BEP'' is used as an abbreviation for binaural-edge pitch and ''BICEP'' is used as an abbreviation for binaural-coherence-edge pitch.

B. Procedure

The listeners were trained using a two-stage method. First, they received training to help them learn the name of each melody. Each of the 10 melodies, played with the pure-tone stimuli, was presented twice. Second, they received training to help them hear the pitches associated with each type of stimulus. They heard each of the 10 melodies, presented twice, first played with the Huggins-pitch stimuli, then with the binaural-edge-pitch stimuli, and finally with the binaural-coherence-edge-pitch stimuli. Throughout all of the training the name of a melody (also recorded on the CD) was announced before its presentation.

The test phase of the experiment was divided into four blocks. In each block, listeners heard each of the 40 possible combinations of melody and type of stimulus, presented in a random order (the ordering was constrained so that neither the same melody nor the same type of stimulus were presented in successive trials). They were required to identify each melody immediately after its presentation and to write their identifications on a score sheet. Feedback (the name of the tune) was provided after a delay of four seconds. The listeners scored their own responses as the experiment proceeded, and the response sheets were later checked for accuracy by the experimenters.[2]

Forty-nine undergraduate students enrolled at the University of Cambridge participated as listeners. The results reported below are based on the responses of 44 listeners, as responses were excluded for one listener who had a self-reported hearing loss and for four other listeners because they performed perfectly (the exclusion of these four sets of results does not affect the statistical analyses reported below).

III. RESULTS AND DISCUSSION

The mean scores across listeners are shown in Fig. 3. The symbols indicate the identification scores for the pure-tone stimuli (squares), Huggins-pitch stimuli (circles), binaural-edge pitch stimuli (upward-pointing triangles), and binaural-coherence-edge-pitch stimuli (downward-pointing triangles). There was no consistent pattern in the identification scores for the different melodies. Although some listeners showed patterns of errors indicating difficulty with specific melodies, these patterns were not consistent across listeners. Overall, each of the melodies was about equal in identifiability.

Three major outcomes can be identified. First, for all the types of stimuli, the identification scores improved from block 1 to block 2 and thereafter remained relatively stable. A Wilcoxon matched-pairs signed-ranks test (Siegel and Castellan, 1988), calculated on the mean responses across the four types of stimulus, showed that mean scores for blocks 1 and 2 differed significantly ($T=39$, $N=32$; $p<0.005$) but that scores for blocks 2 and 3 or for blocks 3 and 4 did not differ significantly ($T=184$, $N=34$, $p>0.05$; and $T=249$, $N=32$, $p>=0.05$, respectively). The initial improvement suggests a small practice effect. We believe that it represents an improvement in learning to attach names to the melodies, because the effect occurs for the pure-tone stimuli also, whose pitches were presumably clear and easily identified.

Second, the identification scores were consistently high in the first block for the three types of dichotic pitch stimulus. The high performance was especially noteworthy for the Huggins-pitch stimuli and binaural-edge-pitch stimuli, for which 41 out of 44 subjects scored better than 60% on the first block of trials (chance corresponds to 10%). This result supports the notion that the sensation of dichotic pitch is relatively immediate and does not require prolonged experience. Furthermore, the pitch appears to be ''musical,'' in that it readily supports melody recognition. It is of interest that the static interaural differences present in our stimuli were sufficient to support melody recognition. This contrasts with recent results of Culling (2000), obtained using stimuli similar to those of Kubovy *et al.* (1974). The latter presented eight continuous sinusoids to each ear, via earphones. The sinusoids had frequencies corresponding to the notes in a musical scale. Seven of the sinusoids were delayed by 1 ms at one ear. The remaining sinusoid was delayed by 1 ms at the other ear, with the result that this component had an ITD that was shifted by 2 ms from the ITDs of the other components. The shifted component was heard to stand out perceptually. A sequence of ITD shifts in different components was clearly heard as a melody. This melody was completely undetectable when listening to the input to one ear alone. Kubovy *et al.* interpreted their results as indicating that differences in relative phase at the two ears can allow an auditory ''object'' to be isolated in the absence of any other cues. However, Culling presented evidence that the transitions in interaural phase were the dominant cue allowing melody recognition. Static differences in interaural phase were found to be a weak cue ''that only a subset of listeners were able to exploit'' (Culling, 2000, p. 1768). Our results indicate that, for noise stimuli, static differences in the interaural phase can be used for melody recognition by the great majority of relatively untrained subjects.

The third major outcome was that identification scores differed across the four types of stimulus. The rank ordering (of highest scores to lowest scores) was: pure tones, Huggins pitch, binaural-edge pitch, and binaural-coherence-edge pitch. This ordering was observed in each of the four blocks as well in the across-block mean. Furthermore, the ordering

TABLE II. Results from a group of Wilcoxon matched-pair signed-rank tests applied to each pairwise comparison of the types of stimuli. In each test the value of T was less than the critical value of T required for a significance level of 0.0083 (used as it is equal to the conventional significance level of 0.05 divided by 6, which was the number of tests); thus each comparison shows a statistically significant effect. ''BICEP'' is used as an abbreviation for ''binaural coherence edge pitch.''

		Type of stimulus		
Type of stimulus	Pure tone	Huggins pitch	Binaural-edge pitch	BICEP
Pure tone	—			
Huggins pitch	$T=38$, $N=32$	—		
Binaural-edge pitch	$T=0$, $N=32$	$T=72$, $N=32$	—	
BICEP	$T=7$, $N=42$	$T=14$, $N=42$	$T=63$, $N=44$	—

was the same or similar in the individual scores for each listener; almost every listener scored worst for the binaural-coherence-edge-pitch stimuli. Table II reports the results of a set of Wilcoxon matched-pairs signed-ranks tests, which show that each of the six pairwise differences between the types of stimuli was statistically significant.

This outcome suggests that all of the pitch sensations produced by the dichotic-pitch stimuli are less salient than the pitch sensation produced by pure-tone stimuli. Furthermore, of the three types of dichotic pitch, Huggins pitch is the most salient, binaural-edge pitch is intermediate, and binaural-coherence-edge pitch is the weakest. This conclusion is further supported by the result that some of the listeners made errors in identification of the binaural-coherence-edge-pitch stimuli but made no errors for the other stimuli. These results are broadly consistent with the analysis described in the Introduction based upon the modified equalization-cancellation model and shown earlier in Fig. 2. However, to make the analysis more rigorous, a model of edge enhancement must be developed in order to account for the existence of the binaural-coherence-edge pitch. This is done in the next section.

It should be noted that there is nothing in the following model of edge enhancement that requires the use of the *recovered* spectrum generated by the modified equalization-cancellation model. It could also be applied to the spectrum of binaural activity, at a fixed internal time delay, generated by the central-spectrum model (e.g., Raatgever and Bilsen, 1986; Frijns *et al.*, 1986), or it could also act an implementation of the (nonmodified) equalization-cancellation explanation of binaural-coherence-edge pitch (Hartmann and McMillon, 2001). We expect that these models will give similar results, but may require different values of the free parameter that determines the relative magnitudes of the spectrum and the edge enhancement. For simplicity, however, and also because the purpose of the present study was not to compare experimentally the various models, we limit the analysis to the recovered spectrum.

IV. A MODEL OF EDGE ENHANCEMENT

Figure 2 shows the recovered spectra for the three types of dichotic pitches used in the experiment. The recovered spectra were calculated as follows. First, bursts of each dichotic-pitch stimulus were synthesized, each of 300-ms duration and 40-dB spectrum level. The parameters of the transition in interaural phase defining the dichotic pitches were the same as those used in the experiment, with the modification that the transition frequency was fixed at 500 Hz. The left and right waveforms of the stimuli were passed through matched 41-channel gammatone filterbanks coded in MATLAB (Patterson *et al.*, 1995; Slaney, 1998), with center frequencies ranging from two equivalent rectangular bandwidths (''ERBs,'' Glasberg and Moore, 1990) below 500 Hz to 2 ERBs above 500 Hz and with filters spaced at 0.1-ERB intervals. The output of each filter was then passed through a model of a high-spontaneous-rate fiber, with a threshold of 45 dB, that computed the probability of firing (Meddis *et al.*, 1990, Table II). The equalization and cancellation processes were combined by subtracting the probability of fiber output for a left channel from the probability of fiber output for the corresponding right channel, as a function of a time delay applied to one channel. Of the set formed by the remainder-after-cancellation at each of the time delays, the remainder with the smallest mean value was measured. This process was repeated for each frequency channel, so giving the recovered spectrum of the residual activation as a function of frequency. The recovered spectra shown in Fig. 2 are means across 25 independent bursts of each dichotic-pitch stimulus.

As may be seen, the mean recovered spectrum of the binaural-coherence-edge-pitch stimuli (dashed line) shows a sloping edge rather than a peak. Next, we describe a model for generating a peak in this spectrum. This model is effectively a computational implementation of a process of lateral inhibition applied to the recovered spectrum. Illustrations of the operation of lateral inhibition, applied to a sloping edge like that seen in the recovered spectrum for binaural-coherence-edge pitch, can be found in von Békésy (1959, Fig. 8) and Small and Daniloff (1967, Fig. 5). The model is based on extracting the second derivative (i.e., the curvature) of the recovered spectrum and inverting its sign. In the transition from a low-level flat portion of the recovered spectrum to a rising edge, the curvature is positive, so the inverse is negative. In the transition from a rising edge to a high-level plateau, the curvature is negative, so the inverse is positive. Thus, the inverse of the curvature has the appropriate properties for simulating the effects of lateral inhibition.

In detail, the first-order derivatives of the recovered spectrum were calculated using

$$\frac{dr_z}{dz} = r_z - r_{z-1}, \tag{1}$$

and then the second-order derivatives were calculated using

$$\frac{d^2r_z}{dz^2}=\frac{dr_{z+1}}{dz}-\frac{dr_z}{dz}=r_{z+1}-2r_z+r_{z-1}, \tag{2}$$

where z is the frequency coordinate (i.e., the channel number), and r_z is the level of the recovered spectrum (i.e., the residual activation) in the zth frequency channel. It should be remembered that the separation of the frequency channels is 0.1 ERBs and so the denominator in these two equations is also 0.1 ERBs. Next, minor fluctuations in the second-order differences were smoothed by averaging the values across ± 4 adjacent channels, corresponding to ± 0.4 ERBs:

$$\left(\overline{\frac{d^2r_z}{dz^2}}\right)=\frac{1}{4+4+1}\sum_{z-4}^{z+4}\frac{d^2r_z}{dz^2}$$
$$=\tfrac{1}{9}(r_{z+5}-r_{z+4}-r_{z-4}+r_{z-5}). \tag{3}$$

We refer to the function defined by Eq. (3) as the ''edge-enhancement function.'' This function was then inverted in sign, scaled, and added to the original recovered spectrum so as to simulate the process of lateral inhibition, giving an ''edge-enhanced recovered spectrum.'' The value of the scaling factor is a free parameter in the model; here a value of 50 was chosen so as to give a peak of reasonable height in the edge-enhanced recovered spectrum for the binaural-coherence-edge-pitch stimuli. The edge-enhanced recovered spectrum, E_z, is thus given by

$$E_z=r_z-50\left(\overline{\frac{d^2r_z}{dz^2}}\right). \tag{4}$$

One consequence of choosing a factor of 50 was that, in the final edge-enhanced recovered spectrum, both the original recovered spectrum r_z and the edge-enhancement function [Eq. (3)] each contribute 50% of the value of the peak of Huggins pitch, although the results are insensitive to the exact value of the constant.

Figure 4 shows the results of applying each step of this ''edge-enhancement'' model to the recovered spectra of the three dichotic-pitch stimuli considered here. In each panel, the solid line represents Huggins pitch, the dotted line represents binaural-edge-pitch, and the dashed line represents binaural-coherence-edge pitch. Panel A shows the same recovered spectra that were illustrated in Fig. 2. Note that there are clear peaks in the functions for Huggins pitch and binaural-edge pitch but not for binaural-coherence-edge pitch. Panel B shows the first-order differences [see Eq. (1)] and panel C shows the unsmoothed second-order differences [see Eq. (2)]. Note that, for both Huggins pitch and binaural-edge pitch, the peak in the recovered spectrum near 600 Hz corresponds to a value for the first difference that is approximately zero and to a value for the second difference that is a local minimum. Panel D shows the smoothed second-order differences; i.e., the edge-enhancement function [see Eq. (3)]. Note that the smoothing process has removed the minor fluctuations that result from the randomness inherent to all noise stimuli. Last, panel E shows the edge-enhanced recovered spectrum [see Eq. (4)]. Note that, for Huggins-pitch and binaural-edge pitch, the frequency location (and the relative height) of the peak in the recovered spectrum is preserved,

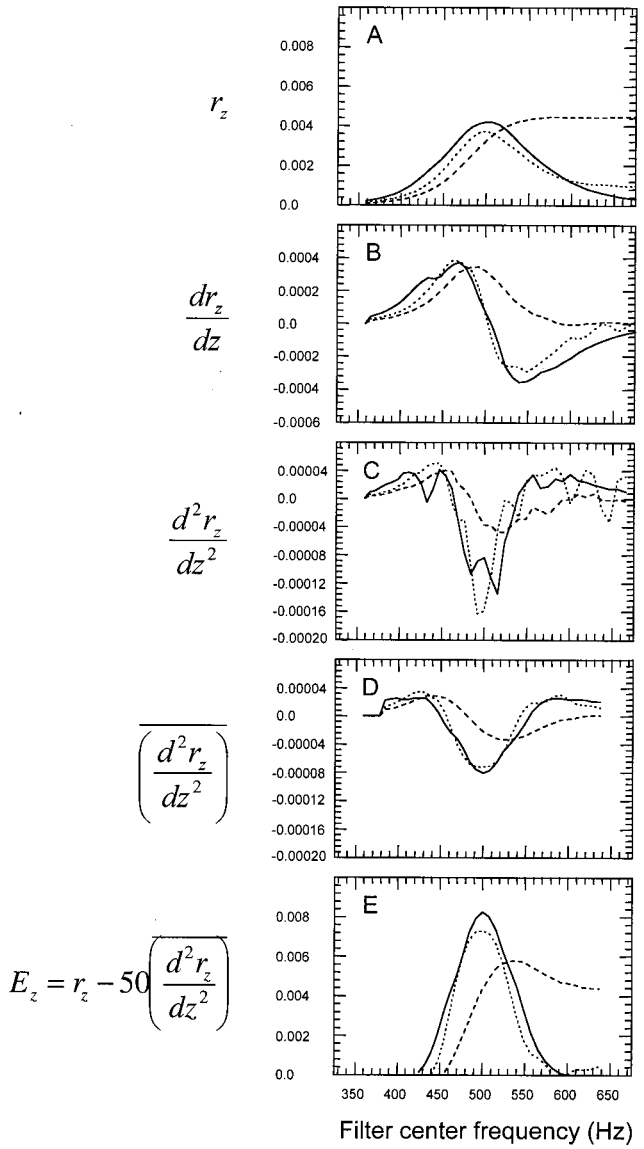


FIG. 4. The results from each step of the ''edge-enhancement'' model. Panel A shows the recovered spectra for a Huggins pitch stimulus (solid line), binaural-edge-pitch stimulus (dotted line), and binaural-coherence-edge-pitch stimulus (dashed line); these are the same as those illustrated in Fig. 2. Panel B shows the first-order differences [see Eq. (1)]. Panel C shows the unsmoothed second-order differences [see Eq. (2)]. Panel D shows the smoothed second-order differences; i.e., the edge-enhancement function [see Eq. (3)]. Panel E illustrates the edge-enhanced recovered spectrum [see Eq. (4)].

but for binaural-coherence-edge pitch, a peak is present in the edge-enhanced recovered spectrum that was not present in the original recovered spectrum.

There are two effects in panel E of Fig. 4 that are of interest. First, the frequency location of the peak in the edge-enhanced recovered spectrum for binaural-coherence-edge pitch is at approximately 540 Hz. This frequency corresponds to a shift of about 8% above the transition frequency of 500 Hz. Such a magnitude of shift is compatible with Hartmann and McMillon's (2001) measurements of the value of the perceived pitch for binaural-coherence-edge pitch.

Second, the magnitude of the peak in the edge-enhanced recovered spectra is greatest for Huggins pitch, marginally

smaller for binaural-edge pitch, and distinctly smaller for binaural-coherence-edge pitch. This ordering of the peak heights resembles the pattern of results in our experiment and is consistent with the assumption that the strength of these dichotic pitches is related to the height of the peak in the enhanced recovered spectrum. It should be noted, however, that this effect is partially determined by the value of the free parameter used in Eq. (4); our choice of 50 was chosen, in part, to obtain this ordering of the three dichotic-pitch stimuli.

V. SUMMARY

(1) We measured identifiability of melodies created using three types of dichotic-pitch stimuli (Huggins pitch, binaural-edge pitch, binaural-coherence-edge pitch) as well as using pure-tone stimuli. The identification scores were consistently high for all of the dichotic-pitch stimuli, demonstrating that the pitch sensations evoked by dichotic-pitch stimuli are ''musical.''

(2) The identification scores were high for the dichotic-pitch stimuli even in the first block of testing, indicating that the sensation of dichotic pitch is relatively immediate and does not require prolonged experience.

(3) Differences in identification scores were observed for the three types of dichotic pitch, Huggins pitch giving the highest scores and binaural-coherence-edge pitch the lowest scores. These differences are consistent with analyses based upon the modified equalization-cancellation model of dichotic pitch, extended to include a model for enhancement of edges in the recovered spectrum. The data are consistent with the idea that the strength of a dichotic pitch is related to the height of a peak in the enhanced recovered spectrum.

ACKNOWLEDGMENTS

We thank Brian Glasberg for his help, and an anonymous reviewer, Frans Bilsen and Sid Bacon (Associate Editor) for their comments in the review process. M.A.A. is supported by a MRC (U.K.) Career Development Award. G.A.M. is supported by a scholarship from Defeating Deafness (U.K.).

[1]Cramer and Huggins (1958) used a bandwidth of 10%, meaning that the width of the interaural phase shift from 90° to 270° was 10% of the center frequency. This definition of bandwidth is usual if, like Cramer and Huggins, the interaural phase shift is constructed using an analog all-pass filter. If, like in the present experiment, the stimuli are constructed digitally in the frequency domain, it is more convenient to define the value of the bandwidth as the width of the interaural phase shift from 0° to 360°. Our bandwidth of 16% thus corresponds to 8% in Cramer and Huggins' terminology.

[2]Listeners were asked to write down their response to each tune *before* the correct answer was given, and then to mark the response as correct or incorrect with a tick or a cross. The experimenter checked that they were doing this as the experiment proceeded. The response sheets were subsequently checked to ensure that the ticks and crosses correctly reflected the tune names that were written down. This was always the case.

Akeroyd, M. A., and Summerfield, A. Q. (**1999**). ''A fully temporal account of the perception of dichotic pitches,'' Br. J. Audiol. **33**, 106–107.

Cramer, E. M., and Huggins, W. H. (**1958**). ''Creation of pitch through binaural interaction,'' J. Acoust. Soc. Am. **30**, 413–417.

Culling, J. F. (**2000**). ''Auditory motion segregation: A limited analogy with vision,'' J. Exp. Psychol. **26**, 1760–1769.

Culling, J. F., Summerfield, A. Q., and Marshall, D. H. (**1998**). ''Dichotic pitches as illusions of binaural unmasking. I. Huggins pitch and the binaural edge pitch,'' J. Acoust. Soc. Am. **103**, 3509–3526.

Durlach, N. I. (**1962**). ''Note on the creation of pitch through binaural interaction,'' J. Acoust. Soc. Am. **34**, 1096–1099.

Durlach, N. I. (**1972**). ''Binaural signal detection: Equalization and cancellation theory,'' in *Foundations of Modern Auditory Theory, Vol. II*, edited by J. V. Tobias (Academic, New York).

Frijns, J. H. M., Raatgever, J., and Bilsen, F. A. (**1986**). ''A central spectrum theory of binaural processing. The binaural edge pitch revisited,'' J. Acoust. Soc. Am. **80**, 442–451.

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. (**1984**). ''Binaural coherence edge pitch,'' J. Acoust. Soc. Am. **75**, S22.

Hartmann, W. M., and McMillon, C. D. (**2001**). ''Binaural coherence edge pitch,'' J. Acoust. Soc. Am. **109**, 294–305.

Julesz, B. (**1971**). *Foundations of Cyclopean Perception* (University of Chicago Press, Chicago, IL).

Klein, M. A., and Hartmann, W. M. (**1981**). ''Binaural edge pitch,'' J. Acoust. Soc. Am. **70**, 51–61.

Kubovy, M., Cutting, J. E., and McGuire, R. M. (**1974**). ''Hearing with the third ear: dichotic perception of a melody without monaural familiarity cues,'' Science **186**, 272–274.

Licklider, J. C. R. (**1959**). ''Three auditory theories,'' in *Psychology, a Study of a Science*, edited by S. Koch (McGraw-Hill, New York).

Meddis, R., Hewitt, M. J., and Shackleton, T. M. (**1990**). ''Implementation details of a computational model of the inner hair-cell/auditory-nerve synapse,'' J. Acoust. Soc. Am. **87**, 1813–1816.

Moore, B. C. J., and Rosen, S. M. (**1979**). ''Melody recognition with reduced pitch and interval information,'' Q. J. Exp. Psychol. **31**, 229–249.

Patterson, R. D., Allerhand, M. H., and Giguère, C. (**1995**). ''Time-domain modeling of peripheral auditory processing: A model architecture and a software platform,'' J. Acoust. Soc. Am. **98**, 1890–1894.

Raatgever, J., and Bilsen, F. A. (**1986**). ''A central spectrum theory of binaural processing. Evidence from dichotic pitch,'' J. Acoust. Soc. Am. **80**, 429–441.

Siegel, S., and Castellan, N. J., Jr. (**1988**). *Nonparametric Statistics for the Behavioral Sciences* (McGraw-Hill, New York).

Slaney, M. (**1998**). ''Auditory Toolbox: A MATLAB toolbox for auditory modeling work, Version 2,'' Technical Report No. 1998-010 (Interval Research Corporation).

Small, A. M., and Daniloff, R. G. (**1967**). ''Pitch of noise bands,'' J. Acoust. Soc. Am. **41**, 506–512.

von Békésy, G. (**1959**). ''Neural funneling along the skin and between the inner and outer hair cells of the cochlea,'' J. Acoust. Soc. Am. **31**, 1236–1249.

Peripheral auditory processing and investigations of the "precedence effect" which utilize successive transient stimuli

Klaus Hartung and Constantine Trahiotis[a)]
Department of Neuroscience and Department of Surgery (Otolaryngology), University of Connecticut Health Center, Farmington, Connecticut 06030

(Received 18 January 2001; revised 10 May 2001; accepted 22 May 2001)

This article addresses how a consideration of peripheral auditory processing can help to understand experiments concerning binaural precedence that employ successive binaural transients. It appears that much of the patterning of the behavioral data is amenable to an explanation based on peripheral interactions that result from auditory filtering and the functioning of auditory hair cells in combination with a binaural model based on cross correlation. A noteworthy aspect of this approach is that it does not include inhibitory mechanisms like those commonly invoked to explain binaural precedence. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1390339]

PACS numbers: 43.66.Pn, 43.66.Ba, 43.66.Rq, 43.66.Qp [SPB]

I. INTRODUCTION

In their classic earphone study concerning what they termed the ''precedence effect,'' Wallach *et al.* (1949) employed pairs of successive binaural transients (or ''clicks''). The word precedence was used to connote their finding that the interaural temporal differences (ITDs) conveyed by the first binaural pair of transients dominated the intracranial location of the unitary image produced by the composite stimulus. Since then, many investigators have employed successive transients in order to investigate a wide variety of stimuli and phenomena that are now collectively referred to as studies of ''binaural precedence'' (e.g., Zurek, 1980; Yost and Soderquist, 1984; Aoki and Houtgast, 1992; Houtgast and Aoki, 1994; Shinn-Cunningham *et al.*, 1995; Tollin and Henning, 1998, 1999). The important empirical and theoretical findings can be found in three comprehensive reviews (Zurek, 1987; Blauert, 1997; Litovsky *et al.*, 1999). Those reviews contain information concerning the precedence effect, in particular, and binaural information processing, in general.

The relative dominance of the first binaural pair of transients has been explained by inferring that the binaural information conveyed by the second pair of binaural transients (or the ''echo'') is somehow suppressed or physiologically inhibited. In fact, it is now common to find words like ''inhibition'' and ''echo suppression'' within descriptions and explanations of the variety of phenomena covered by ''the precedence effect'' (e.g., Lindemann, 1986a, b; Zurek, 1987; Litovsky *et al.*, 1999). This point of view differs from that of Wallach *et al.* (1949) who were careful to differentiate between the dominant role in localization played by the first-arriving sounds and the suppression of echoes.

The principal purpose of this report is to point out, describe, and evaluate how monaural, peripheral, auditory processing may affect behavioral data obtained in binaural ''precedence'' experiments in which ITDs are conveyed by successive pairs of binaural transients. For example, we show that, when the time between the pairs of transients within a single ear or channel is below 2 to 3 ms, as is the case in many of the experiments under discussion, relevant peripheral interactions can occur within an auditory filter. Such interactions result in internal, effective ITDs [and, concomitantly, interaural intensitive differences (IIDs)] that are different from those imposed on the external stimuli. Such interactions occur when the time between the monaural inputs is effectively shorter than the reciprocal of the bandwidth of the auditory filter (see Jeffress and McFadden, 1968). Our approach also incorporates compression and adaptation as characterized by the Meddis hair cell model (Meddis, 1986).

We show that such peripheral interactions are manifested in the binaural cross correlation of the stimuli and that the across-frequency averaged cross correlation can account for the forms of the data obtained in precedence experiments that employ transient or ''short'' stimuli separated by 1 ms or so, a time short enough to produce only one intracranial image.

A noteworthy feature of this new approach is that it does not rely on any central inhibitory mechanisms such as the ones proposed by Lindemann (1986a, b) or Zurek (1987). It is historically interesting that Harris *et al.* (1963) investigated how peripheral auditory processing would affect the internal representation of successive transient stimuli. Their analysis differs from the one presented here because they only considered outputs of one ear. Therefore, they could not address how peripheral interactions would affect the internal effective values of the ITDs and IIDs produced by their stimuli, which is the focus of this study.

II. THE PERIPHERY AND PRECEDENCE

In order to explain our approach, let us begin by considering how successive binaural transients can interact within auditory filters. The auditory filters will be represented by a gammatone filter bank, which is now commonly used in a variety of models to represent peripheral auditory filtering (e.g., Patterson *et al.*, 1995). Our argument can be general-

a)Electronic mail: tino@neuron.uchc.edu

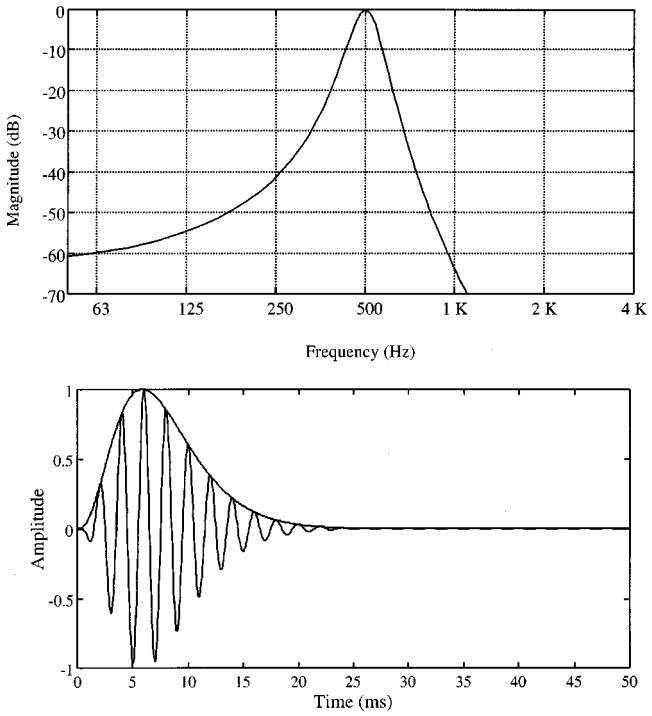


FIG. 1. The upper panel displays the logarithmic amplitude of the frequency response of a gammatone filter centered at 500 Hz and having a 3-dB-down bandwidth of 75 Hz. The lower panel displays the impulse response of this filter along with a line that indicates the envelope of the output.

ized to other types of filters, however, because the interactions being discussed depend heavily upon only the bandwidth of the filters.

In order to understand the within-filter interactions, consider that the duration of the impulse response of a filter is inversely related to its bandwidth. Consequently, interactions are to be expected when a second input is applied while the filter is still responding to the first input (i.e., when the impulse responses of the successive inputs at least partially overlap). Our analysis shows that such interactions result in internal ITDs and IIDs that differ from their external counterparts. It is important to note that the phase response of the filter does not affect the IIDs and ITDs present after filtering. This is so because the phase response of a filter determines the delay imposed on its input. Therefore, any two identical filters assumed to process left and right ear signals, respectively, would impose the same absolute amount of delay on their inputs, but would not impose a differential delay (i.e., an ITD). Our computer simulations revealed that factors like the particular phase response of the filters and the steepness of their rejection slopes have extremely little or no effect on the phenomena being considered here.

Let us now consider the types of interactions that can occur when successive transients are applied to a bank of gammatone filters. We implemented each filter as a cascade of four second-order IIR filters according to the procedure discussed by Slaney (1993). The upper panel of Fig. 1 displays the logarithmic magnitude of the frequency response of one such filter centered at 500 Hz and having a 3-dB-down bandwidth of 75 Hz. The lower panel displays the impulse response of this filter, with a separate line indicating the envelope of the output. Note that the output of the filter produced by a transient input lasts for up to 25 ms or so, and, for this particular value of bandwidth, the maximum output is reached about 6 ms after the occurrence of the impulsive input.

The panels in Figs. 2 and 3 illustrate the essence of our argument. Figure 2 displays the outputs of a filter centered at 500 Hz, like the one illustrated in Fig. 1. Figure 3 displays the output for a filter centered at 676 Hz, which is two ERB units (Glasberg and Moore, 1990) above its counterpart at 500 Hz and has a slightly larger bandwidth. These two filters, although not too different in center frequency, illustrate differing interactions that can occur simultaneously in neighboring spectral regions.

The two upper portions of Fig. 2 display the outputs of a pair of ''left'' and ''right'' filters when the input to each is a pair of successive transients. As indicated at the top, the first pair of binaural transients contains an ITD of 0 μs and the second pair contains an ITD of 200 μs, with the left ear leading. Within each panel, the time between the pairs of transients is *nominally* 1 ms for the leftmost column, 2 ms for the center column, and 3 ms for the rightmost column. For ease of exposition, we identify the different conditions by specifying their nominal interclick interval (ICI) and the ITD imposed on the second pair of inputs. [Of course, one must include the ITD to describe accurately the actual temporal separation between the transients in the (lagging) right ear.]

The instantaneous values of ITD and IID that result from within-filter interactions are plotted in the third and fourth rows of the figure. They were obtained from the analytic signals that resulted from the application of the filter (see, for example, Panter, 1965; Colburn and Isabelle, 2001). In order to obtain the instantaneous values of ITD, we calculated the interaural phase differences from the scalar products and vector products of the left and right analytic signals. The values of ITD were then obtained by transforming these instantaneous differences of phase according to the equation $\mathrm{ITD}=\Delta\Phi/(2\pi f)$, where f is the center frequency of the auditory filter. The values of IID are expressed in dB as 20 times the logarithm of the ratio of the left and right envelope functions.

Beginning at the leftmost column of Fig. 2 with the pairs of transients being separated by 1 ms, it can be seen in the upper two plots of the figure that the ITD of the second pair of transients produces outputs of the two filters that differ from one another. Also note that the output of the left filter is attenuated, re the output of the right filter. As shown in the third and fourth rows of that column, the outputs contain both ITDs and IIDs that change as a function of time. Some instantaneous values of ITD are *negative*, having magnitudes of up to 380 μs. Negative values of ITD signify that the right channel is leading, which contradicts the information conveyed by the second pair of transients. There also occur *positive* instantaneous values of ITD which favor the leading ear by up to 600 μs. Clearly, the 200-μs value of ITD conveyed by the second pair of transients, which is indicated by the arrow, is not well represented by the instantaneous values of ITD measured at the outputs of the filters. Note also that the outputs of the filters result in instantaneous values of IID of

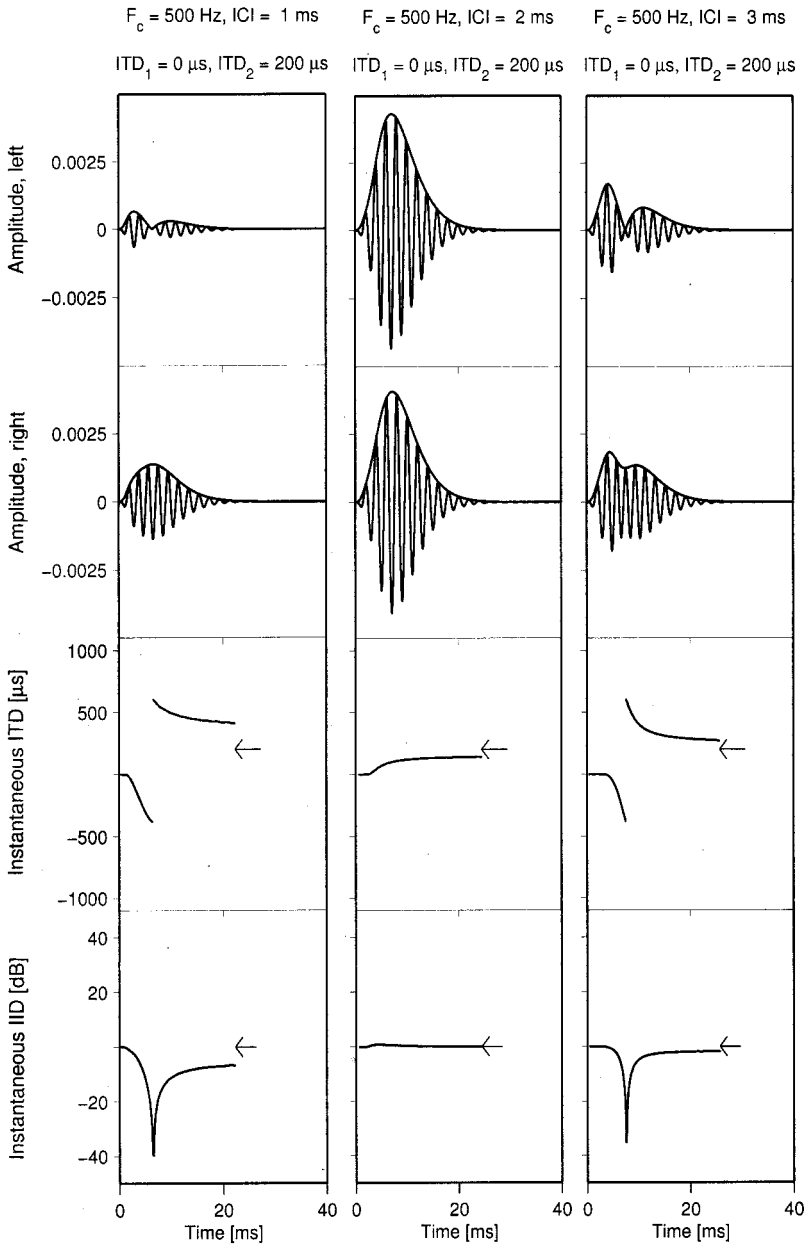


FIG. 2. The outputs of a gammatone filter, like the one shown in Fig. 1, for inputs that are pairs of binaural transients having monaural interclick intervals (ICIs) of either 1, 2, or 3 ms, as defined by the time between the onsets of pairs of monaural inputs. As indicated at the top of the figure, the ITD conveyed by the second pair of inputs was 200 μs, with the left ear leading. The upper portion of the figure displays the outputs of a pair of ''left'' and ''right'' filters when the input to each is the pair of successive transients described at the top of each column. The third and fourth rows of each column show the instantaneous values of ITDs and IIDs as measured after filtering.

up to -40 dB, which reduce rapidly to about -7 dB over the duration of the response. This occurs despite the fact that the IID conveyed by the inputs was 0 dB. Thus, overall, our analysis indicates that the instantaneous ITDs and IIDs at the outputs of the filters are, at times, substantially different in magnitude and/or in sign from those present at the inputs.

The center column of Fig. 2 shows interactions that occur when the time between the pairs of transients is 2 ms. For this case, the magnitude of the output of both left and right filters is larger than was the case when the time between the pairs of transients was 1 ms. This can be understood by considering that the responses of the filter to the successive transients are more or less ''in phase'' (i.e., the outputs add constructively). This is so because the interval of 2 ms between the transient inputs is the period of the 500-Hz center frequency of the filter. In this case, the instantaneous values of ITD are consistently quite small and, in the limit, approach only 140 μs. That is, all of the instantaneous values of ITD are well below the ITD of 200 μs, which is the value of ITD conveyed by the second pair of binaural transients. For this stimulus condition, the magnitudes of the instantaneous values of IID are always less than 1 dB, a value very close to the 0-dB IID carried by the inputs.

Finally, when the time between the pairs of transients is 3 ms (a condition depicted by the rightmost column of Fig. 2), the instantaneous values of ITD and IID are very large. Although the patternings of the instantaneous values of ITD and IID are similar to those found when the interval between the transients was 1 ms, in this case the ''terminal'' value of ITD is about 290 μs. That ITD is larger than the one present in the input, but smaller than the largest ITDs found when the ICI was 1 ms. In this stimulus condition the maximal IIDs are again very large, being on the order of -35 dB.

Figure 3 displays the outputs of a pair of left and right filters centered at 676 Hz. This figure is included to show that auditory filters differing slightly in center frequency can produce dynamic instantaneous values of ITD and IID that differ greatly from one another. Note that the magnitudes of

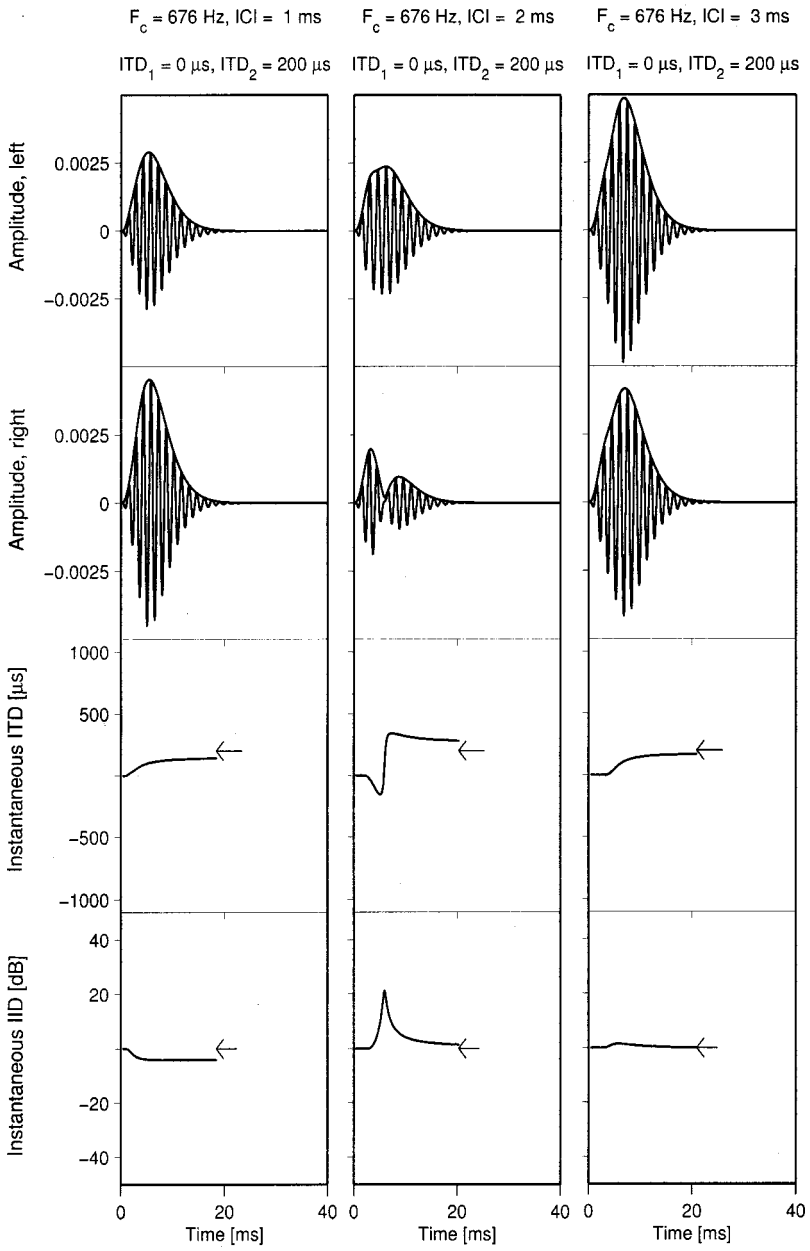


FIG. 3. Same as Fig. 2, but the center frequency of the filter is 676 Hz.

the outputs of the filters centered at 676 Hz that occur when the time between the input transients is 3 ms are fairly similar to the magnitudes of the outputs for the filters centered at 500 Hz when the time between the input transients is 2 ms. Also, the magnitudes of the outputs of the filters centered at 676 Hz that occur when the time between the input transients is 2 ms are fairly similar to the magnitudes of the outputs for the filters centered at 500 Hz when the time between the input transients is 3 ms. In this case, however, the instantaneous values of IID differ from one another in sign. This difference is a consequence of the differing monaural magnitudes of the responses in the left and right filters across the two cases. Thus, it is clear that diverse internal effective ITDs and IIDs are produced by peripheral auditory filtering.

Let us now consider how the outputs of the bank of auditory filters can be included in a cross-correlation-based model of binaural processing in order to account for some of the classical behavioral data concerning precedence. The model is functionally similar to the one described recently by Trahiotis *et al.* (2001), which capitalizes on the work of Stern and Colburn (1978), Blauert and Cobben (1978), and Shackleton *et al.* (1992). The general model provides quantitative, interaural-correlation-based, predictions of binaural detection (e.g., Bernstein and Trahiotis, 1996a, b; Bernstein *et al.*, 1999) and discrimination of ITD (Trahiotis *et al.*, 2001).

For our purposes, pairs of binaural transients were passed through a bank of gammatone filters like the ones discussed earlier. The outputs of the filters served as inputs to the Meddis hair cell model (Meddis, 1986, 1988), which we implemented in the manner described by Meddis *et al.* (1990) and by Slaney (1993). Computer simulations revealed that the Meddis hair cell model improved prediction of the behavioral data. The outputs of the hair cell model are rectified, low-pass filtered, and compressed versions of the bandpass-filtered stimuli which serve as inputs. The inclusion of the hair cell model results in both static and dynamic types of compression. It is important to note (1) that the

dynamic part of the compression serves to reduce the magnitude of the output over time in a manner that functionally provides ''adaptation'' or ''attenuation'' of inputs arriving soon after initial activation and (2) that the reduction of magnitude does not directly affect the instantaneous values of the binaural cues, which, because of prior filtering, are different from those present in the physical stimuli. These two properties differ from transformations produced by central inhibitory processes like those often postulated to account for the precedence effect (e.g., Lindemann, 1986a, b; Litovsky *et al.*, 1999). Yet, both approaches yield the dominant role played by initial inputs.

The filterbank was composed of 14 filters having center frequencies spanning the range from 244.7 to 1690 Hz. The bandwidths of the filters were taken from Glasberg and Moore (1990) and the center frequencies of the filters were spaced by their respective equivalent-rectangular bandwidths (i.e., according to Glasberg and Moore's ''ERB'' function). This resulted in a bank of filters having center frequencies that were spaced at 1-ERB intervals.

A cross correlogram (i.e., cross-correlation surface) was constructed with values of delay (τ) spanning the range from -1500 to 1500 μs in 10-μs steps. Filter-specific cross-correlation functions were computed by taking 30-ms lengths of the filtered signals (a duration long enough to include information at the output of the narrowest filters), shifting them by the correlation delay, multiplying the shifted waveforms, and then integrating the product over the whole time interval. Finally, the filter-specific cross-correlation functions were added together for each correlation delay, so that all 14 channels were summed with equal weighting. This type of across-frequency averaged cross correlogram is consistent with Shackleton *et al.* (1992) and Trahiotis *et al.* (2001) in that it considers performance in terms of the location of the most central peak.

Because we were interested in assessing effects resulting principally from peripheral processing, we did not include either (1) a function that emphasizes activity at or very near midline (Blauert and Cobben, 1978; Stern and Colburn, 1978; Shackleton *et al.*, 1992; Stern and Shear, 1996) or (2) across-frequency weighting such as Stern *et al.*'s (1988) function which describes the relative spectral dominance within binaural processing (Bilsen and Raatgever, 1973).

III. ACCOUNTING FOR PRECEDENCE

Let us now evaluate how the model can account for data obtained in three often-cited precedence experiments employing successive pairs of short stimuli: Wallach *et al.* (1949), Yost and Soderquist (1984), and Shinn-Cunningham *et al.* (1995). In their classic study, Wallach *et al.* reported combinations of ITDs in the first and second pairs of 1-ms-long ''clicks'' that were required for the listeners to report a single intracranial image that was perceived at midline. The pairs of clicks were separated by 2 ms as measured from the beginning of the leading input in the first pair of transients to the beginning of the leading input in the second pair of transients. The data from their two listeners are shown in Fig. 4 along with the predictions from our model. Overall, the predictions are satisfactory in that they fall between the data

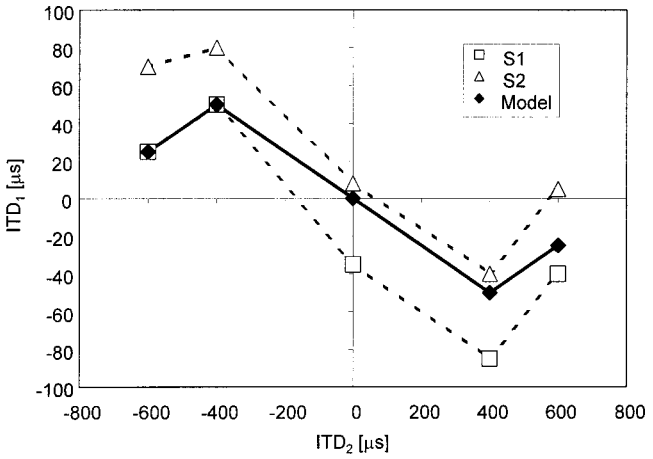


FIG. 4. Combinations of ITD conveyed by the first pair of binaural transients and ITD conveyed by the second pair of transients required to produce a midline intracranial image for the total stimulus, as reported by Wallach *et al.* (1949). The open symbols represent the data obtained for their two subjects and the closed diamond represents the predictions of the model.

obtained from each of the two listeners. The major discrepancy between the predictions and the data stems from the fact that the data are asymmetric around midline and the model's predictions are symmetric. Quantitatively, 89% of the variability in the data obtained from the two listeners can be accounted for by the model.

In order to illustrate the influence of the Meddis hair cell model, cross correlograms are plotted in Fig. 5 both with and without the hair cell model for one of Wallach *et al.*'s stimulus conditions. For this example, the ITD of the first pair of binaural transients is 30 μs and the ITD of the second pair is -600 μs. This stimulus condition is similar to the ones that produced a midline intracranial image for the two listeners.

The top portion of Fig. 5 contains a binaural correlogram (at the left) and its across-frequency average (at the right). These two representations show central activity as measured when the model includes filtering, rectification, and cross correlation, but does not include processing by hair cells. The ''crosses'' in each correlogram indicate the positions of the peaks of the cross-correlation functions computed at each center frequency. The amount of activity within each frequency channel is illustrated by colors. The reddish areas indicate a relatively high level of activity whereas the dark blue areas indicate a relatively low level of activity. Note that the largest amounts of activity occur at frequencies at and below 500 Hz. The peaks of activity for those frequencies occur at an ITD of about -290 μs. Note that there are also two islands of medium activity (light blue) at about 600–700 Hz, one at ITDs around 500 μs, and another at ITDs around -900 μs. Those two peaks of activity in the correlogram illustrate strong interactive effects due to filtering. It is conceivable that such activity could provide useful cues in investigations of precedence that utilize discrimination paradigms. As indicated at the right of the figure, however, the peak of the across-frequency averaged activity for this pair of inputs occurs at about -270 μs. Therefore, the model without hair cells incorrectly predicts that the stimulus would be heard toward the right ear.

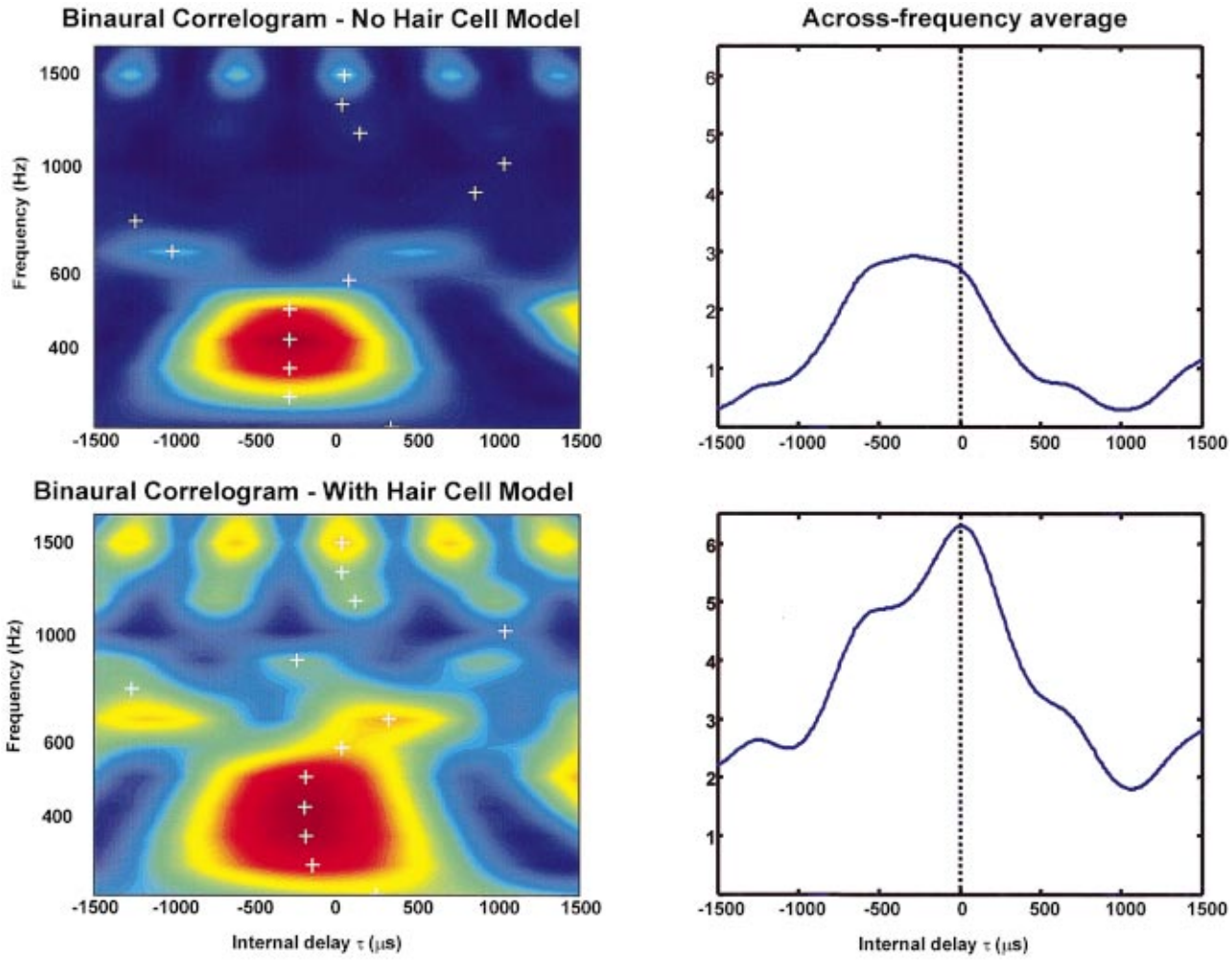


FIG. 5. The left-hand portion of the figure shows cross correlograms computed at the outputs of the gammatone filterbank. The top correlogram was computed after filtering and half-wave rectification of the inputs. The bottom correlogram was computed after filtering and processing via the Meddis hair cell model. The ''crosses'' in each correlogram indicate the positions of the peaks of the cross-correlation functions computed at each center frequency. The amount of activity within each frequency channel is illustrated by colors. The reddish areas indicate a relatively high level of activity whereas the bluish areas indicate a relatively low level of activity. The right-hand portion of the figure shows the across-frequency average of each correlogram shown at the left.

The bottom portion of Fig. 5 shows that including the hair cell model yields an accurate prediction. The addition of the hair cell model results in an increase in activity in the low-frequency region that extends to include 700 Hz. This frequency region displayed relatively lower amounts of activity when the hair cells were not included. Turning to the across-frequency averaged correlogram, note that the expanded model correctly predicts a peak at midline, where the stimuli were perceived by the listeners. A comparison of the correlograms obtained with and without the hair cell model reveals that adding them generally increased relative activity and shifted the peaks of the cross correlation toward the midline for filters centered at and below 700 Hz. The general increase in activity results from the compressive action of the hair cell. The shifting of the peaks of the cross correlation results from adaptation effects which functionally produce an increased relative weighting of the initial inputs to the hair cell. Note that there are virtually no peaks of activity near midline. Nevertheless, across-frequency averaging of the cross correlogram results in a peak at midline, in accord with the behavioral data.

Figure 6 contains the average of the data obtained by Yost and Soderquist (1984) in an experiment similar to the one reported by Wallach *et al.* Yost and Soderquist employed 100-μs-long pairs of ''clicks'' that were nominally separated by 1 ms. The overall patterning of their data also appears to be captured by the model. Quantitatively, only about 40% of the variability in Yost and Soderquist's data are accounted for. The largest discrepancy between the predictions and data occur when the ITD in the second pair of transients is ± 200 μs, a value not included in the Wallach *et al.* study. Presently, we cannot explain the discrepancy between the predictions and the data for that stimulus. However, when we listened to the stimuli we heard images at midline, just as Yost and Soderquist's listeners did. Clearly, the prediction of the model for that stimulus is incorrect. If those two data points are omitted, however, then 85% of the variability in the data is accounted for by the model. That value is very close to the 89% of the variance that was accounted in the Wallach *et al.* data.

Finally, the two panels of Fig. 7 contain data obtained by Shinn-Cunningham *et al.* (1995) in an experiment employing 300-Hz-wide bands of noise centered on 450 Hz (what they term their ''L'' stimulus) and 1250 Hz (what they term their

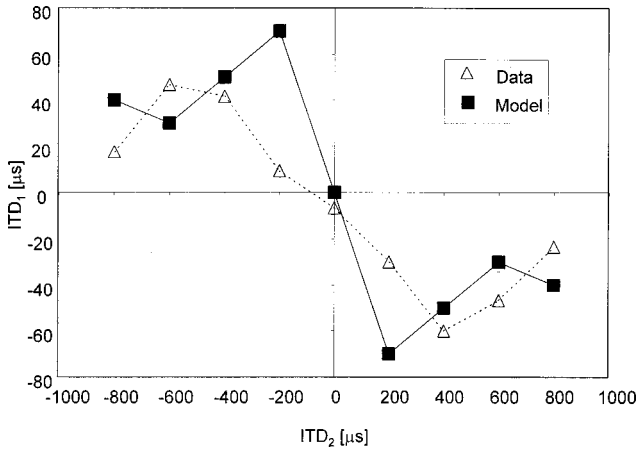


FIG. 6. Combinations of ITD conveyed by the first pair of binaural transients and ITD conveyed by the second pair of transients required to produce a midline intracranial image, as reported by Yost and Soderquist (1984). The open symbols represent the average data obtained from their three subjects and the closed squares represent the predictions of the model.

''H'' stimulus). Both low- and high-frequency bands of noise were temporally shaped by a 3-ms Hanning window to form a ''burst,'' in contrast to the more ''transient'' stimuli employed in the two studies discussed above. Shinn-Cunningham *et al.*'s listeners adjusted the ITD of an acoustic pointer in order to match the intracranial location of the stimulus. In order to evaluate within-filter effects, we only consider data obtained with stimuli composed of components having identical spectra and onsets separated nominally by 1 ms (i.e., Shinn-Cunningham *et al.*'s L-L and H-H conditions).

Visual inspection of Fig. 7 reveals that the fits to the data are good to excellent. Quantitatively, the model accounted for 77% of the data obtained in the L-L condition and 97% of the data obtained in the H-H condition. The fit to the data in the L-L condition appears to be slightly attenuated by an asymmetry in the behavioral data.

IV. GENERAL DISCUSSION

We have presented logical arguments and quantitative analyses of behavioral data that point to the importance of peripheral auditory processing as it affects the perception of successive transients in experiments concerning binaural ''precedence.'' Overall, it appears that a large portion of the variability in such experiments can be accounted for without including special central inhibitory processes and without relying upon ''top-down'' mechanisms in which some sort of selective attention plays a role. This is not to imply that central inhibitory processes and/or selective attention cannot operate in the wide variety of experiments that are included in the term ''precedence.'' Rather, our analysis reveals that considerations of effects due to the periphery are sufficient to account for data obtained in precedence experiments employing successive pairs of transients. The degree to which peripheral auditory processing can account for data obtained with longer stimuli, such as trains of clicks, speech, and music, is an open question.

It is important to note that we have not attempted to account for the data by manipulating values of the parameters of the model in order to maximize the accuracy of the predictions. In fact, we believe one of the appealing aspects of our approach is that only the general nature of the stages of processing was included and no ''curve fitting'' was employed.

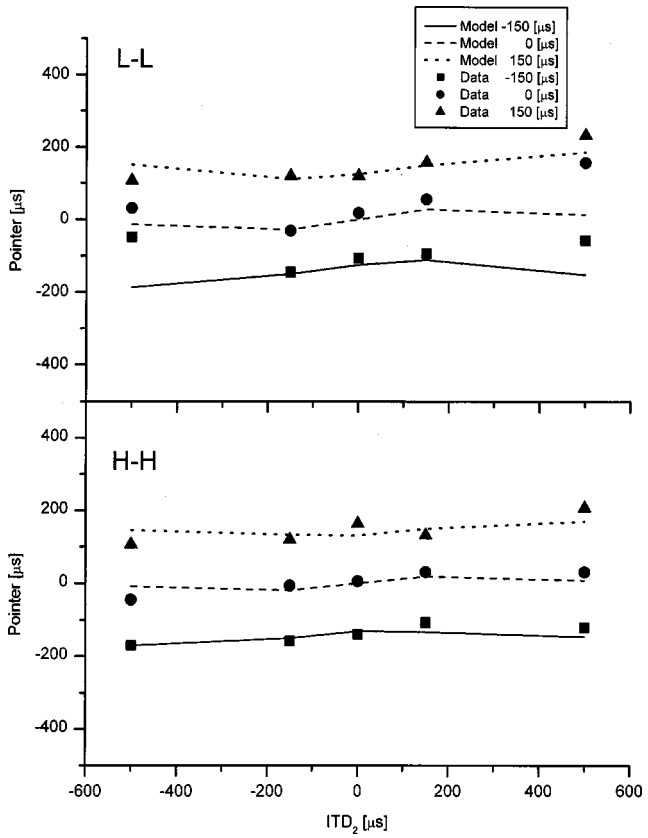


FIG. 7. The ITDs of an acoustic pointer that matched the intracranial position of the bands of noise in an experiment conducted by Shinn-Cunningham *et al.* (1995), as a function of the ITD conveyed by the second pair of binaural stimuli. The letters L and H refer to low- and high-frequency bands of noise. The behavioral data are shown by the closed symbols and the predictions are shown by the lines. The ITDs conveyed by the first pair of binaural stimuli are indicated within the insert at the top right of the figure, along with the symbols that depict them.

In closing, it is appropriate to consider three issues. First, our approach should be clearly differentiated from the one recently put forth by Tollin (1998), which also does not include central inhibitory factors. Tollin does not consider interactive effects *within* peripheral filters, *per se*. Instead, he assumes the interaction of the first and second pairs of binaural transients occurs via a central temporal weighting function. Another difference is that Tollin accounts for precedence by focusing upon a single auditory filter that is construed to be, depending on the particular stimulus, strategically placed such that its spectral components would dominate perception. Our approach is more general and does not depend upon the nature of the activity that occurs in any particular spectral region. In our view, Tollin's approach and ours are not mutually exclusive. Indeed, it may turn out to be the case that both approaches will have to be merged in some manner and, possibly, supplemented by some central inhibition in order to account for the diverse data comprising the binaural precedence literature.

The second issue we wish to address concerns recent

physiological investigations which employed successive binaural transients in attempts to discover physiological correlates of precedence (e.g., Yin, 1994; Fitzpatrick *et al.*, 1995; Litovsky *et al.*, 1999). We believe our findings place an important constraint on the interpretation of such data. Those studies focused upon measurements of responses from single neural units following stimulation by successive pairs of binaural transients like the ones we have considered. A general finding is that neural units which are ''tuned'' to particular values of interaural delay appear to show severely diminished responses to their ''best-ITD'' when priorly stimulated by another pair of binaural transients conveying an ITD to which the neuron does not vigorously respond (what Fitzpatrick *et al.* term the ''worst/best'' case).

Our analysis suggests that the diminished responses obtained when the ITDs were presented in the ''worst/best'' configuration can be accounted for by peripheral auditory processing. It appears that the interaction of the successive transients within the auditory periphery produces internal values of ITDs (and most likely IIDs) that can be ''outside the tuning range'' of the single unit being studied. Said differently, especially for short interclick intervals, one would expect diminished neural responses simply because the internal instantaneous ITDs and IIDs do not provide favorable stimulation.

The third issue concerns the limitation of our findings. In the current literature, the term ''precedence'' is used to connote investigations covering diverse stimuli and psychophysical procedures, and spans a wide variety of phenomena including localization dominance, binaural fusion, echo suppression, and ''build-up'' effects (Litovsky *et al.*, 1999). Some of those studies employ long interstimulus intervals and, therefore, their results cannot be explained adequately by considering only peripheral processing. Clearly, additional, central factors might be more or less influential, depending on the exact nature of each study.

V. CLOSING COMMENTS

We have shown that predictions of binaural performance in three important precedence experiments employing successive pairs of transients can be accounted for by considering effects resulting from peripheral auditory processing in combination with a binaural model based on cross correlation. The approach entails combining stages of processing that, taken singly, would not be sufficient to predict performance. Up until now, explanations of behavioral data concerning precedence have been focused upon central, rather than peripheral, types of mechanisms including central inhibition, echo suppression, and selective attention. In spirit, our approach follows that pioneered by Colburn (1973), who advocated that inputs to a central binaural processor should be considered as external stimuli that have been transformed by peripheral auditory processing. It is noteworthy that the stages in the peripheral model that we employ, and which account for much of the variability in the behavioral data, were only recently amenable to quantitative description. That is, our analysis relies heavily on the relatively recent work of colleagues whose principal interest was understanding the physiological functioning of the monaural peripheral system. We find it gratifying that important studies of binaural precedence can be understood by combining knowledge obtained from prior behavioral and physiological studies of auditory functioning.

ACKNOWLEDGMENTS

This work was supported by Research Grant No. NIH DC-00234 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health. The authors wish to thank Dr. Les Bernstein, Dr. Dan Tollin, Dr. Barbara Shinn-Cunningham, Dr. Sid Bacon, Dr. Steve Colburn, and one anonymous reviewer for their helpful comments. Special thanks are due Dr. Michael Akeroyd for contributing software and suggestions that were necessary for the computer-based analyses.

Aoki, S., and Houtgast, T. (**1992**). ''A precedence effect in the perception of inter-aural cross correlation,'' Hear. Res. **59**, 25–30.

Bernstein, L. R., and Trahiotis, C. (**1996a**). ''On the use of the normalized correlation as an index of interaural envelope correlation,'' J. Acoust. Soc. Am. **100**, 1754–1763.

Bernstein, L. R., and Trahiotis, C. (**1996b**). ''The normalized correlation: Accounting for binaural detection across center frequency,'' J. Acoust. Soc. Am. **100**, 3774–3784.

Bernstein, L. R., van de Par, S., and Trahiotis, C. (**1999**). ''The normalized correlation: Accounting for NoSpi thresholds obtained with Gaussian and 'low-noise' masking noise,'' J. Acoust. Soc. Am. **106**, 870–876.

Bilsen, F. A., and Raatgever, J. (**1973**). ''Spectral dominance in binaural lateralization,'' Acustica **28**, 131–132.

Blauert, J. (**1997**). *Spatial Hearing—The Psychophysics of Human Sound Source Localization* (MIT, Cambridge, MA).

Blauert, J., and Cobben, W. (**1978**). ''Some consideration of binaural cross correlation analysis,'' Acustica **39**, 96–103.

Colburn, H. S. (**1973**). ''Theory of binaural interaction based on auditory-nerve data I. General strategy and preliminary results on interaural discrimination,'' J. Acoust. Soc. Am. **54**, 1458–1470.

Colburn, H. S., and Isabelle, S. K. (**2001**). ''Physiologically based models of binaural detection,'' in *Physiological and Psychophysical Bases of Auditory Function*, edited by D. J. Breebart *et al.* (Shaker, Maastricht, The Netherlands), pp. 161–168.

Fitzpatrick, D. C., Kuwada, S., Batra, R., and Trahiotis, C. (**1995**). ''Neural responses to simple, simulated echoes in the auditory brainstem of the unanesthetized rabbit,'' J. Neurophysiol. **74**, 2469–2486.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.

Harris, G. G., Flanagan, J. L., and Watson, B. J. (**1963**). ''Binaural interaction of click with a click pair,'' J. Acoust. Soc. Am. **35**, 672–678.

Houtgast, T., and Aoki, S. (**1994**). ''Stimulus-onset dominance in the perception of binaural information,'' Hear. Res. **72**, 29–36.

Jeffress, L. A., and McFadden, D. (**1968**). ''MLD's and the phase angle, alpha,'' J. Acoust. Soc. Am. **43**, 164.

Lindemann, W. (**1986a**). ''Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals,'' J. Acoust. Soc. Am. **80**, 1608–1622.

Lindemann, W. (**1986b**). ''Extension of a binaural cross-correlation model by contralateral inhibition. II. The law of the first wavefront,'' J. Acoust. Soc. Am. **80**, 1623–1630.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). ''The precedence effect,'' J. Acoust. Soc. Am. **106**, 1633–1654.

Meddis, R. (**1986**). ''Simulation of mechanical to neural transduction in the auditory receptor,'' J. Acoust. Soc. Am. **79**, 702–711.

Meddis, R. (**1988**). ''Simulation of auditory-neural transduction: Further studies,'' J. Acoust. Soc. Am. **83**, 1056–1063.

Meddis, R., Hewitt, M. J., and Shackleton, T. M. (**1990**). ''Implementation details of a computational model of the inner hair-cell auditory-nerve synapse,'' J. Acoust. Soc. Am. **87**, 1813–1816.

Panter, P. F. (**1965**). *Modulation, Noise and Spectral Analysis* (McGraw–Hill, New York).

Patterson, R. D., Allerhand, M. H., and Giguere, C. (**1995**). ''Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform,'' J. Acoust. Soc. Am. **98**, 1890–1894.
Shackleton, T. M., Meddis, R., and Hewitt, M. J. (**1992**). ''Across frequency integration in a model of lateralization,'' J. Acoust. Soc. Am. **91**, 2276–2279.
Shinn-Cunningham, B. G., Zurek, P. M., Durlach, N. I., and Clifton, R. K. (**1995**). ''Across-frequency interactions in the precedence effect,'' J. Acoust. Soc. Am. **98**, 164–171.
Slaney, M. (**1993**). ''An efficient implementation of the Patterson-Holdsworth auditory filter bank,'' Apple Computer Technical Report #35.
Stern, R. M., and Colburn, H. S. (**1978**). ''Theory of binaural interaction based on auditory-nerve data. IV. A model for subjective lateral position,'' J. Acoust. Soc. Am. **64**, 127–140.
Stern, R. M., and Shear, G. D. (**1996**). ''Lateralization and detection of low-frequency binaural stimuli: Effects of distribution of interaural delay,'' J. Acoust. Soc. Am. **100**, 2278–2288.
Stern, R. M., Zeiberg, A. S., and Trahiotis, C. (**1988**). ''Lateralization of complex binaural stimuli: A weighted image model,'' J. Acoust. Soc. Am. **84**, 156–165.
Tollin, D. J. (**1998**). ''Computational model of the lateralization of clicks and their echoes,'' in *Proceedings of the NATO Advanced Study Institute on Computational Hearing*, edited by S. Greenberg and M. Slaney, pp. 77–82.
Tollin, D. J., and Henning, G. B. (**1998**). ''Some aspects of the lateralization of echoed sound in man. I. The classical interaural dealy-based effect,'' J. Acoust. Soc. Am. **104**, 3030–3038.
Tollin, D. J., and Henning, G. B. (**1999**). ''Some aspects of the lateralization of echoed sound in man. II. The role of stimulus spectrum,'' J. Acoust. Soc. Am. **105**, 838–849.
Trahiotis, C., Bernstein, L. R., and Akeroyd, M. A. (**2001**). ''Manipulating the 'straightness' and 'curvature' of patterns of interaural cross correlation affects listener's sensitivity to changes in the interaural delay,'' J. Acoust. Soc. Am. **109**, 321–330.
Wallach, H., Newman, E. B., and Rosenzweig, M. R. (**1949**). ''The precedence effect in sound localization,'' Am. J. Psychol. **52**, 315–336.
Yin, T. C. T. (**1994**). ''Physiological correlates of the precedence effect and summing localization in the inferior colliculus of the cat,'' J. Neurosci. **14**, 5170–5186.
Yost, W. A., and Soderquist, D. R. (**1984**). ''The precedence effect: Revisited,'' J. Acoust. Soc. Am. **76**, 1377–1383.
Zurek, P. M. (**1980**). ''The precedence effect and its possible role in the avoidance of interaural ambiguities,'' J. Acoust. Soc. Am. **67**, 952–964.
Zurek, P. M. (**1987**). ''The precedence effect,'' in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer-Verlag, New York).

Loudness summation for pulsatile electrical stimulation of the cochlea: Effects of rate, electrode separation, level, and mode of stimulation

Colette M. McKay, Maria D. Remine, and Hugh J. McDermott
Department of Otolaryngology, The University of Melbourne, 384 Albert Street, East Melbourne 3002, Australia

(Received 31 October 2000; revised 19 March 2001; accepted 20 June 2001)

The aim of these two experiments was to gain systematic data on the amount of loudness summation measured for dual-electrode stimuli with varying temporal and spatial separation of current pulses. Loudness summation is important in the implementation of speech processing strategies for implantees. However, the loudness mapping functions used in current speech processors utilize psychophysical data (thresholds and comfortable loudness levels) derived using single-electrode stimuli, and do not take into account the temporal and spatial patterns of the speech processor output. In the first experiment, the current reduction required to equalize the loudness of a dual-electrode stimulus to that of its component (and equally loud) single-electrode stimuli was measured for three electrode separations (0.75, 2.25, and 7.5 mm), three repetition rates (250, 500, and 1000 Hz), and two loudness levels (comfortably loud, and mid-dynamic range). It was found that electrode separation had little effect on loudness summation, except for interactions with level and rate effects at the smallest separation. More current adjustment (in dB) was required for higher rates and lower levels of stimulation. The second experiment investigated the effects of mode (monopolar versus bipolar) and pulse duration on loudness summation. More current adjustment was required in bipolar mode than in monopolar mode at the lower level only. The main effects in both experiments, and their interactions, are consistent with a loudness model in which the neural excitation density is first obtained by temporal integration of excitation at each cochlear place, then converted to specific loudness via a nonlinear relationship, and finally integrated over cochlear place to obtain the loudness. The two important features which affect the loudness relationships in dual-electrode stimulation in this model are the shape of the excitation density function and the amount by which the neural spike probability per pulse is reduced in areas of overlapping excitation due to refractory effects. [DOI: 10.1121/1.1394222]

PACS numbers: 43.66.Cb, 43.66.Ts [SPB]

I. INTRODUCTION

At present, the most common form of electrical stimulation used in cochlear implants is sequential biphasic current pulses. Such stimulation avoids the possibility of direct current summation across different electrode positions. However, there are still physiological and/or perceptual interactions between the effects of current pulses that are separated in time or place. One important percept that is affected by spatial or temporal separation of current pulses is loudness. Loudness is of importance to the implementation of speech processing strategies for implantees. Currently, the loudness mapping functions used in speech processors are based on psychophysical data (thresholds and comfortable loudness levels) derived for constant-rate stimuli on single electrodes. In normal operation, these mapping processes do not take into account the varying temporal and spatial patterns of the speech processor output to dynamically modify intensity levels. The aim of the experiments described here is to gain systematic data on the perceptual effects of temporal and spatial separation of current pulses at different overall levels, with the ultimate goal of predicting the loudness of complex multiple-electrode stimuli.

Early work (Shannon, 1983) showed that loudness summation data could be used to study the direct current summation which occurred in simultaneous stimulation. This direct current summation was the main reason for the adoption of sequential pulsatile stimulation as the normal stimulus waveform. The effect of electrode separation on loudness for dual-electrode stimuli using sequential pulsatile stimulation at a comfortably loud level has been investigated by Tong and Clark (1986), and by McKay and McDermott (1994) with subjects who used the Mini System 22 implant and bipolar stimulation modes. Tong and Clark found that, for two implantees and a repetition rate of 166 Hz, loudness summation increased with increasing electrode separation up to 3 mm. In contrast, McKay and McDermott found that, for four implantees using 250-Hz repetition rate, there was no relationship between loudness summation and electrode separation for separations greater than 0.75 mm.

Loudness summation of acoustic signals has generally been interpreted in terms of auditory filter models (Zwicker and Scharf, 1965; Moore and Glasberg, 1997). In these models, the signal is first passed through a bank of overlapping bandpass filters, and the outputs of the filters are used to determine the spatial pattern of ''excitation'' along the cochlea. (Note that the term ''excitation'' in these acoustic models refers to the output of the auditory filters, and not

neural excitation which is referred to later in this article.) This excitation pattern is converted to a specific loudness pattern, representing the loudness contribution at each point along the cochlea, via a compressive power function, which models the transformation from physical excitation to neural excitation and to loudness. Finally, the specific loudness is integrated across cochlear place to determine the overall loudness of the sound. These models correctly predict that the loudness of a sound of constant total power with multiple components will increase with increasing frequency separation of the components, up to the point where the excitation patterns of the individual components no longer overlap. This is because close signal components will directly combine within an auditory filter before the excitation pattern is determined and the compressive power function applied, whereas components which do not excite the same filter contribute independently to loudness (that is, after the compressive power function is applied). This acoustic model, however, cannot be applied to the electrical case of sequential pulsatile stimulation on two electrodes because (a) pulses on two electrodes are not simultaneous, (b) the two stimulus currents are therefore not directly combined before processing by the auditory system, even if the current distributions spatially overlap, and (c) the acoustic compressive power function will not be applicable because it arises from the nonlinear behavior of basilar membrane mechanics, contributed to by the active process of outer hair cells.

McKay and McDermott (1998) studied the effects on loudness of the temporal parameters of sequential current pulses on the same electrode. In that article, a loudness model was developed which accounted for the effects of temporal separation of the pulses. The two main features of the model were a function which described the relationship between temporal separation of pulses and peripheral neural refractory effects, and a sliding temporal integration window of equivalent rectangular duration (ERD) 7 ms. The model assumed that loudness was related to the total neural excitation within the temporal integration window (for a steady signal of duration greater than 300 ms).

The objective of this article is to extend the temporal model in McKay and McDermott (1998) to account also for spatial separation of individual pulses; that is, where the sequential current pulses are delivered to two different electrode positions. In the experiments reported in this article, the parameter of electrode (spatial) separation was investigated. The interactions of temporal separation (rate of stimulation), mode of stimulation, and overall level of stimulation with loudness summation were also studied.

II. PROCEDURES

A. Subjects and stimuli

A total of 12 adults participated in the experiments. Five were users of the CI22M implant (S1–S5) and seven were users of the C124M implant (S6–S12), both manufactured by Cochlear Ltd. Details of their etiology and implant experience are contained in Table I.

TABLE I. Details of subjects' etiology and implant use.

Subject	Implant type	Length of profound deafness	Implant date	Etiology
S1	CI22	17 years	1988	Trauma
S2	CI22	8 years	1992	Otosclerosis
S3	CI22	20 years	1992	Otosclerosis
S4	CI22	2 years	1996	Progressive/Genetic
S5	CI22	16 years	1988	Otosclerosis
S6	CI24	40 years	1990	Infection
S7	CI24	15 years	1996	Otosclerosis
S8	CI24	4 years	1998	Progressive/Unknown
S9	CI24	1 year	1996	Progressive/Unknown
S10	CI24	15 years	1998	Progressive/Unknown
S11	CI24	1 year	1978/94 (reimplanted)	Trauma
S12	CI24	36 years	1999	Progressive/Genetic

All stimuli used in the experiments were biphasic current-pulse trains of duration 500 ms. The pulse phase duration was fixed for each subject within an experiment, and loudness was varied using the current parameter. Current is varied in these implants in approximately logarithmic current steps. The current steps used in the experiments were converted to current values in μA, using either the individual calibration charts (for CI22M users) or a calibration formula (for C124M users). The interphase gap (between the two phases of each biphasic pulse) was fixed at 43 μs for all stimuli.

Single- and dual-electrode stimuli were used in the experiments. Single-electrode stimuli consisted of constant-current pulse trains of rates 250, 500, or 1000 Hz. Dual-electrode stimuli had two biphasic pulses (one on each of two electrodes) in each repetition period, with repetition rates of 250, 500, or 1000 Hz. The onsets of the first and second biphasic pulses were separated by either 20% or 40% of the repetition period. The particular parameters of the stimuli used for each subject are listed in Table II.

B. The loudness-balance procedure

The loudness-balance procedure was a method of adjustment. To balance two stimuli (A and B), the current of A was first fixed at the required level, and A and B were presented alternating continuously separated by 500-ms silent intervals. The subject was instructed to adjust the current of B with an unmarked rotary knob, or ''up'' and ''down'' buttons, until they considered B to be the same loudness as A. The procedure was then repeated with the current of B fixed at the previously found level, and the subject adjusting the current of A until it had the same loudness as B. The two trials were then repeated (for a total of four trials). When two single-electrode stimuli were being balanced, the average result of the four trials was used to set the currents (to the nearest current step) of these stimuli for subsequent use in the loudness-summation task. The same data were used to set the relative difference in current-step values on the two electrodes when they were used in the corresponding dual-electrode stimuli.

All measurements of loudness summation were obtained by balancing the loudness of a dual-electrode stimulus to that of each of its constituent single-electrode stimuli. When the

TABLE II. Experimental stimulation parameters. CG=common ground mode; BP+1=bipolar mode with ring separation of 1.5 mm; MP=monopolar mode. Electrode numbers refer to the single active intracochlear electrode ring in CG or MP mode, or the most apical of the pair of active rings in bipolar mode. Rings are numbered 1–22 in basal-to-apical direction.

Subject	Experiment	Mode	Pulse duration (μs)	Electrodes	Rates (Hz)	Dynamic ranges of electrodes (dB)	
S1	1	CG	150	20, 19, 17, 10	250/500	250 Hz: 3.2, 3.1, 3.6, 2.6	500 Hz: 4.2, 3.7, 5.4, 4.1
S2	1	BP+1	150	18, 17, 15, 8	250/500	250 Hz: 4.3, 5.1, 2.6, 4.1	500 Hz: 6.1, 4.3, 5.0, 5.1
S3	1	BP+1	150	13, 12, 10, 3	250/500	250 Hz: 4.7, 4.0, 5.3, 2.5	500 Hz: 6.0, 5.7, 7.2, 5.3
S4	1	BP+1	150	18, 17, 15, 8	250/500	250 Hz: 2.8, 2.9, 2.6, 3.0	500 Hz: 3.5, 2.3, 3.4, 3.8
S5	1	CG	150	16, 15, 13, 6	250/500	250 Hz: 4.1, 4.2, 4.3, 5.0	500 Hz: 6.0, 4.9, 4.8, 5.0
S6	1	MP	25	22, 21, 19, 13	500/1000	500 Hz: 4.2, 4.4, 3.9, 3.7	1000 Hz: 6.9, 6.5, 5.3, 6.3
S7	1	MP	37	17, 16, 14, 7	500/1000	500 Hz: 5.3, 5.1, 3.9, 4.7	1000 Hz: 8.6, 6.5, 7.6, 6.0
S8	1	MP	25	20, 19, 17, 10	500/1000	500 Hz: 4.2, 4.6, 3.7, 8.4	1000 Hz: 4.8, 5.5, 6.2, 9.3
	2	MP	25/150	20, 17	500	MP/150: 4.7, 4.7	
	2	BP+1	150	19, 16	500	BP+1/150: 3.0, 4.6	
S9	1	MP	25	19, 18, 16, 9	500/1000	500 Hz: 3.9, 3.2, 3.2, 3.3	1000 Hz: 6.0, 6.7, 6.2, 5.8
S10	2	MP	25/150	19, 16	500	MP/25: 7.7, 8.4 150: 11.4, 10.0	
	2	BP+1	150	18, 15	500	BP+1/150: 8.6, 9.0	
S11	2	MP	25/150	19, 16	500	MP/25: 7.4, 7.4 150: 9.1, 9.3	
	2	BP+1	150	18, 15	500	BP+1/150: 8.6, 7.9	
S12	2	MP	25/150	19, 16	500	MP/25: 5.3, 3.9 150: 6.7, 6.5	
	2	BP+1	150	19, 16	500	BP+1/150: 6.5, 6.7	

loudness of the dual-electrode stimulus was being adjusted by the subject, the two currents were always adjusted in equal current steps, thus maintaining a fixed current-step difference. The results of the eight loudness-balance trials with each dual-electrode stimulus (four trials with each of the two single-electrode stimuli) were retained for statistical analysis. The measure of loudness summation analyzed was the reduction of current required in the dual-electrode stimulus to make it equal in loudness to the single-electrode stimuli. This current reduction was expressed in dB, and also as a percentage of the dynamic range of the particular electrodes used.

C. Experimental procedures

1. Experiment 1: Effect of separation, rate, and level

This experiment investigated the effect of electrode separation, rate, and current level on loudness summation. Nine subjects participated in this experiment, five users of the CI22M implant (S1–S5) and four users of CI24M implant (S6–S9).

Three electrode separations were investigated: 0.75, 2.25, and 7.5 mm (one, three and ten electrode ring spacings, respectively). Note that subject S6 had a separation of 6.75 mm (nine rings) for the largest separation, because the electrode that would have been used for the 7.5-mm separation had to be deactivated during the experimental period. For each subject, each pair of electrodes in the dual-electrode stimuli had the same apical member. The four electrodes selected, their dynamic ranges, and the stimulus parameter values for each of these subjects are listed in Table II.

Two different overall levels were investigated. The first was a level considered to be comfortably loud. The current on the most-apical of the four electrodes was adjusted to several current steps below the maximum tolerable loudness (allowing for current adjustment above and below this level), and the three other single-electrode stimuli were loudness-balanced to this stimulus. The second overall level used was at approximately 50% of the dynamic range. The current on the most-apical electrode was adjusted to be half way (in current steps) between the first level used and the threshold level, and the other three single-electrode stimuli were again loudness-balanced to this stimulus.

Two different rates of stimulation were investigated for each subject. For the CI22M users, repetition rates of 250 and 500 Hz were used for both the single- and dual-electrode stimuli. In the dual-electrode stimuli, the second of the two pulses occurred with a delay relative to the first of 20% of the repetition period for 250 Hz, and 40% for 500 Hz, making a fixed interpulse interval (between the onsets of the first and second pulse) of 0.8 ms. With the CI24M implant, higher rates and smaller interpulse intervals are possible. For users of this implant, repetition rates of 500 and 1000 Hz were investigated. The interval between the first and second pulse in the dual-electrode stimuli was 20% and 40% of the repetition period, respectively, making a fixed interpulse interval of 0.4 ms. The rates and interpulse intervals were chosen to represent typical values present in the output of the SPEAK (Skinner *et al.*, 1994) or ACE (Vandali *et al.*, 2000) speech-processing strategies.

Loudness summation was measured for each of the subjects in all 12 conditions (three electrode separations, two levels, two rates). As there were some differences in the stimuli used with subject groups using the two different implants, the results were analyzed separately for each subject group to test for the effects of electrode separation, rate, and level.

2. Experiment 2: Effect of mode and pulse duration

The aim of this experiment was to see if any overall difference in the amount of loudness-summation for the two groups of subjects in experiment 1 would be explained by differences in the stimuli used, or alternatively was due to subject-dependent factors. Four CI24M users participated in this experiment. As three of the CI24M users in experiment 1 (S6, S7, S9) were no longer available, three new subjects

were recruited (S10–S12): see Tables I and II. There were three differences in the stimuli for the two groups in experiment I: the mode of stimulation for CI24M users was monopolar (using both extracochlear indifferent electrodes), whereas it was common ground or bipolar+1 for the CI22M users; the pulse duration used for CI24M users was 25–45 μs, while for the CI22M users it was 150 μs; and the interval between the first and second pulses in the dual-electrode stimuli was 0.8 ms for the CI22M users, and 0.4 ms for the CI24M users. Using the CI24M implant, it is possible to compare the effect of these stimulus parameter changes in the same subjects. Experiment 2 compared the loudness summation for two of the conditions used in experiment 1 (electrode separation of 2.25 mm and 500-Hz rate, at both levels), using the two sets of stimulus parameters used for each of the subject groups in that experiment. A third set of stimulus parameters, intermediate to those used for the two groups in experiment 1, was also selected: monopolar mode was used with the longer pulse duration (150 μs) and the longer interpulse interval (0.8 ms). When selecting the bipolar electrode to compare with each monopolar electrode, the two rings immediately adjacent to the intracochlear monopolar electrode were selected for the corresponding BP+1 electrode. The only exception was S12, who experienced some facial nerve stimulation with the initially selected bipolar electrode, and so an adjacent bipolar electrode was selected (see Table II). As in experiment 1, the loudness summation was measured as the current reduction required for the dual-electrode stimulus to equalize it in loudness to the two corresponding single-electrode stimuli. To ensure all stimulus-parameter conditions were tested at the same overall loudness, the most-apical of the two single-electrode stimuli used in each of the three stimulus conditions (and which were used as the references for subsequent loudness-balancing tasks) were first loudness-balanced to each other.

The difference in interpulse interval (0.8 versus 0.4 ms) was not expected to have a significant effect on loudness. This expectation was based on previous research investigating the effect on loudness of interpulse intervals (McKay and McDermott, 1998). In that experiment, changing the interval between the two pulses by 0.4 ms (at a repetition rate of 250 Hz) resulted in a current adjustment for equal loudness of less than 0.05 dB for all subjects. The nonsignificant effect of interpulse interval on loudness summation in the present experiment at 500 Hz was confirmed in two subjects (S7 and S8) by comparing loudness summation in monopolar mode with a pulse duration of 25 μs for the two interpulse intervals of 0.4 and 0.8 ms. The mean difference between amount of loudness summation measured for the two intervals was not significantly different from zero in both cases.

The difference in pulse durations could affect the amount of loudness summation. Previous experiments investigating the perceptual effect of pulse duration (McKay and McDermott, 1999) indicated that, for stimuli with the same loudness, the neural excitation patterns may be spatially narrower for shorter pulse durations than for longer pulse durations. If this is correct, and loudness summation for dual-electrode stimuli depends on the extent of spatial overlap of the two excited neural populations, then the use of different pulse durations may lead to different amounts of loudness summation for the same electrode separation and overall loudness level.

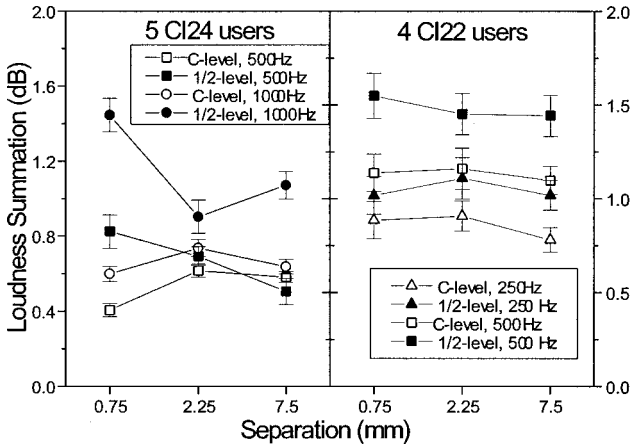


FIG. 1. The reduction in current required to equalize in loudness a dual-electrode stimulus relative to its component single-electrode stimuli at three electrode separations. The left panel shows the results (means and standard errors) for CI24M users (in monopolar mode), and the right panel the results for CI22M users in bipolar mode (see Table II). The symbol shapes represent different rates of stimulation, and the open and closed symbols represent comfortably loud and half-dynamic range levels, respectively.

In a similar way, the mode of stimulation (monopolar versus bipolar) may affect the shape of the neural excitation patterns and thus potentially affect loudness summation across different electrodes. Bipolar stimulation produces a steeper spatial gradient of electrical potential than does monopolar stimulation and hence narrower spatial tuning curves for auditory neurons (Kral *et al.*, 1998). Pfingst *et al.* (1995) proposed that the effects of mode on thresholds at various pulse rates were consistent with a greater number of neurons being activated at threshold, but with lower spike probability, in monopolar mode than in bipolar mode.

III. RESULTS

A. Experiment 1

The average amounts of loudness summation (adjustment in current to equate loudness for dual-electrode stimuli and single-electrode stimuli) for the 12 conditions of electrode separation, rate, and overall level are shown in Figs. 1 and 2, with results for the two subject groups shown in separate panels. In Fig. 1, the loudness summation is expressed as current adjustment in dB. The same data are shown in Fig. 2, with the current adjustment expressed as a percentage of the dynamic range (%DR). For example, if the current adjustment was 1 dB and the dynamic range was 10 dB, the %DR adjustment would be 10%. The data were analyzed with analysis of variance, using a general linear model with separation, rate, and level as fixed factors, and subject as a random factor.

For the CI22 subjects and data in dB, there was no significant main effect of electrode separation ($p=0.17$), but there were significant main effects of level ($p<0.001$) and rate ($p<0.001$), with more loudness summation at the lower level and at the higher rate. There were no significant interactions between the three main factors. When the same data

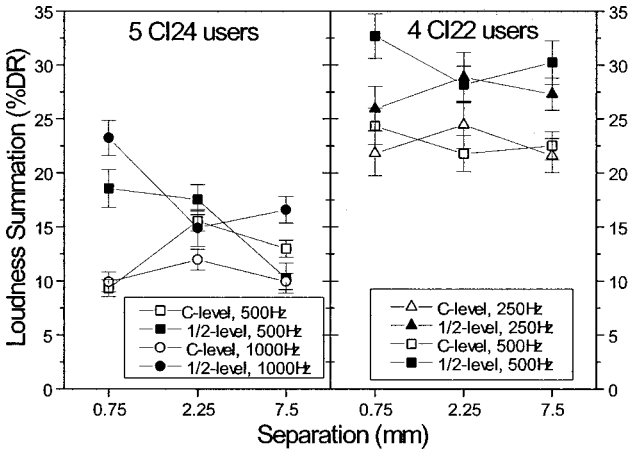


FIG. 2. The same loudness summation data as in Fig. 1, with the current adjustment represented as a percentage of the dynamic range (in dB) of the corresponding single-electrode stimuli.

were expressed as %DR, only level had a significant main effect ($p<0.001$), while the effects of separation and rate were not significant ($p=0.72$ and $p=0.09$, respectively). There was, however, a moderately significant interaction ($p=0.02$) of the separation and rate factors. Inspection of Fig. 2 (right panel) shows that this interaction is due to increased loudness summation for the higher rate at the smallest separation (0.75 mm), compared to the other separations. This was confirmed by analyzing the data separately for the three separations. At all three separations there were significant effects of level ($p<0.001$ in all cases), but at the separation of 0.75 mm, there was also a significant effect of rate ($p=0.003$). In summary, the data from CI22 users, when expressed in dB, show that there was no effect of electrode separation, but significantly more loudness summation at lower levels and higher rates. Expressing the loudness changes in %DR largely negated the rate effect because of the larger dynamic ranges at the higher rate (see Table II). However, even when the data were expressed in %DR, there remained significantly more loudness summation for the higher rate at the smallest separation.

The data from the CI24 users present a more complicated picture. Analysis of variance of the data expressed in dB, with separation, level, and rate as fixed factors and subject as a random factor, showed all factors to have a significant effect. There were, however, highly significant interactions between all three factors, so the data were divided into subgroups to investigate these interactions further. First, the effect of separation was investigated for the four level/rate conditions (Fig. 1, left panel) using ANOVA followed by Tukey's paired comparisons. In all four conditions, there was a significant effect of separation, but the effect was in different directions at the two levels. At C-level, summation at the smallest separation was less than at one or both of the wider separations: for 500 Hz (ANOVA $p<0.001$), summation at 0.75 mm was less than at both 2.25 and 7.5 mm ($p<0.001$ in both cases), and for 1000 Hz (ANOVA $p=0.008$), summation at 0.75 mm was less than at 2.25 mm ($p=0.007$). At the lower level, however, the summation at 0.75 mm was *greater* than at one or both of the wider separations: for 500 Hz (ANOVA $p<0.001$), summation at 0.75 and 2.25 mm was greater than that at 7.5 mm, and for 1000 Hz (ANOVA $p<0.001$), summation at 0.75 mm was greater than at both 2.25 ($p<0.001$) and 7.5 mm ($p<0.001$). In only one condition (500 Hz, lower level) was there a significant difference in summation for 2.25- and 7.5-mm separations, with less summation at 7.5 mm ($p=0.03$).

Next, the data (in dB) were divided into the three separations (0.75, 2.25, and 7.5 mm) to investigate the effects of level and rate. At all three separations, there was a significant effect of level ($p<0.001$, $p=0.02$, and $p<0.001$, respectively) and rate ($p<0.001$, $p=0.003$, $p<0.001$, respectively), with more summation at lower levels and at higher rates. For two separations (0.75 and 7.5 mm) there was also a significant interaction between level and rate ($p=0.001$ and $p<0.001$, respectively), with the effect of level being greater at the higher rate. When the whole set of data were expressed in %DR and reanalyzed, the main effect of rate became nonsignificant, as was the case for CI22 users. However, there still remained significant interaction effects between all three parameters. Analysis of the effect of separation for the four separate level/rate conditions (Fig. 2, left panel), produced the same pattern of separation effect as for the data in dB, except that separation did not reach significance for one condition (C-level and 1000 Hz). Analyzing the effect of rate and level for the three individual separations, there was a significant effect of rate and level at the smallest separation (0.75 mm) ($p=0.04$ and $p<0.001$, respectively), with more summation at higher rates and lower levels (as found for the CI22 data). At 2.25 mm, there was a significant effect of rate ($p=0.03$) and level ($p=0.02$), with more summation at the lower level and *lower* rate. At the largest separation (7.5 mm), neither rate ($p=0.06$) nor level ($p=0.2$) were significant main effects, but there was a highly significant interaction ($p<0.001$), with the effect of level being reversed at the two rates.

To summarize this rather complicated picture, the most consistent effect in all the data (from both subject sets) was that more current adjustment was needed at lower levels to maintain equal loudness. Also consistent across the data was the overall effect of rate: more current adjustment was needed for higher rates. The effect of separation on loudness summation was surprisingly small compared to the effects of level and rate. There was no consistent evidence of increased loudness summation with separation (the only data to show this trend were the C-level data for CI24 users, in which the effect was very small). The interesting feature of the separation parameter is its interaction with level and rate effects: it appears that, at the smallest separation in monopolar mode, the effects of level and rate are greater than at larger separations. Thus, at the lower level, loudness summation was greater at 0.75 mm than at wider separations, but at the higher level, summation was less at 0.75 mm than at wider separations. In both subject groups, the largest effect of level was seen for the higher rate at 0.75 mm.

B. Experiment 2

The results of experiment 2 are shown in Fig. 3, where the two panels show the loudness summation for different modes and pulse durations, with current adjustment ex-

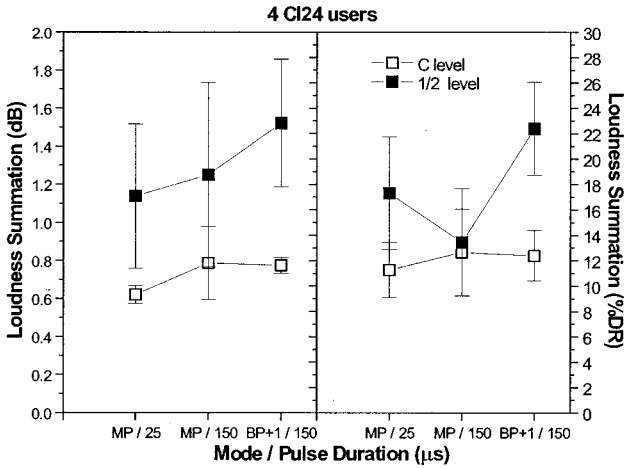


FIG. 3. The reduction in current required to equalize in loudness a dual-electrode stimulus relative to its component single-electrode stimuli for three conditions differing in mode of stimulation and/or pulse duration. All subjects were CI24M users, the electrode separation was fixed at 2.25 mm, and the rate was 500 Hz. The open and closed symbols represent means (and standard errors) across subjects for the comfortably loud and half-dynamic range levels, respectively.

pressed in dB and %DR, respectively. The data for the three conditions at each overall level were analyzed using a general linear model with mode/pulse duration as a fixed factor and subject as a random factor. There was no significant effect of the mode/pulse duration condition on loudness summation at the comfortably loud level ($p=0.25$ and $p=0.5$ for dB and %DR, respectively). However, there was a significant effect of mode/pulse duration condition at the lower level ($p=0.009$ and $p<0.001$ for dB and %DR, respectively). Tukey's pairwise comparisons ($p=0.05$ family significance level) showed that, for both dB and %DR data, the bipolar condition showed more loudness summation than both of the monopolar conditions, which were not significantly different from each other.

IV. DISCUSSION

A. The loudness model

The loudness summation data in these experiments will be discussed with reference to a model of electrical loudness perception which has three main steps: temporal integration of neural excitation; transformation of excitation to specific loudness; and spatial integration of specific loudness to obtain the overall loudness. The three model steps will be discussed in this section, and the application of the model to the interpretation of the present data will be discussed in the later sections.

In the first step of the loudness model, it will be assumed that, at each cochlear place, the neural spikes arising from each successive pulse in the stimulus are summed by the temporal integration window to obtain a spatial distribution of neural excitation (E) versus cochlear place (x), which will be referred to as the excitation density function, $E(x)$. When successive current pulses stimulate the same cochlear place, loudness is determined by the total neural excitation within a time window with ERD of about 7 ms (McKay and McDermott, 1998). It should be stressed that the ERD of this window is associated with temporal resolution tasks (e.g., Viemeister, 1979; Moore *et al.*, 1996), and is not the relatively long integration time (about 300 ms) associated with duration effects on loudness (e.g., Zwislocki, 1969). The total excitation within the window will be determined by both the number of stimulus current pulses within the window and the average neural spike probability for each of these pulses. The spike probability is the probability that a particular neuron will fire when activated by a current pulse (for example, if a neuron fires on every current pulse in a pulse train, its spike probability is 1). In the overlapping part of the excited neural population in dual-electrode stimulation, the effective rate of stimulation is doubled compared to that for neurons driven exclusively by one of the active electrodes. This leads to two counteracting effects on loudness in that neural population (for repetition rates greater than about 100 Hz). First, because the average interpulse interval is halved (relative to the single-electrode case), the average neural spike probability for each stimulus pulse decreases due to refractory factors. Second, the number of stimulus pulses falling within the integration window is doubled. McKay and McDermott (1998) found that when the rate was doubled on a single electrode, the total excitation produced within the time window was increased, because the increase due to doubling the number of stimulus pulses within the window exceeded the decrease due to the fall in average spike probability per pulse.

The second step of the model is to transform the spatial neural excitation pattern to a specific loudness pattern. It will be assumed that the specific loudness arising from each cochlear place, $\mathrm{SL}(x)$, is related to the excitation density at that place by some function, f:

$$\mathrm{SL}(x)=f\cdot E(x). \tag{1}$$

The form of the function f is not known, although its possible features are discussed below. It will be assumed to be an increasing monotonic function, so that equal excitation density will be assumed to provide equal specific loudness, and increasing excitation density will be assumed to provide increasing specific loudness.

In the third model step, the overall loudness is obtained by integrating $\mathrm{SL}(x)$ over cochlear place. This step is identical to the corresponding final step in the acoustic loudness model of Moore and Glasberg (1997). That is, the overall loudness of a signal is obtained by integrating over cochlear place the specific loudness arising at each place. Since the specific loudness for acoustic stimulation is presumed to represent a state more central than the haircell-to-neural-excitation transduction, the process of integrating specific loudness over place to obtain overall loudness is likely to be the same for acoustic and electrical stimulation of the cochlea. It is important to note here that, if f were a linear function, then overall loudness would be dependent only on the total excitation in the integration window [i.e., the integral of $E(x)$ across cochlear place]. If f were not linear, however, then the total loudness would depend on the *shape* of the distribution $E(x)$. That is, the same total amount of excitation would be perceived as different in loudness if it were distributed across cochlear place in a different way. For example, if f were a power function with exponent greater

than 1, then a fixed total amount of excitation would evoke a louder sensation if distributed over a narrower range of cochlear distance.

B. The effect of level on loudness summation

Before discussing the effect of level on loudness summation, it is interesting to examine the implications of the results of these experiments and the loudness model for the relationship between current and loudness. When the current of a stimulus is altered, this may alter both the shape and the magnitude of the neural excitation density function, E. The loudness $L(i)$ evoked by current i is given in the loudness model by

$$L(i) = \int f \cdot E(x,i)\, dx. \tag{2}$$

This general equation represents loudness as a complex function of current if the shape of the excitation density function varies with current and f is a nonlinear function. To separate the effects on loudness of the shape and magnitude of the excitation function, we can write

$$E(x,i) = g(i) E_0(x,i), \tag{3}$$

where $g(i)$ is the total excitation (within the temporal window) for current i, and $E_0(x,i)$ is a normalized shape function whose integral over x is unity.

A simplification of Eq. (2) is possible if the excitation to specific loudness function, f, were a power function. For example,

$$f(y) = ky^n, \tag{4}$$

where k and n are constants.

In this case, the effect of magnitude and shape of the excitation pattern can be separated:

$$L(i) = kg^n(i) \int E_0^n(x,i)\, dx, \tag{5}$$

or

$$\log(L(i)) = n \log(g(i)) + \log\left(\int E_0^n(x,i)\, dx \right) + k', \tag{6}$$

where k' is a constant.

If only the current is varied, and the stimulus location and other stimulus parameters held constant, Eq. (6) describes how loudness grows with current. If the shape of the excitation function is invariant with current level, only the first term is dependent on current. In practice, the approximation of the second term as a constant is reasonable only over small current ranges, and when all other stimulus parameters (including electrode location and subject) are held constant. For instance, it is likely to be a fair approximation for describing how the loudness changed with current adjustment in the loudness-balancing procedure in these experiments. In Eqs. (5) and (6), loudness then becomes a power function of the current-to-total-excitation transform $g(i)$. That is, provided the function f is a power function, and provided that the excitation shape is not current dependent, equal ratio changes of total excitation with level will produce equal ratio changes of loudness, independent of the absolute current level.

To further characterize the way loudness grows with current it is necessary to measure the shape of the current-to-total-excitation function, $g(i)$. It should be noted that, unlike the function $f(E)$, the form of the function $g(i)$ is theoretically highly dependent on stimulus parameters other than current, as well as individual cochlear and electrode parameters. For example, the spatial distribution of surviving stimulable neural elements will contribute to differences among individual implantees and electrode positions in the way that excitation grows with current, as well as possibly contributing to changes in the shape of the excitation pattern with current level. Although the data from the present experiment are insufficient to allow the function $L(i)$ to be determined, a comparison of the loudness summation measurements at the two current levels for each subject allows the relative slopes of this function at the two levels to be inferred.

Our model and data suggest that the loudness of the dual-electrode stimuli in this experiment was always about double that of the equally loud constituent single-electrode stimuli (before current adjustment). This proposition flows from the assumption that the two electrodes would contribute independently to the overall loudness if their corresponding activated neural populations were nonoverlapping, and the observation that the loudness of the dual-electrode stimulus was independent of electrode separation. If so, the current adjustments to equalize the loudness at the two overall levels represented the same loudness change at each level (i.e., $\frac{1}{2}$ loudness). The fact that, on average, more current adjustment in dB was required at lower levels, implies that $L(i)$ cannot be a power function across the whole dynamic range of current. If it were a power function, then equal loudness changes should be produced by equal ratio adjustments in current, and therefore there should not have been a significant level effect when the adjustments were measured in dB.

Recent reports have suggested that loudness may be modeled as an exponential function of current (Zeng and Shannon, 1992; Chattejree *et al.*, 2000). If loudness were an exponential function of current, equal loudness adjustments would result from equal linear current adjustments, and so it would be predicted that, in experiment 1, the dB or ratio adjustment in current would be larger for each subject when overall current level was lower. Given the ratio current adjustment measured at the higher level, and the absolute current at the two overall levels used in experiment 1, it is possible to calculate the ratio current adjustment at the lower level which would be predicted by an exponential function of loudness. The average data for each of the nine subjects at the two overall levels were examined individually to test whether the current adjustments predicted for the lower level by an exponential or power function were equal to the measured values. The results of this analysis are shown in Fig. 4, where, for each subject, the measured value of current adjustment is plotted against the predictions based on either a power or an exponential loudness function. Data which lie on the diagonal line have predictions which match the mea-

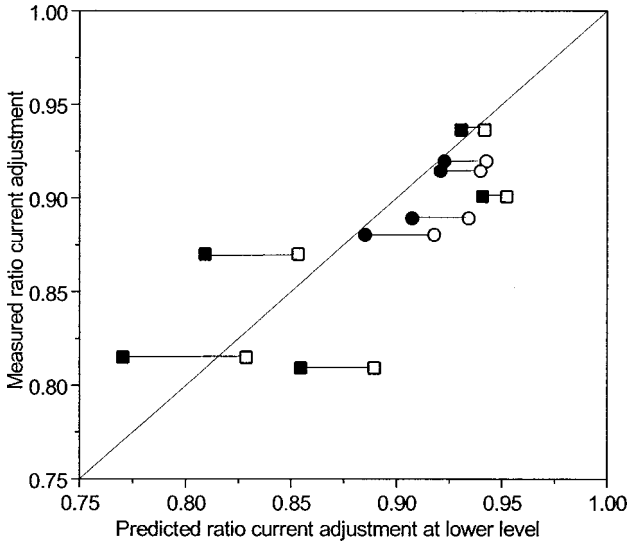


FIG. 4. The measured ratio current adjustment at the lower level (vertical axis) versus the predicted ratio current adjustments (horizontal axis) for simple power-function (open symbols) and exponential-function (closed symbols) models of loudness growth with current. The two predictions for each of the nine subjects in experiment 1 are joined by short horizontal lines. The CI22 subjects' data are represented with squares, and the CI24 subjects' data with circles. The predictions are based on the ratio adjustment at the higher level for each subject and (for the exponential case) the relative absolute currents at each level. The diagonal line represents the cases where predicted and measured ratios are equal.

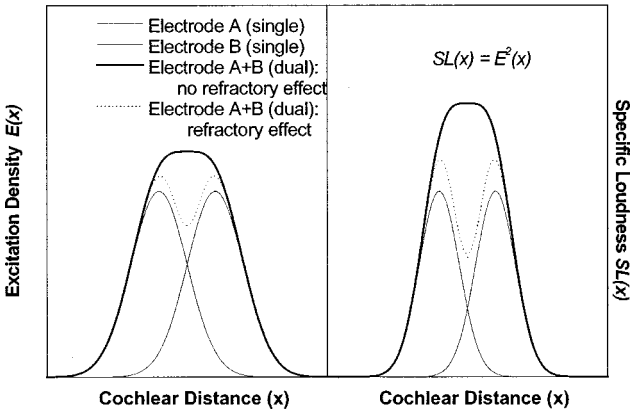


FIG. 5. The left panel shows a schematic representation of the excitation density (total neural excitation within a temporal integration window as a function of cochlear place) for two single-electrode stimuli (thin solid lines) and for the corresponding dual-electrode stimulus (before loudness equalization). A comparison of the thick solid and dotted lines shows the hypothetical result of a reduced spike probability per pulse in the region of overlapping excitation in the dual-electrode stimulus. The right panel shows the specific loudness curves for the excitation patterns in the left panel for the hypothetical transform example shown in the equation.

sured current adjustment. Two general observations can be made about these data: first, the exponential function (closed symbols) predicted more ratio current adjustment at the lower level than did the power function (open symbols), as expected; second, the accuracy of each prediction varied substantially among the subjects. For the five CI22 subjects (square symbols), four predictions for the exponential function and three for the power function were substantially different (by the equivalent of three or more current steps) from the measured values. In two of the CI22 subjects, the exponential function prediction was substantially worse than the power function prediction. For all four CI24 subjects (circle symbols), however, the exponential-function predictions were better than the power-function predictions. In general, our data suggest that, although an exponential function of loudness may allow better predictions of loudness growth with current than a power function in some cases, neither simple function may adequately predict the variations in loudness growth among different subjects or stimulus parameters. This assertion is supported by Eq. (6), if it is acknowledged that both $g(i)$ and $E_0(x,i)$ are likely to be subject and stimulus dependent, and that $E_0(x,i)$ is not likely to be independent of i over the whole dynamic range of the stimulus.

C. The effect of electrode separation

The results of the experiments show that, unlike the corresponding situation in acoustic hearing, the separation of two electrodes has minimal effect on loudness summation compared to other factors such as rate and level of stimulation. This supports the introductory discussion about the differences between acoustic and electric loudness summation for components distributed over cochlear place. However, it is likely that, in both the acoustic and electric cases, the effects on loudness of frequency or electrode separation depend upon the extent of spatial overlap of mechanical or neural excitation at the cochlea or acoustic nerve level. With currently used electrode arrays and stimulus parameters, most stimuli delivered by different electrodes probably produce excitation patterns which overlap to some extent. This is supported by forward-masking patterns obtained using different masker electrodes (for example, see Cohen *et al.*, 1996). It is reasonable to assume that the separation of two electrodes would directly affect the proportion of the total activated neural population which is excited by both electrodes. The process whereby the neural excitation evoked by the two electrodes is combined to contribute to the specific loudness at a particular cochlear place will determine how electrode separation affects loudness summation.

Figure 5 is a schematic diagram to illustrate the loudness model for the case of loudness summation across two separate electrodes, A and B. The left panel shows the individual excitation patterns (after temporal integration) arising from pulses on the two electrodes if they were presented in single-electrode pulse trains (thin solid lines), and the overall excitation pattern when both electrodes are activated in the dual-electrode stimulus. The thick solid line shows the excitation density function that would be obtained if there was no reduction in average spike probability in the overlapping area when both electrodes are activated, and the dotted line shows the hypothetical result of including a refractory effect. The area under this latter curve represents the total excitation [i.e., $g(i)$] within the temporal integration window for the dual-electrode stimulus. Note that the area under the thick solid line equals the sum of the areas under the two individual excitation patterns (thin solid lines). Thus, if specific loudness was linearly related to excitation density (and therefore overall loudness was related only to the total excitation), one would predict that the refractory effect would be the only

factor affecting loudness with varying electrode separation, and that loudness would *always decrease* with increasing overlap of the individual excitation patterns. However, our data do not show this pattern of loudness change with electrode separation, and thus a model where there is a nonlinear transform from neural excitation density to specific loudness would fit the data better.

Such a nonlinear function, *f*, is used in the example of Eq. (6), which illustrates two opposing effects on loudness of varying the electrode separation in dual-electrode stimuli while keeping current constant. As the separation is decreased, the first term decreases because of a reduction in total excitation due to refractory effects, and the second term increases because the excitation pattern shape is less flat and there is a nonlinear transform between excitation density and specific loudness. To illustrate the latter point, imagine two normalized rectangular excitation pattern shapes, one of width 10 mm and constant height of 0.1 spikes/mm, and the other of width 1 mm and height 1 spike/mm. The integral over x is unity for each of these normalized shapes. However, for the example of $n=2$, the integral in Eq. (6) gives values of 0.1 and 1, respectively, for these two shapes. To generalize this illustration, if $n>1$, then the integral in the second term increases as the excitation shape becomes more narrow or peaked.

The effect of this particular transform is illustrated in the right panel of Fig. 5, where the four excitation patterns in the left panel have been transformed to specific loudness via a power function with exponent 2. In this panel, the overall loudness is given by the area under the curves. The area under the thick solid line is now greater than the sum of the areas under the individual-electrode patterns, illustrating that loudness will *increase* with increasing overlap of the excitation patterns when refractory effects are excluded. The area under the dotted line, which represents the specific loudness taking into account refractory effects, approximately equals the sum of the individual-electrode areas, illustrating that refractory effects and a nonlinear transformation from excitation to loudness can work counteractively to keep loudness relatively constant as separation is changed.

Looking at the effect of electrode separation on loudness summation in Figs. 1 and 2, it is clear that, although there is generally very little effect of separation in bipolar modes, two opposite effects of separation are apparent at the smallest separation in monopolar mode. A suitable nonlinear transform between excitation density and specific loudness would predict a loudness increase, or decrease, or no effect as separation decreases, depending on the size of the neural refractory effects.

In the schematic example shown in Fig. 5, the single-electrode excitation density function is modeled as a Gaussian shape centered on the electrode position. The size of the reduction in spike probability shown (dotted line) is approximately that predicted by our earlier work for stimulation rates of 250 Hz (McKay and McDermott, 1998). Thus the transformation function example (excitation squared) predicts effects on loudness which approximate the data obtained (i.e., minimal effect of separation), although further research is needed to determine the precise form of the function. A power function to relate neural excitation to loudness seems plausible and flows from the proposition that the central auditory system processes ratio changes of neural excitation to assess stimulus magnitude. Limits to the range of possible exponent values of such a power function are imposed by the requirement to minimize separation effects when the two excited neural populations overlap. In the extreme case of no overlap, the loudness of a dual-electrode stimulus must be double that of two equally loud individual single-electrode stimuli. At the other extreme of complete overlap of two equally loud stimuli, the total excitation in the dual-electrode stimulus must be a factor of $2^{1/n}$ greater than that in the single-electrode stimuli in order to maintain double the loudness (where n is the exponent of the power function). Our previous work (McKay and McDermott, 1998) showed that the ratio reduction in spike probability when rate was doubled was, on average, a factor of about 0.7 for rates in the range of this experiment, leading to the exponent of 2 used in the example. (The total excitation in the temporal window for the dual-electrode stimulus is increased by a factor of 1.4 over the single-electrode case, and $1.4^2=2$.) Values of 0.6 and 0.8 for the average spike probability reduction lead to exponents of 4 and 1.5 to keep loudness constant with separation.

D. The effect of rate, and its interaction with separation effects

As seen in Figs. 1 and 2, higher rates of stimulation led to greater ratio changes of current being required to equalize the loudness for dual-electrode stimuli, and this rate effect was largely negated when the current change was plotted as a proportion of the dynamic range. As rate was the only parameter investigated in experiment 1 which affected the dynamic range of the stimuli (see Table II), only the rate effect was potentially altered by expressing the data as a percentage of dynamic range rather than in dB.

In our model, there are two features which may contribute to different loudness summation effects at different rates. The first is that the magnitude of the refractory effect may vary with rate, which would lead to an interaction between rate and separation parameters. The data for both subject groups and both data units (Figs. 1 and 2) show such an interaction, with the effect of rate being greater at the smallest separation, especially at the lower levels. The direction of this interaction effect is consistent with there being a smaller drop in spike probability in the overlapping region for higher rates than for lower rates, especially at lower levels.

The second possible influence of rate on the loudness summation data is the effect of rate on loudness growth or excitation growth with current. The data showed a significant main effect of rate, independent of separation, in that more current adjustment was required for higher rates. This main effect of rate may be due to a shallower growth of loudness or excitation with current at higher rates, as suggested also by the increased dynamic ranges at higher rates (see Table II). When the data are expressed as %DR, the data are ''normalized'' for different *average* values of this slope across the whole dynamic range. However, since the slope of $\log(L)$ vs $\log(i)$ is not constant across the dynamic range, this normal-

ization would be expected to over- or undercorrect the data depending on the subject and the level in the dynamic range. This may explain some of the unexpected interactions of rate and level in the %DR data. Using the alternative normalization of %DR in linear current units will produce a similar over- or undercorrection in many cases, because $\log(L)$ versus i also does not have a constant slope for some of the data (see Fig. 4).

E. The effect of mode and pulse duration

The geometry of stimulating electrodes in the cochlea (monopolar or bipolar with various spatial extents, or common ground) has an effect on the shape of the current field and hence on the excitation pattern shape. Pulse duration may also affect the shape of the neural excitation pattern (McKay and McDermott, 1999). Equation (5) predicts that broader/flatter excitation shapes will evoke a percept less loud than narrower ones with an equal total excitation. However, given the counteracting effects of refractoriness and overall excitation pattern shape in dual-electrode stimuli when electrode separation is changed, any differences in excitation shape would be predicted to have little impact on the relative loudness of the (unadjusted) dual-electrode stimulus relative to that of the single-electrode stimuli. Any differences in ratio current adjustments to equalize loudness for different pulse durations and modes should therefore be attributable to differences in the slope of $\log(L)$ vs $\log(i)$, or $\log(g)$ vs $\log(i)$ [given Eq. (6) and our two assumptions of invariant excitation pattern shape over small current changes and a power relationship of excitation to specific loudness]. The results of experiment 2 indicated nonsignificant effects of pulse duration on required current adjustment, at least for the two durations studied. There was, however, less current adjustment needed at lower levels in monopolar mode than in bipolar mode, consistent with a greater slope of $\log(L)$ or $\log(g)$.

This difference in slope for bipolar and monopolar modes at low levels may be related to differences in how the average spike probability might change with level in the two modes. An increase in current will lead to an increase in excitation via two means: by recruiting further neurons that were previously below threshold (thus increasing the spatial spread of excitation), and by increasing the spike probability of neurons which are above threshold and not yet firing at their saturation rate. Our previous work on temporal loudness effects (McKay and McDermott, 1998) predicted the change of average spike probability with rate of stimulation at threshold and comfortably loud levels, and found, in a group of implantees using bipolar modes, that there was no significant effect of level on either the predicted average spike probability, or the way it dropped with increasing rate. This lack of an effect of level on average spike probability was interpreted to mean that the expected increases in spike probability of individual neurons with level were offset by the recruitment of additional neurons close to their thresholds. Given the similarity in absolute average spike probability at different levels, it was not surprising that increasing the rate caused similar reductions in spike probability at different levels.

It is plausible, however, that the use of monopolar stimulation might result in a different pattern of average spike probability change with level than bipolar stimulation. This is because of the different spatial distributions of the electric fields produced in bipolar and monopolar modes. In monopolar mode, the current density decreases less steeply with distance from the electrode than in bipolar mode. If this difference leads to a greater spatial spread of activated neurons, but firing with lower spike probability, then an increase in level at psychophysical threshold in monopolar mode might result in an increase in the average spike probability. Such an increase may be expected because proportionately fewer neurons would be firing at their saturation rate at threshold in monopolar mode than in bipolar mode.

Furthermore, if the average spike probability was lower in monopolar mode at lower levels than at higher levels, then there may also be less refractory effect at low levels, because the average time interval between individual spikes on a particular neuron would be longer. Our previous data support this proposition, because one subject who had a lower predicted average spike probability (0.6 compared to 0.8 for the other three subjects; see Fig. 6 of McKay and McDermott, 1998) also showed a much flatter spike probability versus rate function. A reduction in refractory effects at low levels in monopolar mode would lead, according to the loudness model, to the possibility that loudness summation could increase with decreasing electrode separation at low levels and decrease with decreasing electrode separation at high levels. Thus, the opposing separation effects seen in the present data for monopolar mode are plausibly explained by the loudness model by supposing that, in monopolar mode, average spike probability and the size of its reduction when rate is effectively doubled both increase with increasing level.

V. CONCLUSIONS

These experiments investigated the effects of electrode separation, rate, and level of electric stimulation on the loudness summation for dual-electrode stimuli. It was found that electrode separation had a minimal effect on loudness summation, whereas increasing the rate of stimulation, or lowering the level of stimulation, led to greater ratio current reductions being necessary to maintain a loudness equal to that of the component single-electrode stimuli. There were, however, interactions between electrode separation and the effects of rate and level, with greater rate and level effects for adjacent electrode positions.

The data patterns and their interactions were consistent with a model of loudness comprised of three steps.

(1) The excitation at each position along the cochlea is summed in a temporal integration window with equivalent rectangular duration of approximately 7 ms.
(2) The resultant excitation density function is transformed to specific loudness via a nonlinear relationship, such as a power function with exponent greater than 1.
(3) The overall loudness is calculated by integrating the specific loudness over position in the cochlea.

To apply this model in a more quantitative way, it will be

necessary to investigate the transform of excitation to specific loudness more precisely. In the model, the loudness of a dual-electrode stimulus relative to its component single-electrode stimuli is affected by the interactions which happen in the spatial region where neurons are exposed (above threshold) to electric fields from both electrodes. The data suggest that the effect of neural refractoriness in reducing neural excitation in this overlapping region is largely countered by a nonlinear transform of excitation to specific loudness, so that the typical effects of electrode separation on loudness summation are minimal.

ACKNOWLEDGMENTS

This research was supported by the Garnett Passe and Rodney Williams Memorial Foundation, the Australian National Health and Medical Research Council, the Cooperative Research Centre for Cochlear Implant and Hearing Aid Innovation, and the University of Melbourne Human Communications Research Centre. We wish to thank the cochlear implantees who have given their time and expertise to participate in the experiments. We also thank Brian Moore and an anonymous reviewer for valuable comments on a previous version of the manuscript.

Chatterjee, M., Fu, Q-J., and Shannon, R. V. (**2000**). ''Effects of phase duration and electrode separation on loudness growth in cochlear implant listeners,'' J. Acoust. Soc. Am. **107**, 1637–1644.

Cohen, L. T., Busby, P. A., and Clark, G. M. (**1996**). ''Cochlear implant psychophysics. 2. Comparison of forward masking and pitch estimation data,'' Audiol. Neuro-Otol. **1**, 278–292.

Kral, A., Hartmann, R., Mortazavi, D., and Klinke, R. (**1998**). ''Spatial resolution of cochlear implants: the electrical field and excitation of auditory afferents,'' Hear. Res. **121**, 11–28.

McKay, C. M., and McDermott, H. J. (**1994**). ''Loudness summation for two channels of stimulation in cochlear implants: effects of spatial and temporal separation,'' Ann. Otol. Rhinol. Laryngol. **104** (Suppl. 166), 230–233.

McKay, C. M., and McDermott, H. J. (**1998**). ''Loudness perception with pulsatile electrical stimulation: The effect of interpulse intervals,'' J. Acoust. Soc. Am. **104**, 1061–1074.

McKay, C. M., and McDermott, H. J. (**1999**). ''The perceptual effects of current pulse duration in electrical stimulation of the auditory nerve,'' J. Acoust. Soc. Am. **106**, 998–1009.

Moore, B. C. J., and Glasberg, B. R. (**1997**). ''A model of loudness perception applied to cochlear hearing loss,'' J. Acoust. Soc. Am. **3**, 289–311.

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1996**). ''Detection of decrements and increments in sinusoids at high overall levels,'' J. Acoust. Soc. Am. **99**, 3669–3677.

Pfingst, B. E., Morris, D. J., and Miller, A. L. (**1995**). ''Effects of electrode configuration on threshold functions for electrical stimulation of the cochlea,'' Hear. Res. **85**, 76–84.

Shannon, R. V. (**1983**). ''Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction,'' Hear. Res. **12**, 1–16.

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.

Tong, Y. C., and Clark, G. M. (**1986**). ''Loudness summation, masking, and temporal interaction for sensations produced by electrical stimulation of two sites in the human cochlea,'' J. Acoust. Soc. Am. **79**, 1958–1966.

Vandali, A. E., Whitford, L. A., Plant, K. L., and Clark, G. M. (**2000**). ''Speech perception as a function of electrical stimulation rate using the Nucleus 24 cochlear implant system,'' Ear Hear. **21**, 608–624.

Viemeister, N. F. (**1979**). ''Temporal modulation transfer functions based upon modulation thresholds,'' J. Acoust. Soc. Am. **66**, 1364–1380.

Zeng, F-G., and Shannon, R. V. (**1992**). ''Loudness balance between electric and acoustic stimulation,'' Hear. Res. **60**, 231–235.

Zwicker, E., and Scharf, B. (**1965**). ''A model of loudness summation,'' Psychol. Rev. **72**, 3–26.

Zwislocki, J. J. (**1969**). ''Temporal summation of loudness: an analysis,'' J. Acoust. Soc. Am. **46**, 431–441.

Reconciling frequency selectivity and phase effects in masking

Andrew J. Oxenham[a)]
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Torsten Dau[b)]
Hearing Research Center, Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215

(Received 27 October 2000; revised 10 April 2001; accepted 22 June 2001)

The effects of auditory frequency selectivity and phase response on masking were studied using harmonic tone complex maskers with a 100-Hz fundamental frequency. Positive and negative Schroeder-phase complexes (m_+ and m_-), were used as maskers and the signal was a long-duration sinusoid. In the first experiment, thresholds for signal frequencies of 1 and 4 kHz were measured as a function of masker bandwidth and number of components. A large difference in thresholds between the m_+ and m_- complexes was found only when masker components were presented ipsilateral to the signal over a frequency range wider than the traditional critical band, regardless of the absolute number of components. In the second experiment, frequency selectivity was measured in harmonic tone complexes with fixed or random phases as well as in noise, using a variant of the notched-noise method with a fixed masker level. The data showed that frequency selectivity is not affected by masker type, indicating that the wide listening bandwidth suggested by the first experiment cannot be ascribed to broader effective filters in complex-tone maskers than in noise maskers. The third experiment employed a novel method of measuring frequency selectivity, which has the advantage that the overall level at the input and the output of the auditory filter remains roughly constant across all conditions. The auditory filter bandwidth measured using this method was wider than that measured in the second experiment, but may still be an underestimate, due to the effects of off-frequency listening. The data were modeled using a single-channel model with various initial filters. The main findings from the simulations were: (1) the magnitude response of the Gammatone filter is too narrow to account for the phase effects observed in the data; (2) none of the other filters currently used in auditory models can account for both frequency selectivity and phase effects in masking; (3) the Gammachirp filter can be made to provide a good account of the data by altering its phase response. The final conclusion suggests that masker phase effects can be accounted for with a single-channel model, while still remaining consistent with measures of frequency selectivity: effects that appear to involve broadband processing do not necessarily require across-channel mechanisms. [DOI: 10.1121/1.1394740]

PACS numbers: 43.66.Dc, 43.66.Nm, 43.66.Ba [MRL]

I. INTRODUCTION

In a pair of papers, Smith *et al.* (1986) and Kohlrausch and Sander (1995) showed that two harmonic tone complexes with identical power spectra and very similar flat temporal envelopes could produce masked thresholds for long-duration tones that differed by as much as 20 dB. As these waveforms were based on equations proposed by Schroeder (1970), they have come to be known as Schroeder-phase complexes (Kohlrausch and Sander, 1995; Alcántara *et al.*, 1996; Carlyon and Datta, 1997a,b; Summers and Leek, 1998). These complexes are comprised of a series of equal-amplitude harmonic tones, with the starting phase of component n assigned according to the following equation:

$$\theta_n = \pm \pi n(n-1)/N, \tag{1}$$

where N is the total number of components. Complexes derived using the "+" sign are known as Schroeder positive (m_+), while complexes derived using the "−" sign are known as Schroeder negative (m_-).[1] As pointed out by Smith *et al.* (1986), these stimuli can be thought of either as harmonic tone complexes with flat temporal envelopes, or as a series of repeating linear frequency sweeps, with the m_+ complex producing downward sweeps and the m_- complex producing upward sweeps.

These complexes are interesting stimuli for psychoacoustic experiments for a number of reasons. First, because they both have the same power spectrum but can produce very different masked thresholds, they contradict the power-spectrum model of masking (Fletcher, 1940; Patterson, 1976; Glasberg and Moore, 1990). Second, these complexes might provide some insight into the phase response of the auditory filters. In particular, it has been proposed that auditory filtering, probably already at the level of the basilar-membrane (Recio and Rhode, 2000), alters the waveforms such that m_+ complexes result in a highly modulated, or "peaky," enve-

a)Electronic mail: oxenham@mit.edu
b)Now at: Arbeitsgruppe Medizinische Physik, Universität Oldenburg, 26111 Oldenburg, Germany; electronic mail: torsten.dau@medi.physik.uni-oldenburg.de

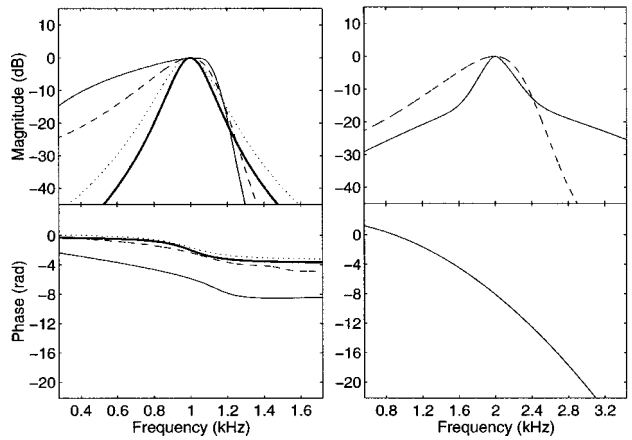


FIG. 1. The magnitude and phase responses of the filters used in the simulations are shown in upper and lower panels, respectively. The left panels show the responses of the Gammatone (heavy solid curve), 60-dB and 40-dB Gammachirp (dashed and dotted curves, respectively), and Strube's transmission-line (light solid curve) filters, all centered at 1 kHz. The upper-right panel shows the magnitude response of the fitted filter (solid curve; see text for details) and the 60-dB Gammachirp filter (dashed curve) centered at 2 kHz, and used in modeling the results from Experiment 3. The phase response shown in the lower-right panel has the same curvature as the m_- complex used in Experiment 3.

lope, while m_- complexes produce a less modulated, flatter, envelope. This in turn places certain constraints on the form of the auditory system's phase response (Kohlrausch and Sander, 1995). Certain models of basilar-membrane filtering (e.g., Strube, 1985; Giguère and Woodland, 1994a) support this view. For instance, Kohlrausch and Sander (1995) show in their Fig. 18 that the output of Strube's (1985) model is more highly modulated in response to an m_+ complex than to an m_- complex. This is in contrast to the response of a Gammatone filter (Patterson *et al.*, 1995), which shows essentially the same degree of modulation in response to both (Kohlrausch and Sander, 1995, Fig. 17). One reason for the qualitative success of BM models in accounting for the perceptual difference between m_+ and m_- complexes is that the phase response of these models has a negative curvature throughout most of the passband. In the case of the m_+ complex, this negative curvature compensates for, or ''cancels,'' the positive curvature of the complex, leading to a filtered waveform in which the starting phases of all the components come close to coinciding, giving a peaky envelope. In contrast the Gammatone filter, although generally very successful in many psychoacoustic applications, has a curvature that changes from being negative to positive at the center frequency (CF), meaning that it can never completely compensate for the curvature of the m_+ complex (Kohlrausch and Sander, 1995). The lower left panel of Fig. 1 illustrates this property by plotting the phase response of a segment of the (linear) Strube transmission-line model with a CF of 1 kHz (solid light curve), compared with the response of a fourth-order Gammatone filter with the same CF (solid heavy curve). The other two curves (dotted and dashed) are from Gammachirp filters (Irino and Patterson, 1997), discussed later.

While BM models provide a satisfying qualitative account, one troubling aspect is that their frequency selectivity is much poorer than that normally associated with the human auditory system. To illustrate this, the upper-left panel of Fig. 1 shows the magnitude responses of the respective filters. Again, the transmission-line and Gammatone filters are denoted by the light and heavy solid lines, respectively. The frequency selectivity of the Giguère and Woodland's (1994a, b) nonlinear model at medium and high levels is similar to that of the Strube (1985) transmission-line filter. While it has been speculated that the phase response of the BM filters, rather than the magnitude response, is primarily responsible for the good description of the m_+/m_- difference by these models (Kohlrausch and Sander, 1995; Alcántara *et al.*, 1996), this has not been demonstrated. An indication that broader frequency selectivity may be necessary to account for large m_+/m_- differences can be found in the data of Carlyon and Datta (1997b). They found that thresholds were affected by masker components that fell outside what is normally considered the passband of the auditory filter. For instance, thresholds in their m_+ condition increased when components half an octave and more away from the signal frequency of 1.1 kHz were removed, even though the phase and amplitude of the more central components were unchanged. Using a model (Giguère and Woodland, 1994a) with much broader frequency selectivity than the Gammatone filter, Carlyon and Datta were able to provide a reasonable description of their results. It seems unlikely that a similarly good description would have been possible with a Gammatone filter, even if its phase response had been altered to provide negative curvature throughout the passband.

Carlyon and Datta (1997b) attributed the effect of the masker bandwidth to the absolute number of components present, as this limits the maximum peakiness of the envelope. However, it is possible that the limiting factor was not the number of components, but rather the bandwidth of the masker, relative to the signal frequency. It may be, for instance, that the presence of ''off-frequency'' components is necessary to elicit a large m_+/m_- difference, as suggested by Buss *et al.* (1998). This explanation is more likely if the effects of nonlinear interactions, such as suppression (Oxenham and Plack, 1998), or across-frequency mechanisms, such as in comodulation masking release (CMR) (Hall *et al.*, 1984), play a role in determining thresholds. The role of such mechanisms could explain why it is necessary when using a single-channel model to assume broader frequency selectivity than is necessary in most other situations. No attempts have been made so far to reconcile frequency selectivity and seemingly broadband phase effects within the same model.

The first experiment studies the effect of masker bandwidth on thresholds in m_+ and m_- maskers. By using two signal frequencies (1 and 4 kHz), it was possible to dissociate the effects of an absolute number of components and relative bandwidth. In some conditions, by presenting the off-frequency components to the ear contralateral to the signal, we were able to examine the possible role of higher-level, across-channel mechanisms in determining thresholds.

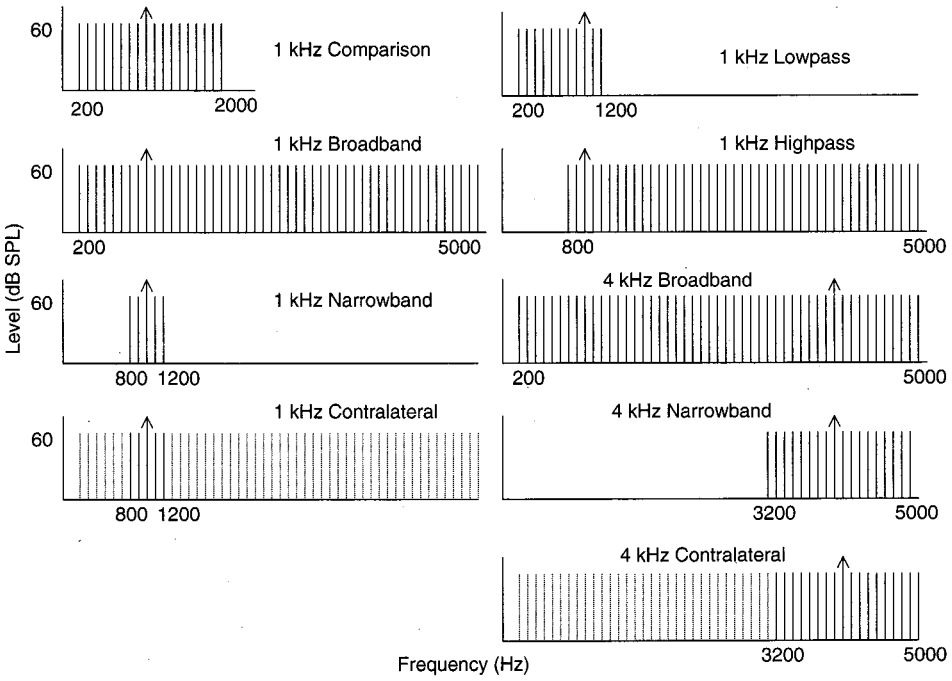


FIG. 2. Schematic diagram of the stimuli used in Experiment 1. Dashed lines denote components presented contralaterally.

II. EXPERIMENT 1. EFFECTS OF NUMBER OF COMPONENTS AND MASKER BANDWIDTH

A. Stimuli and conditions

The masker was a 320-ms (total duration) harmonic tone complex with a fundamental frequency (f_0) of 100 Hz, gated on and off with 10-ms raised cosine ramps. The level of the masker was set to 60 dB SPL per component. The sinusoidal signal was temporally centered within the masker and had a total duration of 260 ms, gated with 30-ms raised-cosine ramps. The signal frequency (f_s) was either 1 or 4 kHz. In all but one condition, the masker was derived from a harmonic tone complex with equal-amplitude frequency components from 200 to 5000 Hz, with the phase of each masker component selected based on Eq. (1) with $N=49$. One comparison condition was run using a 1-kHz signal, in which the masker was derived using components from 200 to 2000 Hz ($N=19$). For every condition, signal thresholds were measured in both m_+ and m_- configurations. The signal was always added in phase with the masker component at the signal frequency. Figure 2 shows a schematic diagram of the power spectra of the conditions tested. In certain conditions some of the masker components were eliminated, but the starting phases of the remaining components were not changed. The dotted lines indicate components presented to the ear contralateral to the signal.

The 1-kHz narrowband condition is similar to one tested by Carlyon and Datta (1997b). Based on their results, we expect the m_+/m_- difference to be reduced relative to the broadband condition. The 4-kHz narrowband masker has about the same relative bandwidth as the 1-kHz narrowband masker, but the same number of components as in the 1-kHz comparison condition. If, as suggested by Carlyon and Datta (1997b), the number of components is critical, then the 19 components of the 4-kHz narrowband masker should be sufficient to produce a large m_+/m_- difference. On the other hand, if a relatively wide masker bandwidth is necessary regardless of the number of components, then the m_+/m_- difference should be small in the 4-kHz narrowband condition.

The two contralateral conditions present the narrowband masker and the signal to one ear, and the remaining components from the broadband masker to the other ear. If higher-level across-channel processing is partly responsible for the m_+/m_- difference, then we might expect the contralateral components to have some effect. On the other hand, if the difference is mediated solely by peripheral mechanisms, then the contralateral components would not be expected to have any effect on thresholds.

The 1-kHz highpass and lowpass conditions test which range of components contribute more to the m_+/m_- difference. The 1-kHz comparison condition was included to provide a direct comparison with previous studies, many of which have used very similar parameters (Kohlrausch and Sander, 1995; Carlyon and Datta, 1997a), and with the 4-kHz narrowband condition, which used the same number of components.

In pilot tests, it was noticed that distortion products could be heard in the 4-kHz narrowband condition and that they were modulated by the presence of the signal. Because of this, a lowpass pink noise (2.5-kHz cutoff frequency) was added at a level of 38 dB SPL per 1/3 octave band to all the 4-kHz conditions in order to mask any distortion products. The stimuli were generated digitally and played out using a LynxOne sound card at a sampling rate of 32 kHz. The stimuli were passed through a programmable attenuator (TDT PA4) and headphone buffer (TDT HB6) before being presented to the listener via Sennheiser HD 580 headphones. The stimuli (except for the contralateral masker components) were presented to the listener's left ear. Listeners were seated in a double-walled sound-attenuating booth.

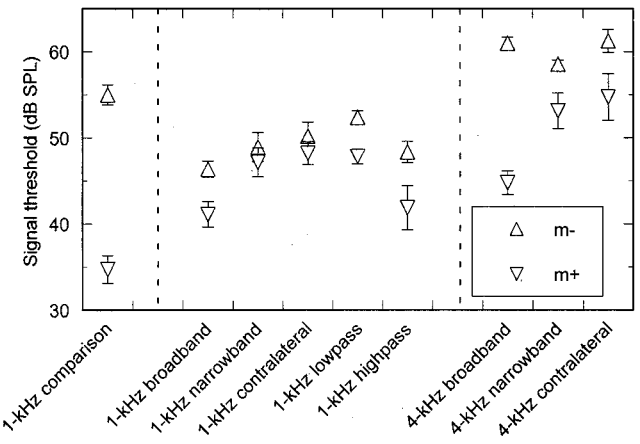


FIG. 3. Mean data from Experiment 1. Error bars denote ±1 standard error of the mean.

B. Procedure

An adaptive three-interval three-alternative forced-choice (3AFC) procedure was used in conjunction with a 2-down 1-up tracking rule to estimate the 70.7% correct point on the psychometric function (Levitt, 1971). Each interval in a trial was separated by an interstimulus interval (ISI) of 500 ms. The intervals were marked on a computer monitor and feedback was provided after each trial. Listeners responded via the computer keyboard or mouse. The initial step size was 5 dB, which was reduced to 2 dB after the first 2 reversals. The threshold was defined as the mean of the remaining 8 reversals. Four threshold estimates were obtained from each listener in each condition.

C. Subjects

Four listeners participated as subjects in this experiment. Two were the authors, one was a graduate student in the first author's laboratory, and the other was paid for her services. The age of the subjects ranged from 24 to 34 years. All had thresholds of less than 15 dB HL at octave frequencies between 250 and 8000 Hz. The two less experienced subjects received at least 6 h training before data were collected.

D. Results

The trends of the four listeners were very similar, and so only the mean data from Experiment 1 are shown in Fig. 3. Thresholds in the presence of the m_+ and m_- maskers are shown as downward-pointing and upward-pointing triangles, respectively. The error bars denote ±1 standard error of the mean across subjects.

In the comparison condition, the difference between the two masker types is about 20 dB, which is consistent with previous studies using similar stimulus parameters (Kohlrausch and Sander, 1995; Summers and Leek, 1998). The difference in the 1-kHz broadband condition is much smaller (5.3 dB). The increase in thresholds for the m_+ complex, relative to the comparison condition, is probably in part because the sweep rate (the change in group delay as a function of frequency, or phase curvature) of the m_+ complex does not match the curvature of the auditory filters at 1 kHz as well as does the reference condition. The Schroeder-phase complex produces a linear frequency sweep, meaning that the phase curvature ($d^2\theta/df^2$) is constant and independent of frequency (f). The curvature of the m_+ complex is determined by f_0 and the number of components (N) (Kohlrausch and Sander, 1995):

$$\frac{d^2\theta}{df^2} = \frac{2\pi}{Nf_0^2}. \tag{2}$$

Increasing the number of components from 19 to 49 therefore leads to a decrease in the curvature by a factor of more than 2.5.

The *decrease* in thresholds for the m_- 1-kHz broadband condition, relative to the comparison condition, probably cannot be explained in these terms. Instead, it is helpful to consider that the proportion of time within each masker cycle in which the instantaneous frequency falls within the bandpass region of the auditory filter centered at 1 kHz is reduced when the bandwidth of the stimulus is increased. This means that, regardless of the local curvature of the complex, the frequency sweep will excite the auditory filter for a smaller proportion of the period, leading to a modulated envelope for the m_- complex, and hence to lower thresholds.

Unfortunately, the 1-kHz broadband condition does not provide a very good baseline condition, as the m_+/m_- difference is already rather small. Nevertheless, as expected, the m_+/m_- difference for the 1-kHz narrowband condition is even smaller. The m_+/m_- difference for the 1-kHz contralateral condition is about the same as for the narrowband condition, suggesting that the contralateral components did not influence performance. The 1-kHz lowpass condition shows thresholds that are elevated with respect to the broadband condition for both m_+ and m_- complexes, while the highpass results are very similar to those of the broadband condition.

The 4-kHz broadband condition showed a substantial m_+/m_- difference of about 16 dB. This is similar to the difference found by Summers and Leek (1998) for a similar stimulus configuration. The m_+/m_- difference for the 4-kHz narrowband condition was much smaller (5.4 dB), despite the fact that the number of components was the same as in the 1-kHz comparison condition. This suggests that masker components interacting over a relatively wide frequency region may be required to produce a large m_+/m_- difference. As at 1 kHz, the 4-kHz contralateral condition does not appear to differ from the 4-kHz narrowband condition, again suggesting no influence of the contralateral masker components.

In summary, a large m_+/m_- difference requires the presence of masker components over a frequency range wider than a traditional critical band (e.g., Scharf, 1970). This makes it seem intuitively unlikely that single-channel models, with sufficient frequency selectivity to predict other psychoacoustic data, could be used to predict these differences. However, as intuition can be misleading in such cases, quantitative modeling was used to test the ability of current models to predict the data. The following section uses a va-

riety of filter models, combined with the detection model of Dau *et al.* (1997a), in an attempt to describe the data from Experiment 1.

III. MODEL PREDICTIONS FOR EXPERIMENT 1

A. Description of the model

The model used here is based on the single-channel version of the model described by Dau *et al.* (1997a, b). This model has been successful in predicting a variety of simultaneous- and forward-masking data, as well as modulation-detection data. It is assumed that the signal is detected by a filter centered at the signal frequency and only the output from that filter is processed further. Briefly, the filtered stimulus is half-wave rectified and passed through a series of feedback loops, which simulate certain aspects of auditory-nerve adaptation. The output from the adaptation loops is passed through a bank of modulation filters. The output from the modulation filterbank is compared with an internal template of the signal, and decisions are derived using a version of an optimal detector. The model's ability to describe detection in the presence of temporally fluctuating maskers, and the ability of the detector to make use of information in the masker valleys, make it a good candidate for predicting the present data. All model parameters were identical to the ones used by Dau *et al.* (1997a).

A number of different front-end filters were used with the model. The first was the linear transmission-line model of Strube (1985). This was used in the first version of the model (Dau *et al.*, 1996a,b) and a nonlinear version (Strube, 1986) has also been used to predict data on vowel identification using Schroeder-phase stimuli (Alcántara *et al.*, 1996).[2] The second was a fourth-order Gammatone filter, which has been used in numerous models of auditory processing (e.g., Meddis and Hewitt, 1991; Dau *et al.*, 1997a; Ellis, 1999; Godsmark and Brown, 1999). Third, two versions of the Gammachirp filter (Irino and Patterson, 1997) were tested. The filter parameters were taken from Table II of Irino and Patterson (1997), averaged across all subjects at each center frequency. The two versions used parameters corresponding to signal levels of 40 and 60 dB SPL, which lie near the extremes of the signal levels measured in Experiment 1. The magnitude and phase responses of all four filters at 1 kHz are shown in the left-hand panels of Fig. 1. Finally, a filter was tested which combined the phase response of the Strube filter with the magnitude response of the Gammatone filter. This was done to test the idea that the phase response, rather than the magnitude, was critical in being able to predict the m_+/m_- difference (Kohlrausch and Sander, 1995). The 3-dB bandwidths and the equivalent rectangular bandwidths (ERBs) of these filters are given in Table I.

Simulations were run separately for each filter type. The stimuli were calculated in the same way as in the actual experiment and the model acted as a subject, being presented with three intervals, one of which contained the signal. The adaptive procedure was also the same as that used in the actual experiment. The predictions are the mean of four threshold estimates for each condition. Predictions were not made for the contralateral conditions, as the results would have been identical to those in the narrowband conditions.

TABLE I. Description of filters used in the simulations of Experiment 1.

	Center frequency (Hz)	3-dB bandwidth (Hz)	ERB (Hz)
Strube transmission-line filter	1000	365	407
	2000	701	803
	4000	1217	1452
Gammachirp 60 dB	1000	214	249
	2000	516	582
	4000	959	1114
Gammachirp 40 dB	1000	153	175
	2000	374	426
	4000	701	800
Gammatone	1000	116	133
	2000	208	241
	4000	406	457

B. Model results

The predictions from the five filter versions are shown in Fig. 4. As expected based on previous studies (Kohlrausch and Sander, 1995; Alcántara *et al.*, 1996; Carlyon and Datta, 1997b), the transmission-line model is able to predict an m_+/m_- difference in the comparison condition, although it is smaller than that observed experimentally (11.5 dB compared with 20 dB). The correct trend is also observed in some of the other conditions. Some discrepancies are apparent, however. For instance, the 1-kHz lowpass and 1-kHz broadband predictions are very similar and are different from the predictions for the 1-kHz narrowband and 1-kHz highpass conditions. This is expected, given the highly asymmetric magnitude response of the filter, where much more weight is given to frequency components below the CF than above. However, these trends were not observed in the data. The predicted m_+/m_- difference of 8 dB in the 4-kHz broadband condition is again less than the observed value of 16 dB. Also, the predicted difference between the 4-kHz broad- and narrowband conditions consists mainly of a decrease in the predicted m_- threshold. This can probably be explained by the reduction in the masker energy falling within the wide filter passband as the masker bandwidth is reduced, and because the elimination of many components results in a more modulated waveform. In contrast, the data show the opposite effect, with the main difference between the 4-kHz broad- and narrowband conditions being an increase in the m_+ threshold. The pattern of results for the Gammachirp filter at 60 dB SPL is very similar to that of the transmission-line model.

Both the 40-dB Gammachirp and the Gammatone filters fail to show a substantial m_+/m_- difference or any effect of stimulus bandwidth. Because the change in the Gammachirp parameters results in a change in both the magnitude and phase characteristics of the filter, it is not possible to say which is responsible for the difference in the predictions between the two versions. The final simulation, combining the Gammatone filter's magnitude response with the

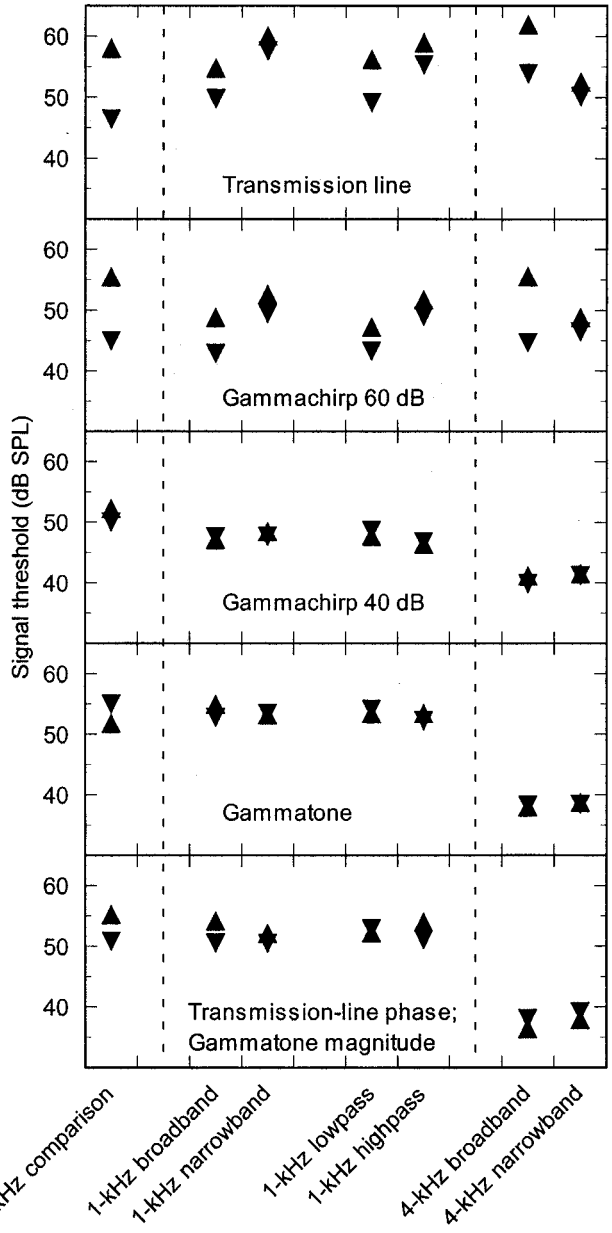


FIG. 4. Predictions of Experiment 1 using different auditory filter models. As in Fig. 3, upward- and downward-pointing triangles denote thresholds with the m_- and m_+ maskers, respectively.

transmission-line filter's phase response, answers that question. This hybrid filter shows an m_+/m_- difference of around 4.5 dB in the comparison condition, which lies in between the predictions of the transmission-line and Gammatone filters. Thus, even with a phase response that can in principle be used to predict an m_+/m_- difference, the passband of the Gammatone filter seems to be simply too narrow to allow a substantial difference to be predicted. It is interesting, however, that the small difference (3.6 dB) predicted in the 1-kHz broadband condition becomes 2 dB smaller in the 1-kHz narrowband condition. This implies that even with the frequency selectivity assumed by the Gammatone filter, components well outside its 3-dB bandwidth can still contribute to reducing thresholds. This is similar to the findings of Verhey *et al.* (1999).

C. Predictions using the temporal-window model

None of the filters tested provided very convincing fits to the data in conjunction with the model of Dau *et al.* (1997a). This could be an inherent problem with all the filters tested, or it could be a limitation of the detection model. To help address this, the data were simulated using the temporal-window model (Moore *et al.*, 1988; Plack and Moore, 1990; Oxenham and Moore, 1994; Oxenham and Plack, 2000). Stimuli were filtered using Strube's transmission-line filter, as that provided the best fits to the data in Experiment 1. The stimuli were then half-wave rectified, subjected to a power-law nonlinearity, and passed through a sliding temporal integrator, or temporal window. Threshold was determined by the maximum ratio between the window output due to the masker alone and the output due to the masker plus signal. A variety of window shapes and durations were tested in combination with a number of different power-law nonlinearities ($y=x^p$), ranging from very compressive ($p=0.1$) to energy detection ($p=2$). None of these combinations improved on the predictions of the Dau *et al.* model. This provides some indication that the failure of the model may be due to the initial filtering stage, rather than the particular post-filtering model we chose to use.

D. Discussion

A comparison of Figs. 3 and 4 shows that none of the filters can provide a completely satisfactory account of the data, at least when used in conjunction with the detection model of Dau *et al.* (1997a) or with the temporal-window model. The two broadest filters, Strube's transmission-line model and the 60-dB Gammachirp, provide the most convincing fits. This is somewhat disturbing, as the frequency selectivity associated with both filters is much broader than that exhibited by the Gammatone filter, which is generally believed to provide a good description of auditory frequency selectivity at low and medium sound pressure levels (Patterson, 1976; Glasberg and Moore, 1990; Derleth and Dau, 2000). In contrast to the speculation of Kohlrausch and Sander (1995), combining the amplitude response of the Gammatone filter with the phase response of the transmission-line model does not seem to be sufficient to produce predictions in line with the data.

There is a possibility, albeit an unattractive one, that the effective frequency selectivity of the auditory system is somehow different for harmonic tone complexes than for noise. Lentz *et al.* (1999) found similar auditory filter shapes when measuring thresholds both in noise and in tone complexes with logarithmically spaced components, although the estimated filter bandwidths for a profile-analysis task using the same complex-tone stimuli were considerably larger. To our knowledge, no direct measures of frequency selectivity have been made using harmonic tone complexes. The second experiment was designed to estimate frequency selectivity in noise and in harmonic tone complexes, with fixed or random phases, using a variant of the notched-noise method (Patterson, 1976; Glasberg and Moore, 1990).

IV. EXPERIMENT 2. FREQUENCY SELECTIVITY IN NOISE AND COMPLEX TONES

A. Methods

Many of the stimulus parameters and procedures were the same as in Experiment 1. Primarily those that differed are mentioned here. The signal frequency in this experiment was 2 kHz. This was chosen so that a masker f_0 of 100 Hz could still be used but the relative spectral notch could be manipulated in a more fine-grained way than would be possible at 1 kHz. The starting phase of the signal was random and was different in each trial. This was done because many conditions did not include a masker component at the signal frequency, and in order to make thresholds in the noise and tone complex maskers as comparable as possible. The harmonic tone complex masker was created with components from 100 Hz to 4000 Hz. In certain conditions, components around the signal frequency were eliminated. In the first part, only m_+ and Gaussian-noise maskers were tested. The noise masker had the same bandwidth as the tone complex and had a spectrum level of 40 dB SPL. This level was chosen as it was approximately equal to the average spectrum level of the tone complex (60 dB SPL per component, spaced 100 Hz apart). In the second part, three other maskers were also tested. These were an m_- complex, a cosine-phase complex, and a random-phase complex. In the random-phase complex, the phase of each component was taken from a uniform distribution on each interval of a trial. Spectral notches, placed symmetrically around the signal frequency, were introduced into the maskers. The distances between the signal frequency and the edge of the noise (or the nearest masker component) were 0.05, 0.1, 0.2, 0.3, and $0.4f_s$. For the harmonic tone complex maskers, the spectral notches were created by removing 1, 3, 5, 7, and 9 spectral components, respectively. For the noise masker, the notch was created by setting the magnitude values of the noise's Fourier transform to zero within the notch. Therefore, the steepness of the attenuation slopes at the cutoff frequencies was limited only by the gating of the maskers. The stimuli were presented diotically over Sennheiser HD 580 headphones.

An adaptive three-interval 3AFC procedure was used in conjunction with a 2-down 1-up tracking rule to estimate the 70.7% correct point on the psychometric function (Levitt, 1971). Each interval in a trial was separated by an ISI of 500 ms. The intervals were marked on a computer monitor and feedback was provided after each trial. Listeners responded via the computer keyboard or mouse. The initial step size was 5 dB, which was reduced to 2 dB after the first 4 reversals. The threshold was defined as the mean of the remaining 6 reversals. Three threshold estimates were obtained from each listener in each condition. Three of the four listeners from Experiment 1 (SD, TD, and AO) participated in the first part of this experiment. Only the authors (TD and AO) participated in the second part.

B. Results

The mean data from three listeners, using only the m_+ complex and the Gaussian-noise masker, are shown in the left panel of Fig. 5. Error bars denote ± 1 standard error of the mean across listeners. For the broadband maskers with no notch, the mean threshold in the m_+ condition is about 26 dB lower than in the noise condition. This difference is totally eliminated for all other conditions, and is even slightly reversed at the 0.05 deviation, where only the masker component at the signal frequency was removed. This can probably be understood in terms outlined by Duifhuis (1970, 1971). He found that a high-order harmonic in a pulse train was individually audible only if it was *not* present in the spectrum. This seemingly paradoxical result can be understood if one considers that the operation of removing a component is identical to adding one in antiphase to the original. The time waveform of such stimuli can be seen to be a pulse train with the addition of a sinusoid with a frequency of the removed component (see Fig. 5 of Duifhuis, 1970). In our case, the removal of the single component was equivalent to the addition of a 60-dB SPL sinusoidal masker at the signal frequency, which presumably had the effect of "filling up" the temporal valleys of the internal m_+ masker waveform. In the 0.05 notch condition, a tonal component at the signal frequency was heard in all intervals using the m_+ masker. As the signal had a random starting phase, the addition of the signal could result in an increase, decrease, or no change in the perceived level of that component, depending on the exact phase and level relationship between the masker component and the signal. Detecting a random level change in a tone is somewhat different from detecting the presence of an otherwise absent tone (Green, 1967), which may explain why the mean m_+ threshold for the 0.05 notch is actually higher than that for the noise.

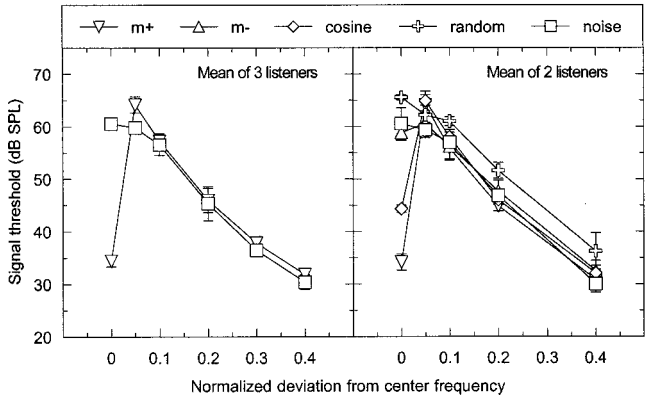


FIG. 5. Mean results from Experiment 2. The left panel shows the mean of three listeners with Gaussian noise (squares) and an m_+ tone complex (triangles) as maskers. The right panel shows the mean from a subset of two of the three listeners for a variety of different maskers (see the legend and the text for details).

Disregarding the no-notch and the 0.05 conditions, the data provide no evidence for different effective frequency selectivity in the presence of an m_+ tone complex. This finding was extended to other harmonic tone complexes, using the two authors as listeners, as can be seen in the right panel of Fig. 5. In the broadband (no-notch) condition, the m_+ complex (downward-pointing triangles) produces the least masking, with the cosine complex (diamonds) producing somewhat more masking, and all the other maskers producing similar amounts, with thresholds ranging from around 58 to 66 dB SPL. In the 0.05 notch condition, the m_+ and co-

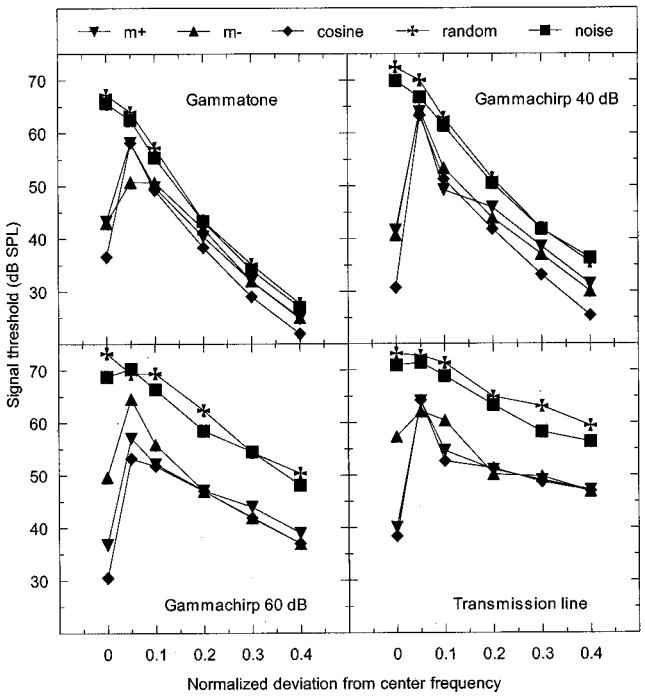


FIG. 6. Model predictions of Experiment 2.

sine maskers both produce the highest thresholds, presumably because both result in the percept of a Duifhuis pitch. Again, except for the 0 and 0.05 conditions, there is no evidence for differences in effective frequency selectivity in the presence of the different maskers. One trend, which was apparent for both listeners, is that the random-phase masker produces the most masking in almost all conditions, even though the dependence of thresholds on notch width is very similar to that found in the other conditions.

These results are reassuring: the idea that different stimuli require different filters would have been unparsimonious and difficult to justify within a physically realizable system. Nevertheless, the frequency selectivity demonstrated is rather good; increasing the total notch width from 200 Hz to 1600 Hz resulted in a mean threshold drop of about 30 dB. It seems unlikely that such good frequency selectivity could be achieved with either the transmission-line or the 60-dB Gammachirp filter. This was tested quantitatively in the following section.

C. Model predictions

The model was identical to the one described in Sec. III A. All four filters were tested again, using the stimuli from Experiment 2. This time, the Gammachirp filter parameters were taken from the 2-kHz parameters shown in Irino and Patterson (1997). The predictions are shown in Fig. 6. Considering only conditions with spectral notches, it is the Gammatone and 40-dB Gammachirp filters that provide the more convincing fits. Both show an average decrease in thresholds of around 30 to 35 dB, compared with around 20 dB for the 60-dB Gammachirp and 15 dB for the transmission-line filter. Furthermore, both of the broader filters predict much higher thresholds in the noise and random-phase conditions than in any of the other conditions. These differences are not observed in the data. None of the four filters predicts lower thresholds in the m_+ condition than in the cosine condition, while the data show about a 10-dB advantage.

The 60-dB Gammachirp filter proved to be somewhat too wide to account for the data from Experiment 2. The discrepancy between the 60-dB Gammachirp predictions and the data is probably primarily due to our method of measuring frequency selectivity, rather than a failure of the Gammachirp filter. We chose to keep the masker spectrum level constant while varying the level of the signal to achieve threshold, as has been done in most studies of frequency selectivity (Patterson, 1976; Patterson and Nimmo-Smith, 1980; Moore and Glasberg, 1983; Glasberg and Moore, 1990). In contrast, the parameters of the 60-dB Gammachirp filter (Irino and Patterson, 1997) were derived from the data of Rosen and Baker (1994), in which the signal level was held constant at 60 dB SPL, while the masker level for each notch width was varied to achieve threshold. Rosen and Baker (1994) and Rosen *et al.* (1998) have argued on both theoretical and empirical grounds that it is more appropriate to hold the signal level constant, thereby assuming that the filter shape is governed by the output level of the filter. The earlier method (e.g., Patterson and Nimmo-Smith, 1980) implicitly assumes that the filter shape is governed by the input sound level to the filter, i.e., the total sound energy, regardless of its frequency content. A recent study of Glasberg and Moore (2000) supports the conclusion of Rosen and colleagues that the filter shape is more accurately defined by keeping the signal level fixed, i.e., that the filter shape is determined by the filter output level. If the fixed-signal method provides a more accurate estimate of filter shape at a given sound level, then our results from Experiment 2 represent a composite filter shape, averaged across levels ranging from 60 to 35 dB SPL. This is expected to be considerably narrower than the ''true'' filter shape for a constant output level of 60 dB SPL, which is the basis of the 60-dB Gammachirp filter.

D. Discussion

Certain conclusions can be drawn from the data and simulations of Experiments 1 and 2. First, none of the filters correctly predicts the pattern of results in the no-notch condition. For instance, in contrast to the data, all filters resulted in lower predicted thresholds for the cosine-phase masker than for the m_+ masker. Also, whereas the data showed very little difference in thresholds between the noise masker and the m_- masker, the simulations predicted threshold differences of between 15 and 30 dB. These discrepancies suggest that the phase response of even the transmission-line or 60-dB Gammachirp filters may not be appropriate for modeling the phase response of the human auditory system. Second, the magnitude response of the Gammatone filter is too narrow to account for the phase effects observed in Experiment 1, but provides a good account of the frequency selectivity measured in Experiment 2. One interpretation of this, which is consistent with the work of Rosen and colleagues (Rosen and Baker, 1994; Rosen *et al.*, 1998), is that the ''true'' frequency selectivity in the presence of the maskers

used in Experiment 1 is considerably poorer (broader) than that implied by Experiment 2. Accordingly, the Gammatone filter with an ERB set to approximate the ERB as defined by Glasberg and Moore (1990), is probably too narrow to account for frequency selectivity with broadband stimuli at moderate sound levels of around 60 dB SPL. The considerably wider 60-dB Gammachirp filter is probably a better approximation.

V. EXPERIMENT 3. CONTRIBUTIONS OF INDIVIDUAL MASKER COMPONENTS

As discussed earlier, the assumption of the fixed-masker or fixed-signal notched-noise paradigm is that the input level or the output level of the filter, respectively, drives changes in filter shape. In fact, both methods (fixed-signal or fixed-masker level) provide a linear approximation to a nonlinear system and so it is highly unlikely that either approach is strictly correct. For instance, both approaches fail to take account of the contributions of suppression to masking (Delgutte, 1990; Moore and Vickers, 1997; Oxenham and Plack, 1998). In order to bypass the input- vs. output-level debate, it is necessary to find a method of measuring filter shape in which both the input and output levels of the filter remain roughly constant. This was achieved in Experiment 3. Noticing that removing the masker component at the signal frequency from the m_+ masker was sufficient to eliminate the m_+/m_- difference, we reasoned that the effect of removing any other single component from the masker could provide a measure of that component's effectiveness, and hence an indirect measure of its attenuation, relative to the central component. Experiment 3 therefore measured the effect of removing a single masker component from m_+ and m_- complexes as a function of the component's frequency.

A. Methods

1. Effect of component frequency

The procedure was the same as in Experiment 2, as was the signal (2 kHz) and the method of stimulus generation and presentation. The masker was a harmonic tone complex with an f_0 of 100 Hz and a level of 60 dB SPL per component. The components were in either m_+ or m_- phase, derived from a complex with components from 100 to 4000 Hz. The experiment investigated the effect of removing a single masker component as a function of the frequency separation between the masker component and the signal. For the m_+ complex, the separations tested were 0, $\pm$100, $\pm$200, $\pm$400, $\pm$800, and $\pm$1200 Hz. For the m_- complex, the effect was considerably smaller (as expected from Experiment 2) and so only separations of 0 and $\pm$400 Hz were tested. Thresholds in the broadband conditions were also tested, as were thresholds for a bandpass condition where only components between 800 and 3200 Hz (i.e., $\pm$1200 Hz) were present.

2. Effect of component level

As discussed in Sec. IV B, the absence of a given component can be treated as an addition of that component in the antiphase. The reasoning behind this experiment was that in the m_+ condition, the effect of the absence (or antiphasic addition) of a given component would decrease as the frequency separation between the component and the signal increased. This can be treated as a measure of the attenuation of a given component, relative to the component at the signal frequency. In other words, plotting the signal threshold as a function of the missing component frequency provides something akin to a masking pattern. However, the masking pattern provides only an indirect measure of filter attenuation. To derive the effective attenuation, it is necessary to know how signal thresholds change as a function of the masker component level at the signal frequency. This was done in the second part of the experiment, using the same signal and m_+ complex as in the first part, by adding a masker component at the signal frequency in antiphase to the original component at that frequency and measuring thresholds as a function of the antiphasic component level. Setting the level of the antiphasic component to be the same as that of the other masker components (0 dB) was equivalent to removing the masker component at that frequency completely. The three listeners who participated in Experiment 2 completed both parts of this experiment.

B. Results and discussion

The individual results from the second part of the experiment were fairly clear-cut. All three listeners showed a relationship between the antiphasic component level and the signal threshold that was very close to linear, at least down to a component level of -24 dB. Individual regression lines for the data down to and including -24 dB had slopes of -1.1, -0.87, and -0.97 for listeners AO, TD, and SD, respectively. The linear relationship between the signal threshold and effective on-frequency component level suggests that we can interpret changes in threshold as a function of signal–component frequency separation directly in terms of filter attenuation, down to levels of -24 dB. These results are plotted in the upper-left panel of Fig. 7. Thresholds in the presence of the m_+ and m_- complexes are shown as downward- and upward-pointing triangles, respectively. The dashed and dotted lines denote thresholds for the broadband m_+ and m_- complexes, respectively. Not shown are thresholds in the presence of the m_+ complex with components from 800 to 3200 Hz only (but with the same phases as in the broadband condition). They were not significantly different from the broadband m_+ thresholds for any listener (t-test, $p>0.05$), suggesting that components below 800 Hz or above 3200 Hz do not contribute to determining thresholds at a signal frequency of 2000 Hz.

In the m_- complex, the effect of removing a component is relatively small (about 5 dB) in the on-frequency condition, and is negligible at a frequency spacing of $\pm$400 Hz. This is presumably because the m_- waveform, once passed through the auditory filters, is not highly modulated and so there are no significant temporal valleys for the additional sinusoid to ''fill up.'' In contrast, removing the on-frequency component in the m_+ complex results in a threshold increase of more than 25 dB. The effect of removing a component decreases with increasing frequency separation, as expected, so that the effect is negligible at the largest separations of 1.2 kHz (0.6 in the normalized units of Experiment 2).

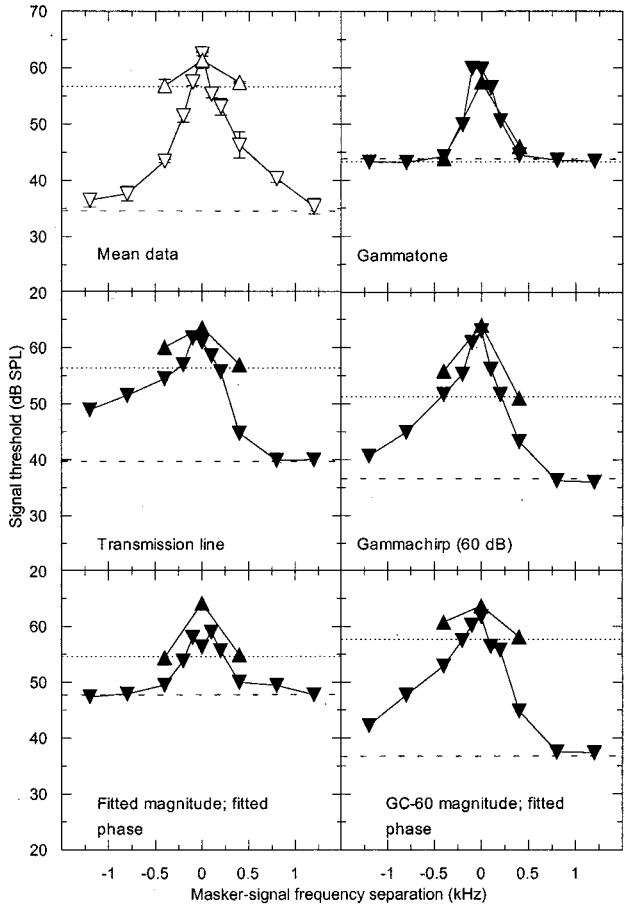


FIG. 7. Mean data (upper-left panel) and simulations from Experiment 3. Thresholds are plotted as a function of the frequency of the missing masker component. Upward- and downward-pointing triangles denote thresholds with the m_- and m_+ masker, respectively. Dotted and dashed lines denote thresholds for the m_- and m_+ maskers, respectively, with no missing components.

As a linear relationship was found between the on-frequency component level and signal threshold for attenuations up to 24 dB, the m_+ thresholds plotted in the upper-left panel of Fig. 7 can be interpreted directly in terms of auditory filter attenuation for all but the 1.2-kHz separation points. However, thresholds for the missing on-frequency component itself may require a more careful interpretation. In this case, a tonal component at the signal frequency was heard in all intervals. As discussed in Experiment 2, the addition of the signal could result in an increase, decrease, or no change, in the perceived level of that component, depending on the exact phase and level relationship between the masker component and the signal. This may explain the sharp increase in thresholds for the central data point.

Although the method used here has the advantage that the input and output level of the filter remain roughly constant, it has the disadvantage that it is not possible to rule out ''off-frequency listening'' (Patterson and Nimmo-Smith, 1980); O'Loughlin and Moore, 1981). Especially at higher signal levels, it may be that the signal is detected best in filters tuned away from the signal frequency. Because of this, the estimates of the filter bandwidth even with this new method may be somewhat too narrow. This is discussed further below.

C. Model predictions

The basic model tested here was the same as that used for Experiments 1 and 2. The filters were the transmission-line filter, the Gammatone, and a 60-dB Gammachirp, all centered at 2 kHz. The predictions are shown in the remaining panels of Fig. 7.

The transmission-line filter shows a highly asymmetric masking pattern, as would be expected from its magnitude response, which is similar to that shown for the 1-kHz transmission-line filter (Fig. 1). The decrease in thresholds above the signal frequency for the m_+ condition is similar to that observed in the data: thresholds are about 16 dB lower at the 0.4-kHz separation than at the peak in both the data and the predictions. However, the decrease in thresholds for components lower than the signal in frequency is far too shallow in the predictions: at the -1.2-kHz separation, predicted thresholds are only 12 dB lower than at the peak, compared with 26 dB in the data.

The 60-dB Gammachirp provides a reasonably good account of the frequency selectivity, thereby validating its use in Experiment 1, and supporting the idea that its failure in Experiment 2 was due to in part to a changing of the effective auditory filter shape in the experiment as the spectral notch widened. However, the tuning is still somewhat too broad on the low-frequency side. For instance, at the -0.4-kHz separation, predicted thresholds are 11 dB lower than at the peak, compared with 19 dB in the data. Some of this discrepancy may be due to off-frequency listening, not accounted for within the model.

The Gammatone filter (upper right panel) provides a reasonable account of frequency selectivity in the tip region of the filter: the decrease in m_+ thresholds is predicted to be about 16 dB for the 0.4- and -0.4-kHz separations, which is in line with the data. The small dynamic range of the predictions makes it impossible to evaluate the suitability of the Gammatone's response at wider frequency separations. As in the previous two experiments, the Gammatone fails to predict any real difference between the masked thresholds of the two complexes.

While the frequency selectivity shown by the 60-dB Gammachirp filter is reasonable, the predicted threshold difference between the broadband m_+ and m_- maskers is still too small. After these simulations had been completed, a new version of the Gammachirp filter was published (Irino and Patterson, 2001). Predictions with this new filter were in fact worse than for the original Gammachirp filter: the predicted difference in thresholds for the broadband m_+ and m_- maskers was only 7 dB, compared with 13 dB for the original Gammachirp, and 22 dB in the experimental data. As the magnitude responses of the old and new Gammachirp filters are very similar, it seems likely that the poorer predictions of the new Gammachirp filter are due to its phase response.

VI. DERIVING A NEW FILTER

The relatively good predictions using the 60-dB Gammachirp filter provide some hope that it may be possible to find a quasi-linear filter with reasonable frequency selectivity that can also be used to describe the phase properties of the

auditory system. However, even this filter exhibits frequency selectivity that is somewhat broader than that observed in the data, while predicting an m_+/m_- difference that is too small. It may be possible to ''fine-tune'' the phase and magnitude response of the filter so that predictions better match the data. In this section, we derive a magnitude response from the data of Experiment 3 and we combine it with an ''optimal'' phase response.

A. Magnitude response

The mean data from the m_+ complex in Experiment 3 were used to define the filter's magnitude response. In the second part of Experiment 3 it was found that the linear relationship between the on-frequency masker component level and the signal held for levels at and above -24 dB. Also, as discussed in Sec. IV B, the high threshold for the missing on-frequency component may be due to factors that did not play a role in any of the other conditions. For these reasons, data from the ± 1.2-kHz and the on-frequency conditions were excluded from the fit. The remaining mean data from the m_+ condition were fitted with a rounded exponential (roex) function (Patterson *et al.*, 1982). The version of the roex filter used here is known as the roex (p,w,t) filter and is described by two rounded exponential functions (one for the peak, and one for the tail) on either side of the function with a weighting factor, w, determining the relative contribution of each. It was assumed that the weighting function and the ratio between the two roex functions were the same on both sides of the filter, so the function had four free parameters, p_u and p_l to describe the slopes above and below the peak of the filter, respectively, w to describe the breakpoint, and tf, which is the factor by which the constant describing the tail of the filter is smaller than that describing the peak. Using a multi-dimensional minimization routine with a least-squares criterion, the best-fitting roex function was found. The final parameters were $p_l=33.11$, $p_u=26.74$, $tf=0.224$, and $w=0.0402$. The fit is rather good, with a rms error of only 0.78 dB, which is not surprising, given that 4 parameters were used to fit only 8 data points. The resulting ERB of the filter is 308 Hz. This value is somewhat higher than 241 Hz, as proposed by Glasberg and Moore (1990) and implemented in the Gammatone filter, but is still considerably lower than the ERB of the 60-dB Gammachirp filter (582 Hz). The derived magnitude response of the filter, based on the roex equations, is shown in the upper right panel of Fig. 1 (solid curve).

B. Phase response

With the magnitude response of the filter defined, the next step was to specify the phase response. Neither the absolute values of phase as a function of frequency, nor the slope of the function (the first derivative, corresponding to the group delay) can be determined behaviorally, the latter because delays in subsequent processing and reaction time would always swamp any delay introduced by cochlear filtering itself. The types of signal used in this study can, however, be employed to study the curvature, or second derivative, of the phase response (Kohlrausch and Sander, 1995). This is equivalent to the slope of the function relating group delay to frequency at a given CF. Given that our stimuli have constant curvature, our data cannot be sensitive to changes in curvature with frequency. It is therefore assumed that the phase curvature of the auditory filters can be approximated as being constant within the filter passband. Inspection of phase-response data from the basilar membrane (Ruggero *et al.*, 1997; Rhode and Recio, 2000) suggests that this may be a reasonable first approximation. Also, the frequency glides observed by Carney *et al.* (1999) in the impulse responses of auditory-nerve fibers were approximated by straight lines on a frequency–time plot, thereby implying a constant frequency sweep rate, or constant phase curvature. Another assumption is that the lowest predicted thresholds in the complex will occur when the phase curvature of the filter is equal in magnitude, but opposite in sign, to the phase curvature of the m_+ complex (Kohlrausch and Sander, 1995), thereby transforming the filtered m_+ complex into a complex in which all components have the same starting phase. In other words, the lowest thresholds in the m_+ condition will be predicted when the filter has a curvature equal to the m_- complex. Note that this filter curvature does not guarantee the largest predicted m_+/m_- *difference*, as this depends not only on the peakiness of the m_+ envelope, but also on the flatness of the m_- envelope. However, pilot simulations allowing the filter phase curvature to vary showed that the predicted m_+/m_- differences using the optimum phase were less than 0.5 dB larger than the differences predicted simply by using the phase curvature of the m_- complex. Therefore, the curvature of the m_- complex from Experiment 3, namely $-5\times10^{-6}\pi$ rad/Hz^2, was used to define the new filter's phase curvature. The realization of this curvature, as used in our simulations, is shown in the lower right panel of Fig. 1.

C. Predictions

The derived magnitude and phase response of the new filter were combined and an impulse response was calculated by applying an inverse Fourier transform. The new filter was then incorporated into the model of Dau *et al.* (1997a). The resulting predictions of the data from Experiment 3 are shown in the lower left panel of Fig. 7. While the frequency selectivity is in line with the data for thresholds above the asymptotic threshold (dashed and dotted lines), the model fails to describe either the large m_+/m_- difference or the large effect of adding an on-frequency antiphasic component at the signal frequency. Thus, even using a filter with a fitted frequency response and optimized phase response, the model of Dau *et al.* does not successfully predict the data.

One reason for this failure might be that our estimate of frequency selectivity is in error. For instance, as previously mentioned, off-frequency listening was not accounted for in the filter fitting procedure. As discussed in many previous studies (e.g., Patterson, 1976), this would lead to our estimate of the filter bandwidth being too narrow. Whatever the reason, it seems clear that the derived frequency selectivity is too narrow to allow the model to predict a sufficiently large m_+/m_- difference. For this reason, we returned to the 60-dB Gammachirp filter. The parameters of the Gammachirp filter (Irino and Patterson, 1997) were selected to fit

data obtained by Rosen and Baker (1994) using the notched-noise technique, which is designed to minimize off-frequency listening. As discussed earlier, the Gammachirp filter predicts somewhat broader tuning than observed in Experiment 3, but that may be primarily due to the model not taking off-frequency listening into account. The predicted m_+/m_- difference was also too small with the 60-dB Gammachirp filter, but that may be due to the filter's nonoptimal phase response. The next simulation combined the magnitude response of the 60-dB Gammachirp filter with the optimal phase response, as described above.

The predictions from combining the 60-dB Gammachirp filter's magnitude response and the fitted phase response are shown in the lower right panel of Fig. 7. The fit to the data is rather good. In particular, these are the first simulations of this study in which the m_+/m_- difference found in the data is matched well by the model. The predictions are an improvement on the predictions of the original 60-dB Gammachirp, primarily because of the higher predicted thresholds in the m-conditions. This results in an overall predicted m_+/m_- difference of about 21.7 dB, as opposed to 13.3 dB for the original 60-dB Gammachirp. A mean m_+/m_- difference of 22 dB was observed in the data. The results therefore provide the first evidence that predictions from a single-channel model may indeed be able to quantitatively match the results from such experiments, even without taking into account nonlinear effects such as suppression.

It is important to note that a single-filter model, with a bandwidth sufficiently narrow to predict traditional measures of frequency selectivity (Rosen *et al.*, 1998), was still able to predict the effects of components extending well beyond the traditional critical band. This is probably due to the highly modulated response to the m_+ masker, so that even highly attenuated components can affect the short-term power of the masker in its low-level epochs. This may explain the findings of Lin and Hartmann (1997), who reported that the strength of the Duifhuis (1970) pitch was influenced by components well outside the traditional critical band. Lin and Hartmann suggested that their results were evidence for the influence of across-channel processing. In contrast, our data and simulations suggest that their results may in fact be accounted for using a single-channel model.

At least two discrepancies between the data and the predictions remain. First, in Experiment 3 it was found that removing masking components below 800 Hz and above 3200 Hz had no effect on masked thresholds for the 2000-Hz signal. In the simulations, predicted thresholds in the m_+ masker increased by 2.5 dB when these components were removed. This suggests that the frequency selectivity assumed by the Gammachirp filter is still somewhat too broad: just as the constant masker level paradigm (Glasberg and Moore, 1990) may underestimate auditory filter bandwidth, the constant probe level paradigm (Rosen *et al.*, 1998) may overestimate it somewhat. Second, the data from Experiment 3 suggest a rather symmetric filter shape, whereas the shape of the 60-dB Gammachirp filter and its predictions are rather asymmetric. Overall, though, these discrepancies do not appear to represent fundamental failings of the model, and could probably be reduced with further fine-tuning of the filter shape.

D. Effects of peripheral compression

It has been shown by Carlyon and Datta (1997a,b) and Summers and Leek (1998) that stimulus level plays a major role in determining the m_+/m_- difference. Carlyon and Datta found that the difference was greatly reduced at low levels, an effect that could be due to narrower filters, more linear peripheral processing at low levels, a change in the filter's phase response at low levels, or a combination of these. Summers and Leek found that the m_+/m_- difference became somewhat smaller at very high levels, and was greatly reduced or even absent in listeners with sensorineural hearing loss, suggesting that cochlear nonlinearity may play an important role, and filter bandwidth (which should if anything be broader at high levels and in impaired listeners) is not the sole determining factor. The detection model used in the present study does not include any static nonlinearity, such as is observed on the basilar membrane (BM), and instead introduces time-dependent nonlinearities within the feedback loops. It is not clear to what extent these can be substituted for the effects of cochlear compression. To address this issue, further simulations were carried out using a version of the temporal-window model (Moore *et al.*, 1988), which incorporates a static compressive nonlinearity prior to temporal integration (Oxenham and Moore, 1994; Plack and Oxenham, 1998).

The temporal-window model was tested with the final model (60-dB Gammachirp magnitude response together with the ''optimal'' phase response). In Experiments 2 and 3, the signal was added in random phase to the masker. The temporal-window model in its current form is a deterministic model, and so cannot predict thresholds for randomly varying stimuli. For this reason, it was assumed that the signal could be approximated as adding, on average, in quadrature phase to the masker. This is the same as adding the intensities of the two stimuli and is basically the same procedure that was used by Oxenham *et al.* (1997). After filtering, the envelopes of the masker and signal were extracted using a Hilbert transform. The envelopes were squared to produce intensity-like quantities and then added before being subjected to a power-law compression (p). The value of p ranged from 1 (energy detection) down to 0.16, which is the value found by Oxenham and Plack (1997) at mid levels in their behavioral measure of BM compression. The stimuli were then passed through a temporal window of various shapes and durations. The smallest window was a Hanning window with a 2.5-ms half-amplitude duration. The longest was a rectangular window that spanned an entire period of the masker waveform (10 ms). Detection was assumed to occur at the instant in time when the ratio between the window output due to the masker-plus-signal and that due to the masker was maximal. This ratio was assumed to be the same for all conditions.

Predictions of Experiment 3 using the temporal-window model could be made similarly good to those of the Dau *et al.* model with the right combination of window shape and nonlinearity. The effect of nonlinearity was dependent on the

window size: the wider the temporal window, the more effect had the compression. This can be understood in terms of the differences between simultaneous and nonsimultaneous masking when using compression (Oxenham *et al.*, 1997). For very short windows, such as the 2.5-ms Hanning window, the predicted m_+/m_- difference with no missing components was larger than that observed in the data and was very similar for both the maximum ($p=0.16$) and minimum ($p=1$) compression applied (37.8 and 35.6 dB, respectively). For the 10-ms rectangular window, the predicted m_+/m_- differences were 26.8 and 0 dB for $p=0.16$ and $p=1$, respectively. The 0-dB prediction is the obvious outcome of an energy detector, integrating over a complete period of the masker, as both m_+ and m_- maskers have the same overall energy. Finally, we tested the double-exponential window employed by Plack and Oxenham (1998) and Oxenham and Plack (2000) to account for forward masking. This window, which has an equivalent rectangular duration of approximately 8 ms, predicted m_+/m_- differences of 28.4 dB and 0.2 dB for $p=0.16$ and $p=1$, respectively.

In summary, when using temporal integration windows of the shape and duration used in previous studies (Oxenham and Plack, 2000), the effects of peripheral compression are dramatic: ''normal'' compression can produce predicted m_+/m_- differences in excess of 25 dB, while linear processing produces essentially no difference. This finding provides quantitative support for the idea that differences in within-channel peripheral compression are sufficient to account for the differences in masker phase effects between normal-hearing and hearing-impaired listeners (Summers and Leek, 1998).

VII. SUMMARY

In Experiment 1 thresholds were measured in the presence of m_+ and m_- maskers at signal frequencies of 1 and 4 kHz. The two main findings were: (i) large threshold differences between m_+ and m_- maskers require a relatively wide masker bandwidth, irrespective of the number of masker components; and (ii) contralateral masker components have no effect on thresholds, suggesting no influence of across-channel mechanisms, at least beyond the level of binaural interaction.

Experiment 2 used a variant of the notched-noise technique to show that the effective frequency selectivity of the auditory system is the same for noise and harmonic complex tones. However, following Rosen and Baker (1994), it was concluded that the technique of varying the signal level while keeping the masker spectrum level constant probably underestimates auditory filter bandwidth at a given masker level.

Experiment 3 introduced a new method of estimating frequency selectivity, in which the masker level and frequency content remain roughly constant in all conditions, thereby bypassing any debate about whether filter input or output level governs filter shape. However, even this new method may underestimate the auditory filter bandwidth, as it does not take into account the possible effects of off-frequency listening.

The results from all three experiments were simulated using a detection model proposed by Dau *et al.* (1997a) with a variety of front-end filters, including the Gammatone, the Gammachirp, and the transmission-line model proposed by Strube (1985). None of these filters was able to provide a convincing fit to the data. The data and simulations suggest that none of the current models of auditory filtering successfully matches both the frequency selectivity and the phase response of the human auditory system. The Gammatone filter has a bandwidth that is too narrow to account for the present data, regardless of its assumed phase response. The transmission-line filter bandwidth was too broad to account for results from frequency–selectivity experiments. The Gammachirp filter, while having a sufficiently wide bandwidth, does not mimic the phase response of the auditory system with sufficient accuracy to predict the large m_+/m_- differences found in the data.

The data from Experiment 3 were well predicted by combining the magnitude response of the Gammachirp filter with a phase response derived from the m_- stimulus. This shows that phase effects can, in principle, be accounted for with a single-channel model. More generally, some effects that appear to require across-channel processing (Lin and Hartmann, 1997) may in fact be accounted for with an appropriate single-channel model. Good predictions were also obtained using the same filter in conjunction with the temporal-window model. Eliminating the static compression from the temporal-window model was sufficient to completely eliminate any predicted m_+/m_- masking differences. The importance of peripheral compression in these simulations may explain why hearing-impaired listeners, who are thought to have little or no peripheral compression, also show very small m_+/m_- masking differences.

Finally, as all the tested models of peripheral filtering failed to produce quantitatively accurate predictions, these data promise to provide strong constraints for testing future models of peripheral auditory processing.

ACKNOWLEDGMENTS

This work was supported by NIH/NIDCD Grant No. R01 DC 03909 (A. J. O.) and by the Max Kade Foundation and the Deutsche Forschungsgemeinschaft (DFG) (T. D.). Stephan Ewert provided help with the programming. Comments by the reviewers, Larry Feth and Van Summers, greatly improved the manuscript.

[1]In fact, Smith *et al.* (1986) used the opposite designation. However, all studies since then have used the convention applied here.

[2]At levels at which a difference between m_+ and m_- thresholds is predicted, the response of Strube's nonlinear model is very similar to that of his linear version.

Alcántara, J. I., Holube, I., and Moore, B. C. J. (**1996**). ''Effects of phase and level on vowel identification: Data and predictions based on a nonlinear basilar-membrane model,'' J. Acoust. Soc. Am. **100**, 2382–2392.

Buss, E., Grose, J. H., and Hall, J. W. (**1998**). ''Fast-acting compression of auditory stimuli: A test with narrowband stimuli,'' Assoc. Res. Otolaryngol. Abs., Vol. 21, p. 798.

Carlyon, R. P., and Datta, A. J. (**1997a**). ''Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system,'' J. Acoust. Soc. Am. **101**, 3636–3647.

Carlyon, R. P., and Datta, A. J. (**1997b**). ''Masking period patterns of Schroeder-phase complexes: Effects of level, number of components, and phase of flanking components,'' J. Acoust. Soc. Am. **101**, 3648–3657.
Carney, L. H., McDuffy, M. J., and Shekhter, I. (**1999**). ''Frequency glides in the impulse responses of auditory-nerve fibers,'' J. Acoust. Soc. Am. **105**, 2384–2391.
Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997a**). ''Modeling auditory processing of amplitude modulation. I. 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,'' 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.
Delgutte, B. (**1990**). ''Physiological mechanisms of psychophysical masking: Observations from auditory-nerve fibers,'' J. Acoust. Soc. Am. **87**, 791–809.
Derleth, R. P., and Dau, T. (**2000**). ''On the role of envelope fluctuation processing in spectral masking,'' J. Acoust. Soc. Am. **108**, 285–296.
Duifhuis, H. (**1970**). ''Audibility of high harmonics in a periodic pulse,'' J. Acoust. Soc. Am. **48**, 888–893.
Duifhuis, H. (**1971**). ''Audibility of high harmonics in a periodic pulse. II. Time effects,'' J. Acoust. Soc. Am. **49**, 1155–1162.
Ellis, D. P. W. (**1999**). ''Using knowledge to organize sound: The prediction-driven approach to computational auditory scene analysis and its application to speech/nonspeech mixtures,'' Speech Commun. **27**, 281–298.
Fletcher, H. (**1940**). ''Auditory patterns,'' Rev. Mod. Phys. **12**, 47–65.
Giguère, C., and Woodland, P. C. (**1994a**). ''A computational model of the auditory periphery for speech and hearing research. I. Ascending path,'' J. Acoust. Soc. Am. **95**, 331–342.
Giguère, C., and Woodland, P. C. (**1994b**). ''A computational model of the auditory periphery for speech and hearing research. II. Descending paths,'' J. Acoust. Soc. Am. **95**, 343–349.
Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.
Glasberg, B. R., and Moore, B. C. J. (**2000**). ''Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise,'' J. Acoust. Soc. Am. **108**, 2318–2328.
Godsmark, D., and Brown, G. J. (**1999**). ''A blackboard architecture for computational auditory scene analysis,'' Speech Commun. **27**, 351–366.
Green, D. M. (**1967**). ''Additivity of masking,'' J. Acoust. Soc. Am. **41**, 1517–1525.
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.
Irino, T., and Patterson, R. D. (**1997**). ''A time-domain, level-dependent auditory filter: The gammachirp,'' J. Acoust. Soc. Am. **101**, 412–419.
Irino, T., and Patterson, R. D. (**2001**). ''A compressive gammachirp auditory filter for both physiological and psychophysical data,'' J. Acoust. Soc. Am. **109**, 2008–2022.
Kohlrausch, A., and Sander, A. (**1995**). ''Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets,'' J. Acoust. Soc. Am. **97**, 1817–1829.
Lentz, J. J., Richards, V. M., and Matiasek, M. R. (**1999**). ''Different auditory filter bandwidth estimates based on profile analysis, notched noise, and hybrid tasks,'' J. Acoust. Soc. Am. **106**, 2779–2792.
Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Lin, J.-Y., and Hartmann, W. M. (**1997**). ''On the Duifhuis pitch effect,'' J. Acoust. Soc. Am. **101**, 1034–1043.
Meddis, R., and Hewitt, M. (**1991**). ''Virtual pitch and phase sensitivity studied of a computer model of the auditory periphery. I: Pitch identification,'' J. Acoust. Soc. Am. **89**, 2866–2882.
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., Glasberg, B. R., Plack, C. J., and Biswas, A. K. (**1988**). ''The shape of the ear's temporal window,'' J. Acoust. Soc. Am. **83**, 1102–1116.
Moore, B. C. J., and Vickers, D. A. (**1997**). ''The role of spread of excitation and suppression in simultaneous masking,'' J. Acoust. Soc. Am. **102**, 2284–2290.
O'Loughlin, B. J., and Moore, B. C. J. (**1981**). ''Off-frequency listening: effects on psychoacoustical tuning curves obtained in simultaneous and forward masking,'' J. Acoust. Soc. Am. **69**, 1119–1125.
Oxenham, A. J., and Moore, B. C. J. (**1994**). ''Modeling the additivity of nonsimultaneous masking,'' Hear. Res. **80**, 105–118.
Oxenham, A. J., Moore, B. C. J., and Vickers, D. A. (**1997**). ''Short-term temporal integration: Evidence for the influence of peripheral compression,'' J. Acoust. Soc. Am. **101**, 3676–3687.
Oxenham, A. J., and Plack, C. J. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.
Oxenham, A. J., and Plack, C. J. (**1998**). ''Suppression and the upward spread of masking,'' J. Acoust. Soc. Am. **104**, 3500–3510.
Oxenham, A. J., and Plack, C. J. (**2000**). ''Effects of masker frequency and duration in forward masking: Further evidence for the influence of peripheral nonlinearity,'' Hear. Res. **150**, 258–266.
Patterson, R. D. (**1976**). ''Auditory filter shapes derived with noise stimuli,'' J. Acoust. Soc. Am. **59**, 640–654.
Patterson, R. D., Allerhand, M. H., and Giguère, C. (**1995**). ''Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform,'' J. Acoust. Soc. Am. **98**, 1890–1894.
Patterson, R. D., and Nimmo-Smith, I. (**1980**). ''Off-frequency listening and auditory filter asymmetry,'' J. Acoust. Soc. Am. **67**, 229–245.
Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). ''The deterioration of hearing with age: frequency selectivity, the critical ratio, the audiogram, and speech threshold,'' J. Acoust. Soc. Am. **72**, 1788–1803.
Plack, C. J., and Moore, B. C. J. (**1990**). ''Temporal window shape as a function of frequency and level,'' J. Acoust. Soc. Am. **87**, 2178–2187.
Plack, C. J., and Oxenham, A. J. (**1998**). ''Basilar-membrane nonlinearity and the growth of forward masking,'' J. Acoust. Soc. Am. **103**, 1598–1608.
Recio, A., and Rhode, W. S. (**2000**). ''Basilar membrane responses to broadband stimuli,'' J. Acoust. Soc. Am. **108**, 2281–2298.
Rhode, W. S., and Recio, A. (**2000**). ''Study of mechanical motions in the basal region of the chinchilla cochlea,'' J. Acoust. Soc. Am. **107**, 3317–3332.
Rosen, S., and Baker, R. J. (**1994**). ''Characterising auditory filter nonlinearity,'' Hear. Res. **73**, 231–243.
Rosen, S., Baker, R. J., and Darling, A. (**1998**). ''Auditory filter nonlinearity at 2 kHz in normal hearing listeners,'' J. Acoust. Soc. Am. **103**, 2539–2550.
Ruggero, M. A., 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.
Scharf, B. (**1970**). ''Critical bands,'' in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York).
Schroeder, M. R. (**1970**). ''Synthesis of low peak-factor signals and binary sequences with low autocorrelation,'' IEEE Trans. Inf. Theory **16**, 85–89.
Smith, B. K., Sieben, U. K., Kohlrausch, A., and Schroeder, M. R. (**1986**). ''Phase effects in masking related to dispersion in the inner ear,'' J. Acoust. Soc. Am. **80**, 1631–1637.
Strube, H. W. (**1985**). ''A computationally efficient basilar-membrane model,'' Acustica **58**, 207–214.
Strube, H. W. (**1986**). ''The shape of the nonlinearity generating the combination tone $2f_1-f_2$,'' J. Acoust. Soc. Am. **79**, 1511–1518.
Summers, V., and Leek, M. R. (**1998**). ''Masking of tones and speech by Schroeder-phase harmonic complexes in normally hearing and hearing-impaired listeners,'' Hear. Res. **118**, 139–150.
Verhey, J. L., Dau, T., and Kollmeier, B. (**1999**). ''Within-channel cues in comodulation masking release (CMR): Experiments and model predictions using a modulation-filterbank model,'' J. Acoust. Soc. Am. **106**, 2733–2745.

Spatial and temporal factors in auditory saltation

D. P. Phillips[a] and S. E. Hall
Hearing Research Laboratory, Department of Psychology, Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada

(Received 18 April 2001; accepted for publication 28 June 2001)

This report describes three experiments on auditory saltation, studied with click stimuli presented at interclick intervals (ICIs) from 30 to 240 ms. In experiment 1, subjects rated the strength of the saltation illusion evoked by trains of six monaural clicks (i.e., three presented to one ear followed by three to the other ear), and six dichotic clicks on which were imposed either a 500-μs interaural time difference or a 9-dB interaural level difference (ILD). The interaural disparity of the clicks was reversed at the midpoint of the train. Subjects reported equivalent strengths of saltation for the dichotic clicks, but weaker saltation for the monaural ones. These data indicated that saltation is supported by dichotic clicks. regardless of the stimulus manipulation used to generate the lateralized images. In experiment 2, subjects rated the strength of the saltation evoked by six click trains lateralized, in separate trials, by ILDs ranging from 9 to 30 dB, and by a train of monaural clicks. In all ILD conditions, the level of the clicks at the ''quiet'' ear was above click-detection threshold when presented monaurally. Saltation weakened with increasing ILD, and approximated that seen with monaural clicks when the ILD was 30 dB. These data indicated that for the range of ICIs used here, saltation weakened as the stimuli became more strongly lateralized. In experiment 3, the number of dichotic clicks preceding the disparity reversal was, in separate trials, varied from 3 to 10, and subjects were asked to detect the presence of spatial stationarity in the click train. By plotting the subjective ratings as a function of the length of the leading click train, it was shown that the temporal window within which the saltation effect operates varies between listeners, but is usually less than about 350–400 ms. [DOI: 10.1121/1.1396329]

PACS numbers: 43.66.Lj, 43.66.Mk, 43.66.Pn, 43.66.Qp [LRB]

I. INTRODUCTION

Nearly 30 years ago, Geldard and Sherrick (1972; Geldard, 1975) described an illusion of cutaneous sensation. Five mechanical taps of equal duration and interstimulus interval were delivered successively to a locus near the wrist, to a point near the mid-forearm and to a point near the elbow. Depending on the interstimulus interval and number of taps, the resulting percept was of an orderly succession of taps not only at the points of contact, but also at points between them. The illusion was originally described as a ''cutaneous rabbit'' (Geldard and Sherrick, 1972), because the perceptual experience was likened to a very small rabbit hopping up one's arm. A phenomenologically similar illusion has since been shown to be present in the visual (Geldard, 1976; Lockhead *et al.*, 1980) and auditory modalities (Bremer *et al.*, 1977; Thurlow and Oneson, 1984; Hari, 1995), and thus seems to be based on a very general mechanism of spatio-temporal perception (Shore *et al.*, 1998). The generality of the illusion across modalities has led to it being referred to as ''sensory saltation.''

The importance of this illusion is at least twofold. First, it is a case in which the perception of events late in a sequence somehow is able to influence the perception of events earlier in the sequence. Despite being a very general property of perception, this process has received little attention, and so merits exploration in its own right. Second, the lengths of the interstimulus intervals capable of supporting the illusion in the auditory modality appear to distinguish normal and dyslexic readers, with the illusion persisting at longer intervals in the impaired readers (Hari and Kiesilä, 1996). This suggests that the mechanisms(s) accessed by the illusion may be related to those implicated in accounts of developmental language disorders rooted in an auditory temporal processing problem (for review, see Tallal *et al.*, 1993; Farmer and Klein, 1995; Phillips, 1999).

The purpose of the present article is to report three experiments on auditory saltation using dichotic click stimuli. The first two experiments concern the spatial properties of the click stimuli capable of supporting the illusion. The third experiment examines the duration of the temporal window over which the perceptual integration process (i.e., the ''writing'' of the percept) is exerted in the dichotic paradigm. Our purpose is thus to provide a more detailed characterization of the auditory saltation effect than has previously been available. The completeness of this characterization will guide future explicit interrogation of the mechanisms that mediate the illusion.

II. EXPERIMENT 1

A. introduction

Hari (1995) and Shore *et al.* (1998) have described a robust auditory saltation illusion obtained using short trains of dichotic click stimuli. Both studies used trains of eight dichotic clicks, the first four of which were lateral-

[a]Author to whom correspondence should be addressed; electronic mail: ears@is.dal.ca

ized to one side by the imposition of an interaural time difference (ITD) favoring one ear, while the last four clicks were lateralized to the opposite side by an equivalent ITD favoring that ear. The strength of the illusion was greatest for short interclick intervals (ICSs; 30–60 ms) and was at its weakest for ICIs greater than about 120 ms. For effective click rates, subjects routinely reported an orderly, continuous progression of apparent click (intracranial) location for successive elements of the train. The illusion did not weaken with repeated exposure. Interestingly, both studies reported that the illusion was weak or absent for control conditions of monaural click trains (in which the initial clicks are presented to one ear and the final clicks are presented to the other ear), sometimes even at the highest click rates; rather, the percept was of four clicks lateralized to one side, followed by four clicks lateralized to the other, as was the description of percepts aroused by dichotic clicks presented with very long ICIs.

The studies by Hari (1995) and Shore *et al.* (1998) were thus in agreement that binaural stimulation is a prerequisite for a strong saltation effect. Both studies, however, revealed this dependence using clicks lateralized with an interaural temporal disparity (ITD). It is thus unclear whether auditory saltation requires dichotic stimuli *per se*, or clicks lateralized specifically by ITDs. That is, would the motion illusion be generated by clicks lateralized with interaural level differences (ILDs)? The fact that saltation occurs reliably using free-field click stimuli (Bremer *et al.*, 1977; Phillips *et al.*, 2001) does not answer the question, because it is unclear whether the location percepts of the free-field clicks were based on ITDs, ILDs, or both. Furthermore, it might be argued that monaural clicks (after Hari, 1995; Shore *et al.*, 1998) represent an extreme case of ILDs, and that the failure to obtain strong auditory saltation with monaural stimuli would thus suggest that clicks lateralized on the basis of ILDs would be ineffective in generating the motion illusion.

The purpose of experiment 1 was to address this issue directly. For a single set of nine listeners, we compared the strength of the saltation effect evoked by monaural clicks, clicks lateralized on the basis of 500-μs ITDs, and clicks lateralized on the basis of 9-dB ILDs.

B. Methods

1. Subjects

A total of nine adult subjects (two male) participated in some or all of the experiments. Ages ranged from 22 to 44. Six subjects were experienced listeners. Nine listeners participated in experiments 1 and 3. Five of those listeners also participated in experiment 2. Audiograms to 8.0 kHz were obtained on all but one listener. All but one of those tested had normal hearing to 8.0 kHz. The exceptional male listener had a mild loss at 8.0 kHz.

2. Stimuli

All stimuli were made up of clicks, each composed of a single positive-going rectangular pulse produced in Macromedia's SOUNDEDIT™. Each pulse had a duration of two samples (45.35 μs). Binaural clicks with an interaural time

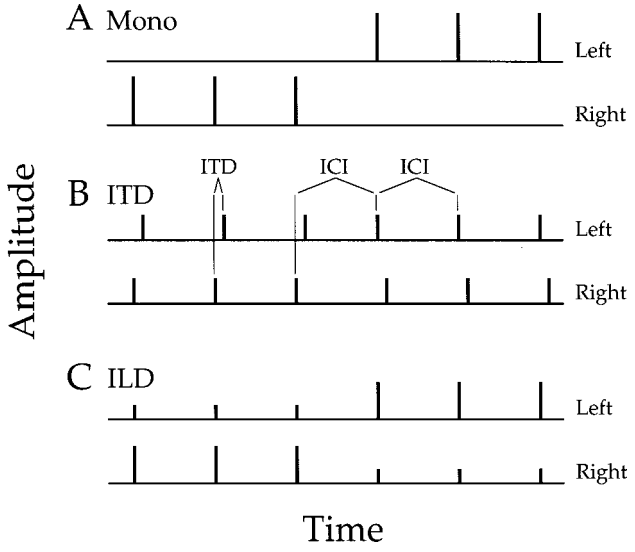


FIG. 1. Schematic illustration of the general structure of the stimuli used in experiment 1. Detailed description in the text.

difference (ITD) were produced by imposing a 500-μs onset delay between the members of the otherwise-identical click pair in the left and right channels. Binaural clicks with an interaural level difference (ILD) were produced by altering the relative amplitudes of the otherwise-simultaneous click pair.

Each complete stimulus consisted of an equally spaced train of six clicks, at various interclick intervals (ICIs). ICI here is defined as click onset to click onset. Figure 1 shows the three types of click trains used in this study. Additional stimuli were studied in experiment 3, in which the number of leading clicks (prior to the change in lateralization) ranged from three to ten.

Panel (A) shows a monaural (''mono'') stimulus, which is simply one in which the first three clicks come from one channel and the last three from the other channel (in this case, R→L). Mono stimuli were included as controls to experiments 1 and 2. Panel (B) shows an ITD stimulus, in which the first three of the binaural click pairs are right-channel leading, followed by three left-leading click pairs. Note that the amplitude of each individual rectangular pulse is exactly one-half that of the pulses in the mono stimulus. This was done in order to equate roughly the loudness of the clicks in the two types of trains. Panel (C) shows an ILD stimulus. As in panels (A) and (B), this stimulus train begins with three click pairs whose perceived spatial location is on the right (because of level disparity this time), followed by three on the left. To maintain roughly equal loudness with the other click types, the amplitudes of the pulses (the ratio of which was calculated to provide a given ILD expressed in dB) were engineered to fall symmetrically around the same 50% pulse amplitude of the ITD pulses. In experiment 1, and in experiment 2 that follows, the ILDs were generated by setting the levels of the signals to the two ears symmetrically around a constant mean amplitude. This procedure is not mathematically rigorous, and might have been expected to result in some variation in subjective loudness. Following the work of Keen (1972), departures from a constant loudness

would have been most significant only for large ILDs. The authors and a graduate student, therefore, expressly listened for variations in binaural loudness in pilot testing. No such variation was obvious to these listeners. Note also that since ILDs of only one magnitude were used in each trial, any such variations would have occurred between trials (and thus not have affected clicks within a trial). Our anecdotal experience is that the saltation effect is, in any case, robust over significant variations in overall stimulus level.

Stimuli of each type, direction (L→R and R→L), and ICI combination required were saved as ''snd'' resources to HYPERCARD™ stacks programed to run each experiment. The order of stimulus presentation was pseudorandomized by the program prior to each block.

3. General procedure

Subjects were seated in a sound-attenuating room in front of a Macintosh Applevision AV monitor, and keyboard with mouse. The Macintosh PowerPC 8600 CPU was located outside the room to provide a very quite environment. Sounds were delivered through Koss TD/65 headphones. The sound level in the Macintosh control panel was initially set to one deemed comfortable by the authors, and remained at this level for the duration of the studies (approximately 67 dB sound-pressure level, A-weighted; Bruel & Kjaer instruments).

Stimulus trials were subject-initiated by button click and responses recorded in the same manner. Subjects were informed at the outset that this was a study of the quality of their perception of the stimuli, and that there were no right or wrong answers. For stimulus conditions generating the perceived motion, all subjects reported a clear sensation of clicks taking an orderly intracranial trajectory across the midline plane.

4. Procedure specific to experiment 1

Before beginning this experiment, we wished to determine an ILD to use that would be in some manner ''equivalent'' to an ITD of 500 μs. We chose to use a method of cue competition to determine the ILD that would exactly cancel out the lateralization percept generated by an ITD of 500 μs [time-intensity trading, after Deatherage and Hirsh (1959) and Harris (1960)], as follows. Stimuli consisting of eight binaural click pairs at an ICI of 90 ms were created. They were either left-leading or right-leading for the entire duration of the eight clicks. Copies of these two base stimuli were created, and on them were imposed interaural level differences favoring right or left, in steps of 1 dB ranging from 6 to 13 dB. This resulted in 16 stimuli. Three listeners (the authors plus a graduate student) listened to four repetitions of each stimulus in random order and responded ''left'' or ''right'' according to the perceived lateralization of the click train. Proportion of responses not matching the actual ITD of the stimulus was plotted as a function of ILD. For all three Ss, the resulting curves rose monotonically from a probability of 0 at 6 dB to a probability of 1 at 13 dB, and crossed 0.5 at ILDs of 9.2, 9.4, and 9.7, respectively. On the basis of this result, an ILD of 9 dB was chosen. The resulting

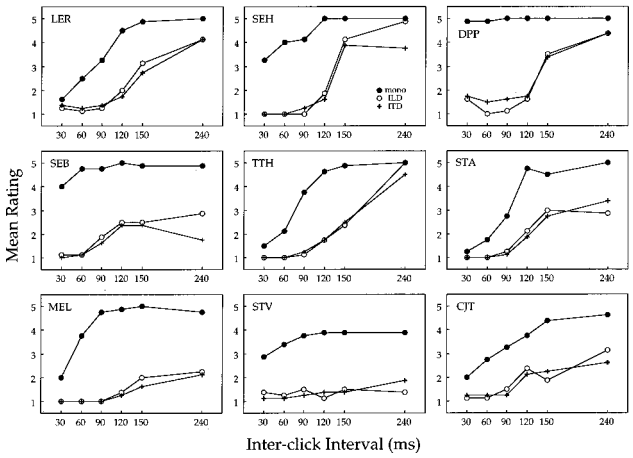


FIG. 2. Data from experiment 1. Each panel shows data for one subject, identified by a three-letter code. Mean rating of the continuity of perceived motion is plotted as a function of interclick interval. The parameter is condition (mono, ILD, ITD) as specified in the legend in the top middle panel.

''trading ratio'' was 55.6 μs/dB, which is in the range reported previously for click stimuli by Deatherage and Hirsh (1959) and Harris (1960).

Click trains consisting of mono, ITD (500-μs), and ILD (9-dB) types with ICIs of 30, 60, 90, 120, 150, and 240 ms were studied. A total of 144 trials in random order was performed, in one session, for each subject (6 ICIs×2 directions ×3 types×4 repetitions). Trial order was randomized across all conditions and repetitions. Subjects were allowed to listen to a stimulus as many times as they wished prior to responding. Subjects responded with a number from 1 to 5. A response of ''1'' indicated that the perceptual experience was of a smooth spatial progression of clicks from the point of origin to the end point, with clicks evenly spaced throughout their excursion. A response of ''5'' indicated that the first three clicks were clearly localized at the spatial location of origin, and the last three at the spatial end point, with none of the clicks located along the trajectory between the two locations. Intermediate ratings indicated imperfectly smooth progression such that clicks did not seem to be evenly spaced along their trajectory, and yet neither were they anchored perfectly at their end points. For example, the following experience might be rated a ''3:'' the first three clicks march one after another towards the midline, with the third seeming perhaps halfway to the midline, while the last three clicks resume their progression in the analogous location on the other side—there is a clear spatial gap between the first and last three, even though all six progressed spatially to some extent. In practice, this judgment was for all listeners a very natural and easily performed one, and all but one subject used the entire range of ratings in their responses.

C. Result and discussion

Data collected were mean ratings of the continuity of the perceived motion generated by each stimulus condition (6 ICIs for each of the monaural, ITD, and ILD stimulus configurations). These data are shown, separately for each of the nine listeners, in Fig. 2. The general features of the data were as follows. First, for every listener, the strength of the motion

illusion was greater for the dichotic stimuli than for the monaural ones, and this was true for all ICIs. Five of the nine listeners showed some saltation effects (ratings less than 5) for monaural stimuli, but only at the shortest ICIs. At the longer ICIs, the monaural click trains were usually assigned ratings close to 5.0, indicating zero saltation and a single discontinuity in the perceived locus of the clicks as the laterality of the stimulus train switched at its midpoint. These data are comparable to those previously presented by Hari (1995) and Shore *et al.* (1998).

Second, in all listeners, the continuity of the perceived motion for the dichotic click trains was greatest for those stimuli with the three shortest ICIs. Indeed, mean ratings were close to 1.0 (i.e., perfect saltation) for ICIs of 30, 60, and 90 ms. This indicates that the listeners typically used the full width of the motion rating scale. Listeners differed in exactly how much the saltation effect weakened as ICIs were lengthened. In four listeners (LER, SEH, DPP, TTH) the illusory motion almost disappeared at the longest ICIs (240 ms). In three listeners, the decline in the continuity of the perceived motion was less dramatic, and in one listener (STV, the one listener who did not use the full range of the 5-point rating scale), there was little deterioration in the saltation even at the longest ICI. Both the greater strength of the illusion at short ICIs, and the individual difference in dependence of saltation continuity on ICI, are consistent with previous descriptions of the illusion (Hari, 1995; Shore *et al.*, 1998).

The major new finding of experiment 1 lies in a comparison of the responses to the ITD and ILD stimuli. Despite the marked individual differences described above, the mean ratings of motion illusion strength for the ITD and ILD stimuli were virtually identical for almost all ICIs in each of the eight listeners. The curves for the ITD and ILD conditions are close in each listener; indeed, they cross at least once, and usually twice for every participant. These data indicate that, despite individual differences in patterns of sensitivity to click ICI, and despite the fact that the ITD and ILD stimuli were not carefully equated for the strengths of the image lateralization they effected, the ITD and ILD stimuli were equivalently effective in generating illusory auditory motion. Experiment 1 thus extends previous evidence by showing that the lateralization of the stimuli used to generate the illusion may be based on either ITDs or ILDs.

III. EXPERIMENT 2

A. Introduction

Experiment 1 revealed that dichotic clicks lateralized on the basis of either ILD or ITD supported a robust saltation effect, and revealed that monaural stimuli were weak or ineffective stimuli for the illusion. The dichotic clicks differ from monaural ones not only in that sound stimulation arrives at both ears, but in that the extent to which the dichotic clicks are lateralized is far less extreme than is the case for the monaural clicks. Experiment 2 was an attempt to sort out whether it was the binaurality or the extent to which the clicks were lateralized that was responsible for supporting the illusion. The general design of experiment 1 was repeated, but using monaural clicks (control) and dichotic clicks which, in different trials, were varied in ILD from 9 to 30 dB. Importantly, for all dichotic stimulus conditions, the level of the clicks at the ''quiet'' ear was above detection threshold for monaural clicks at that ear. The dichotic clicks with the smaller ILDs were readily lateralized intracranially, while those with the larger ILDs (25.6 dB, 30 dB) were more often perceived as ''in'' or ''at'' one ear, i.e., at the extreme of the laterality continuum. Thus, by using a range of ILDs wide enough to span those which were weakly lateralized (small ILDs) to those that were at the extreme of the laterality continuum (30-dB ILDs), while maintaining the binaurality of the stimuli, we were able to determine which property of the clicks best supported the saltation effect.

B. Methods

1. Subjects and apparatus

Seven normal-hearing listeners participated in experiment 2. Five of these listeners also had participated in experiment 1. The equipment used was the same as that in experiment 1.

2. Procedure specific to experiment 2

Click trains consisting of mono- and five values of ILD (9.0, 15.1, 21.2, 25.6, and 30.2 dB), with ICIs of 30, 60, 90, 120, 150, and 240 ms, were studied. The ILDs were generated by setting the levels of the signals to the two ears symmetrically around a constant mean amplitude. As mentioned above (experiment 1), this procedure is not mathematically rigorous, and might have been expected to result in some variation in subjective loudness (Keen, 1972). The authors and a graduate student therefore expressly listened for variations in binaural loudness in pilot testing. No such variation was obvious to these listeners. Because the stimulus amplitude ratios (required for a given ILD) were generated by setting the levels of the signals to the two ears symmetrically around a constant mean amplitude, in practice the decrease in amplitude applied to the ear that received the less intense signal was greater, in dB, than was the increment in amplitude applied to the ear that received the more intense signal. Fortuitously, this asymmetry grew as the ILD grew, and in form, if not in exact value, it matches the asymmetry of decreases and increases reported by Keen (1972) to be required to maintain roughly constant binaural loudness.

For what follows, we emphasize that even in the clicks with the largest ILDs, the level of the click at the ''quiet'' ear was audible when presented monaurally. A total of 144 trials in random order was performed, in one session, for each subject (6 ICIs$\times$2 directions$\times$6 types [5 ILD conditions, plus mono]$\times$2 repetitions). Subjects were allowed to listen to a stimulus as many times as they wished prior to responding. The response was a rating of 1 to 5 as described for experiment 1. Preliminary testing of two listeners (DSF, STV) revealed that ratings had not reached values near 5 for ICIs of 240 ms, and for those subjects, the experiment was run again, but including 480-ms ICI conditions.

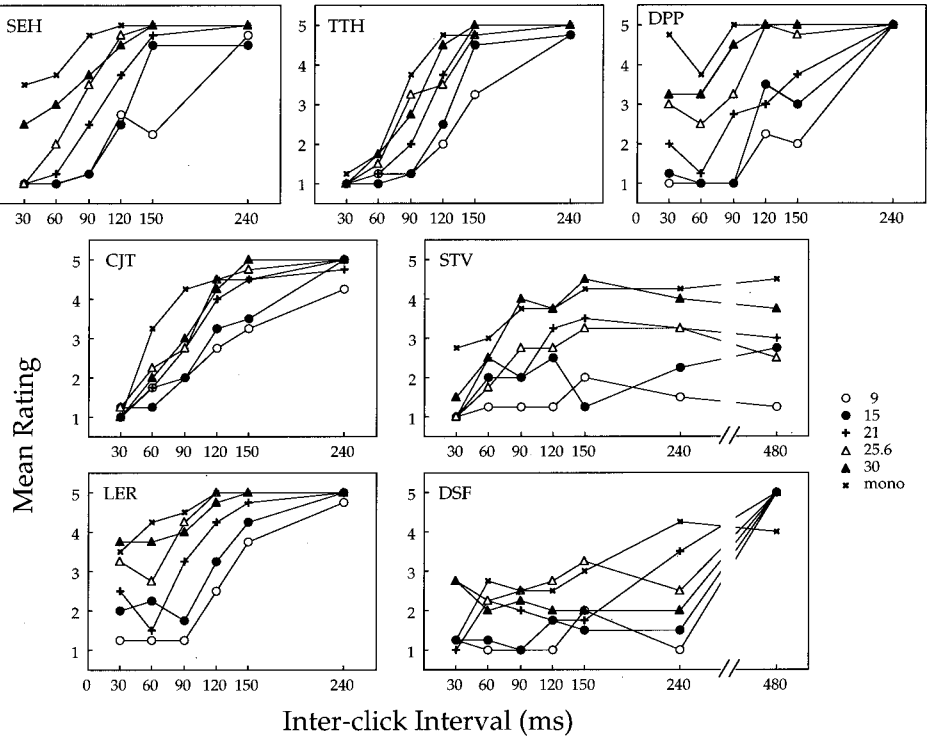


FIG. 3. Data from experiment 2. Each panel shows data for one subject, identified by a three-letter code. Mean rating of the continuity of perceived motion is plotted as a function of interclick interval. The parameter is condition (mono, and magnitude of the ILD—in dB) as specified in the legend at lower right.

C. Results and discussion

Data collected were ratings of motion continuity. Mean ratings(averaged over repetitions and directions of motion) are shown separately for the seven listeners in Fig. 3. Mean rating has been plotted as a function of ICI, with condition as the parameter. For each listener, monaural clicks produced the *weakest* saltation effect (lines joining x's). In most of the listeners, monaural clicks again elicited high ratings (weak saltation) except at the very shortest ICI. For each listener, the *strongest* saltation, i.e., the saltation supported at the longest ICIs, was seen with the 9-dB ILD stimuli (lines joining unfilled circles). Increments in ILD toward 30 dB caused an orderly change in the rating function so that the larger the ILD, the more closely the function approximated that seen with monaural stimuli, i.e., the saltation effect declined in strength as the size of the ILD increased.

Experiment 2 thus confirmed that clicks lateralized on the basis of ILDs were effective in supporting auditory saltation. Even using the largest ILD (30 dB), the stimulus at the quiet ear had a level which was audible when presented alone, so the fact that the stimuli were binaural was not sufficient to optimize auditory saltation. Importantly, however, the greater the ILD, the closer the rating functions approximated those seen with monaural stimuli. This means that it may be the extent to which the clicks are lateralized, rather than whether they are binaural or monaural, which determines the strength of the saltation effect. This finding is of interest for at least two reasons. First, an ILD of about 30 dB is perhaps towards the maximum that might be generated by the interaction of click stimuli and the human head [after Middlebrooks' (1992) measurements using narrow-band noises]. This suggests that dichotic clicks supporting percepts at the extreme of the laterality continuum are not as effective in supporting the saltation effect as are clicks with smaller ILDs. This is not to suggest that dichotic stimuli with very large disparities do not support discriminable lateralized percepts, because there is evidence that they can (e.g., Bernstein and Trahiotis, 1985; Domnitz and Colburn, 1977; Mossop and Culling, 1998). A complexity in this general argument is that naturally occurring ILDs are typically associated with ITDs favoring the same ear. In this sense, the present experiment actually used ''unnatural'' combination of ILD and ITD (after Gaik, 1993) because the ILDs were not accompanied by any nonzero ITD. Had the clicks been subjected to ILDs in combination with ITDs, the subjective laterality of the stimuli would have been enhanced, so that the resulting range of most effective ''ILDs'' might have been narrower than that seen here.

There is a second reason for interest in the outcome of experiment 2. The declining strength of the saltation effect with increasing click ILD expressed itself as a shift in the most effective ICIs towards lower values (faster rates). Restated, this means that with increases to the perceived eccentricity range spanned by the leading and trailing clicks, there was some tendency for higher click rates to be required to maintain the illusory stimulus movement. This suggests that there may be constraints on the perceived angular ''velocity'' of the illusory motion. The present data on this point contained too much variance to delineate the orderliness of this effect, but it may be worthy of future examination.

IV. EXPERIMENT 3

A. Introduction

Sensory saltation is an expression of the influence on the percepts evoked by early members of a stimulus sequence by the percepts aroused by later elements in the same sequence. That is, a perceptual consequence of late members of the

stimulus train is a modification of the percepts that would otherwise have been aroused by early members of the stimulus sequence. This raises the question of how far back in time the ''integration'' or ''perceptual writing'' process extends.

This is a nontrivial question. In the first place, the temporal window over which this perceptual writing occurs likely represents some kind of processing time or integration period for temporally resolved sensory information to form a percept. Disturbances to the width of this window might be directly relevant to the genesis of acoustically- or phonologically based language disorders (after Hari and Kiesilä, 1996). That is, if this temporal window were too wide, then it is conceivable that the percepts emerging from it would be prone to excessive distortion by processing within it (e.g., auditory saltation). Accordingly, knowledge of the width of this temporal window in normal listeners becomes an important benchmark for future studies of impaired ones. In this regard, the saltation effect serves as a potent reminder that perception is not an on-line, event-by-event, veridical mental elaboration of the stimulus. Rather, it is a time-delayed, highly processed representation of the physical events (e.g., the ''multiple drafts'' model: Dennett, 1991). The auditory saltation paradigm provides a means for assessing the width of the time window over which the processing of sensory information (i.e., the writing of the percept) operates. The purpose of experiment 3 was to provide an empirical estimate of the width of this window.

Based on the results of experiments 1 and 2, we reasoned that the last three clicks of a train were sufficient to alter the perception of the clicks that immediately preceded them. In turn, this meant that we could vary the number of clicks in the train preceding the laterality-reversed, terminal three clicks and ask listeners to report on the existence of any *stationarity* of the leading clicks in the train. We expected that the longer the leading sequence of clicks (i.e., the longer the sequence of clicks preceding the laterality reversal), the greater would be the likelihood of listeners reporting a period of locus stationarity at the beginning of the click train. By studying this phenomenon for a wide range of click ICIs (and thus a wide range of click rate and click number conjunctions), we were able to derive an estimate of the approximate duration of the processing window shaping the auditory saltation effect.

B. Methods

1. Subjects and apparatus

The nine listeners who participated in experiment 1 participated in experiment 3. The equipment and general procedures were the same as those for experiment 1.

2. Procedures specific to experiment 3

Click trains with the following values of ICI were studied: 30, 45, 60, 75, 90, 105, 120, 150. Click lateralization was achieved by imposing a 500-μs ITD on the dichotic signals. In addition, the number of leading clicks (i.e., the number of clicks preceding the laterality reversal) ranged from three (as in the other two experiments) to ten. A total of 512 trials in random order was performed, in one session, for each subject (8 ICIs $\times$2 directions$\times$8 types$\times$4 repetitions). Subjects were allowed to listen to a stimulus only once prior to responding. The response was either ''1'' or ''2'': A response of ''1'' indicated that the perceived progression through space of the component clicks in the stimulus train began immediately. This rating of 1 is not equivalent to the rating of 1 defined in experiments 1 and 2, because the smoothness of the motion (or the extent to which the component clicks were evenly spaced) is not an issue, merely that the perceived motion, whatever its quality, began right away. A rating of 2 indicated that ''stationarity'' was detected. Subjects were to respond with a 2 if two or more clicks at the beginning of the train were perceived to have been located at the point of origin. Again, the quality of the motion, or evenness of the spatial progression of the component later in the train, was not relevant to the response; indeed, a frequent experience with some of the higher number of leading click conditions was that several clicks would be perceived clearly anchored at the origin, followed by a perfectly smooth experience of illusory motion in the latter portion of the train.

C. Results and discussion

Data collected were dichotomous reports of the detection of spatial stationarity at the beginning of the click train. These scores were averaged over repetitions and directions of perceived motion. For the nine subjects who participated in this experiment, these mean data are plotted as a function of ICI in Fig. 4. In each panel of Fig. 4, the parameter is the number of clicks in the train prior to the reversal of click laterality. The spreads of rating functions showed great individual differences. However, a careful scrutiny reveals that high ratings (i.e., detections of stationarity) were most common for click trains with longer ICIs and larger prereversal click numbers.

To obtain a measurement of the duration of the perceptual writing process, the following analysis was employed. The data in Fig. 4 were replotted as a function of the duration of the leading click train (defined as the interval between the onset of the first click of the leading train and the onset of the trailing click triplet), less one click. We reasoned that any report of stationarity must have been based on no less than one ICI. This means that the window duration of the perceptual integration or writing process could, on any given trial, have been as long as the interval between the second click of the leading click train and the first click of the trailing click triplet. Because the ICIs were always constant within a trial, this is equivalent to the duration of the leading click train less one click. The leading click train lengths over which ratings intersect a midrange score (i.e., ''1.5'') provide the estimates of the duration of the temporal window over which the perceptual integration process operated in this study.

The replotted data are shown in Fig. 5, separately for each listener. Again, the parameter is the number of clicks in the train preceding the stimulus disparity reversal. For all listeners, the family of rating functions had roughly a sigmoidal disposition, with ratings of 1 (no stationarity) being reported when the duration of the leading click train was short, and ratings of 2 (detected stationarity) being reported

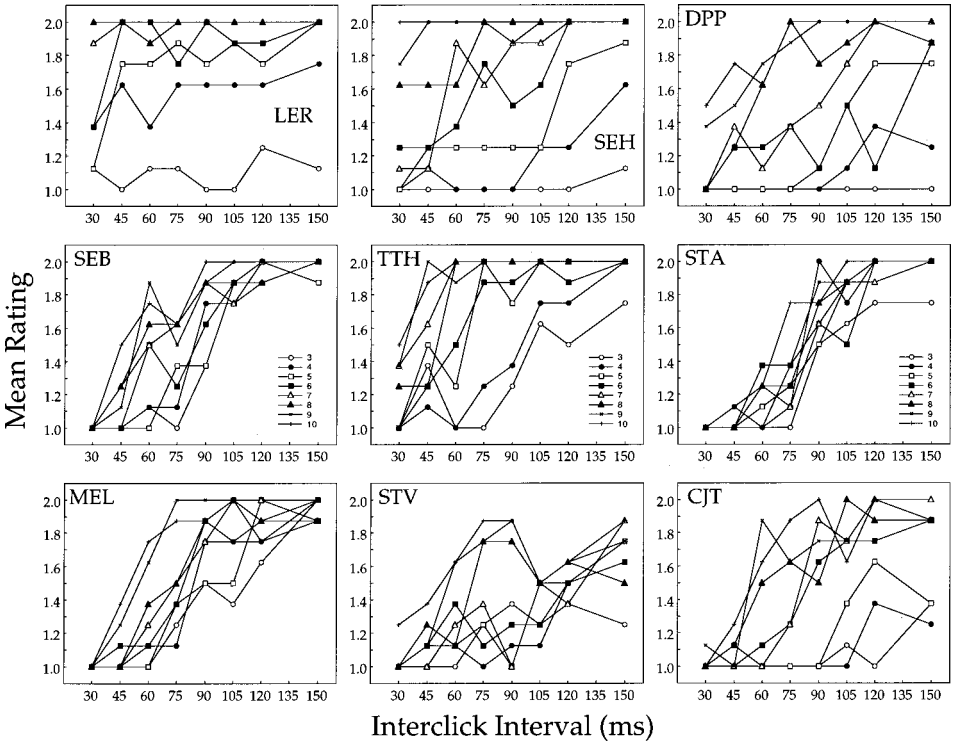


FIG. 4. Data from experiment 3. Each panel shows data for one subject, identified by a three-letter code. Mean rating for the detection of stationarity is plotted as a function of interclick interval. The parameter is the number of clicks preceding the laterality reversal (legends in middle row of panels).

when the leading click train was long. There was considerable individual variation in the range of train durations over which ratings climbed (from 1 towards 2, and in the width of that range. Note that for most listeners, the family of curves in the transformed data is more tightly grouped than that in the raw data (Fig. 4). This greater compactness suggests that responses were dominated by the length of the leading click train. There was, however, at least one exception to this generality. Subject STA's rating functions (middle right in Figs. 4 and 5) were more compact in the raw data than in the transformed set. This suggests that STA's responses were more influenced by click ICI (or rate) than by the duration of the leading train. The data for subjects SEB and MEL (left middle and lower) were, in this regard, somewhat between those for STA and the rest of the listeners, because the families of curves are only slightly more compact in the transformed data than in the raw data. These individual differences are important, because they suggest that some subjects demonstrated responses that appeared to be influenced more by click rate than by train duration.

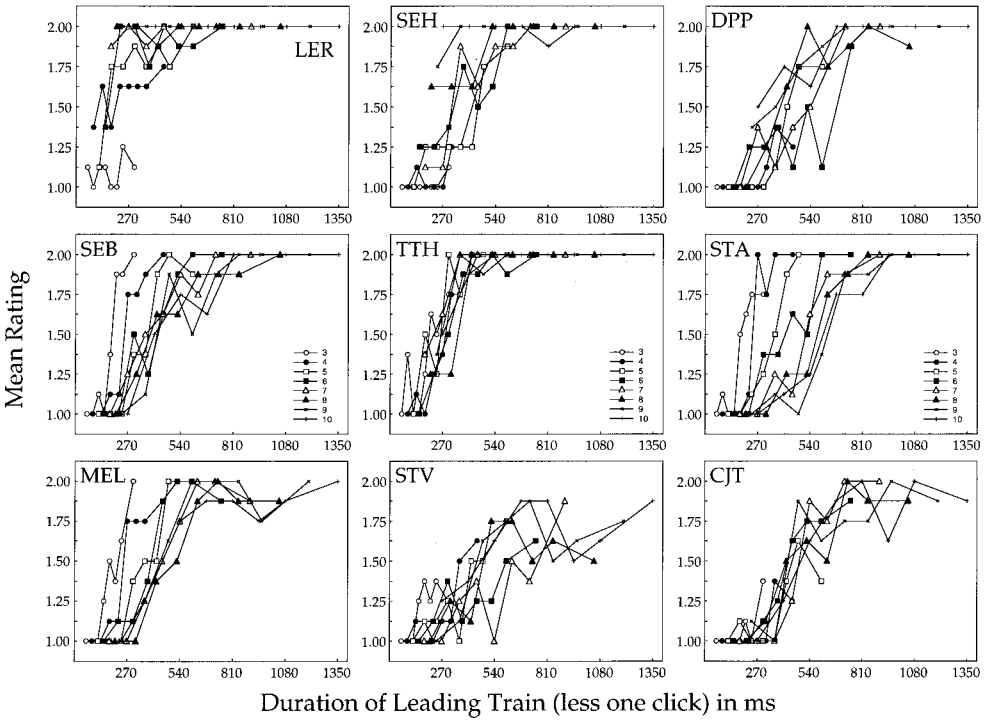


FIG. 5. Data from experiment 3. Each panel shows data for one subject, identified by a three-letter code. Mean rating for the detection of stationarity is plotted as a function of the *duration* of the click train, minus one click (in ms), preceding the laterality reversal. The parameter is the *number* of clicks preceding the laterality reversal (legends in the middle row of panels).

Parenthetically, a similar account might be made for individual differences in the data of experiments 1 and 2, i.e., subjects' responses might have been influenced not only by the perceived continuity of motion, but by, for example, click rate. It is, however, the fact that the other stimulus parameters had greater effects on the responses that gives us confidence in our design. Thus, in experiment 1, if click rate had been the dominant factor driving the ratings, then the separation between the data for monaural and dichotic click trains would not have been so evident. In experiment 3, while it is clear that individual subjects differed in the extent to which their responses were determined by click rate versus train length, it remains the case that for most of them, the curves in the transformed data were more closely grouped than those in the untransformed data. It is this closer grouping that constitutes evidence for some involvement of a temporal window of limited width.

It is an arbitrary choice as to how to derive a measurement of temporal window width from these data, particularly because we make no assumptions about the shape of that window. One was to estimate the window width is simply to read, by interpolation, the leading train duration associated with a mean rating of 1.5 in the center of each family of curves. This analysis provides estimates as short as 250 ms (listener LER) to as long as about 500 ms (listener STV). For the six listeners with the most compact families of rating curves (LER, SEH, SEB, TTH, MEL, CJT), the estimated duration of the perceptual integration or writing window was less than about 350–400 ms.

V. GENERAL DISCUSSION

The purpose of this study was to explore two spatial properties and one temporal property of auditory saltation as studied dichotically. Experiment 1 revealed that dichotic click stimuli produced robust saltation effects, whether the perceived laterality of the clicks was produced by the imposition of an ITD or an ILD. Experiment 2 provided evidence that for dichotic clicks lateralized on the basis of ILDs, the saltation effect decreased in strength with increases in the size of the ILD, and therefore with increases in the extent to which the clicks were lateralized. Experiment 3 provided a preliminary estimate of the width of the temporal window within which the perceptual integration or writing process operates, and revealed it to be something less than about a third of a second.

The fact that saltation was supported by clicks lateralized on the basis of either ILDs or ITDs (experiment 1) suggests that the illusion operates on a spatial representation regardless of the genesis of that representation in interaural time or interaural level processing. This is compatible with observations from free-field studies of auditory saltation, which revealed that saltation for clicks issued from horizontal-plane sources centered at extreme azimuths in a single hemifield (i.e., opposite one ear)—where spatial location processing is probably based as much on monaural spectral processing as on interaural level or time differences (e.g., Slattery and Middlebrooks, 1994)—is just as vivid and robust as it is for locations centered on the auditory midline (Phillips *et al.*, 2001).

In this regard, note that in the present dichotic experiments, and in our preliminary free-field observations (Phillips *et al.*, 2001), the auditory saltation illusion traverses the midline. This property distinguishes the auditory effect from its analog in the visual and somatosensory modalities where the saltation effect does not cross the midline (Geldard, 1976; Geldard and Sherrick, 1986). The reasons for this discrepancy are unclear. One possibility is that the forebrain auditory system, unlike the visual and somatosensory ones, contains large populations of neurons with spatial receptive fields spanning both left and right sensory hemifields (''omnidirectional'' cells, after Middlebrooks and Pettigrew, 1981; see Phillips and Brugge, 1985, for review). It is thus conceivable that individual neurons with spatial receptive fields spanning the motion end points are in some way required to support the illusion (see also Geldard and Sherrick, 1986).

Auditory saltation is an illusion of sound source motion in which the perception of the late elements in the stimulus train influences the perception of the earlier elements in the train. The behavioral expression of this effect is the percept of successive clicks in the train being spatially relocated in a path between the actually stimuluated locations. Saltation is not a continuous motion of the kind stimulated by continuous variations in an ongoing stimulus (e.g., binaural beats, or genuine source motion in the free field), but neither is the source of sound itself continuous in time. It is for this reason that the perceived motion has a step-like quality: it is as if a single auditory object undertakes a continuous motion between two locations and emits a click periodically as it does so. Saltation is also distinguished from ''phi motion'' in vision by two features. First, saltation requires repetitive stimulation of at least one location, while phi motion does not. Second, the saltation motion percept has a step-like quality, while the motion percept prompted by phi motion stimuli is more continuous (Geldard and Sherrick, 1986).

The purpose of experiment 3 was to provide an estimate of the temporal window width within which the perceptual process operates. The rating functions for each listener showed some dispersion across leading train length (Fig. 5), which suggests that the temporal ''border'' between clicks affected and unaffected by the perceptual writing process may be somewhat gradual or blurred, at least over the time scale investigated here. Given that, more quantitative approaches to the measurement, as might be achieved by curve fitting, may be misplaced because the precision of the measurement may impute and inappropriate temporal precision to the window width. It is, in any case, unlikely that such measurements would be outside the range suggested by visual inspection of the data presented here, i.e., a window width in the range of a third of a second or less in most of our listeners.

The present estimate of the temporal window width has a striking number of counterparts in other measures of central auditory processing ''times.'' Considering first tasks that can be executed monaurally, recent attempts to measure the temporal window of auditory integration using electrophysiological measures have provided estimates on the order of 170 ms (Yabe *et al.*, 1998) to about 200 ms (Nagarajan *et al.*, 1999). In the speech perception domain, the extent to which

spectral slices of speech sounds can be temporally offset without significant impairment of discrimination performance is on the order of 140 ms, and 50%-correct word recognition is seen with asynchronies of more than 200 ms (Arai and Greenberg, 1998). In this regard, there is evidence to suggest that the functional unit of speech perception is not the phoneme but the syllable, whose mean duration is again on the order of 200 ms (Greenberg, 1999). In the realm of binaural hearing, depending on the task, and on the listener, the temporal window for the detection of a change in binaural stimuli (''binaural sluggishness'') can be of the same order, i.e., as low as 50 ms and as high as 300 ms (Akeroyd and Summerfield, 1999). It does not follow from these apparent correspondences that a single temporal mechanism is being tapped by the experiments. It does, however, suggest that there is a window, on the order of 200–300-ms wide, within which there is a great deal of perceptual processing executed, and to which the listener probably has little direct access.

The present experiments were not specifically designed to answer questions about mechanisms underlying auditory saltation, but is noteworthy that the design of effective stimuli (repetitive, binaural transients that are easily lateralized) might render perceptual responses to them susceptible to ''binaural adaptation'' (after Buell and Hafter, 1988; Hafter and Buell, 1990; see also Hafter and Dye, 1983). By this term, they mean that under conditions of repetitive stimulation, the usefulness of the interaural information extracted from the elements of the train declines, until the system is ''reset'' by the introduction of some form of a change in the stimulus train. The description of the general stimulus conditions giving rise to binaural adaptation, and recovery from it, is not far removed from that of the conditions for effective auditory saltation. The main difference is that the ICIs used in Hafter *et al.*'s studies were usually shorter than those used in the present study (less than about 30 ms). With that caveat, the work on binaural adaptation raises the possibility that the mechanisms mediating auditory saltation might take place at a lower level of processing than has previously been assumed.

ACKNOWLEDGMENTS

This work was supported by grants from NSERC of Canada to D.P.P. We thank Susan Boehnke and Leanna Rutherford for their participation, and for many helpful discussions of this work. Thanks are also due to Drs. L. Bernstein and W. Grantham, and to an anonymous reviewer, for helpful comments on an earlier version of this article.

Akeroyd, M. A., and Summerfield, A. Q. (**1999**). ''A binaural analog of gap detection,'' J. Acoust. Soc. Am. **105**, 2807–2820.

Arai, T., and Greenberg, S. (**1998**). ''Speech intelligibility in the presence of cross-channel spectral asynchrony,'' Proc. IEEE Conf. Acoust. Speech Signal Proc., Seattle, 933–936.

Bernstein, L. R., and Trahiotis, C. (**1985**). ''Lateralization of low-frequency, complex waveforms: The use of envelope-based temporal disparities,'' J. Acoust. Soc. Am. **77**, 1868–1880.

Bremer, C. D., Pittenger, J. B., Warren, R., and Jenkins, J. J. (**1977**). ''An illusion of auditory saltation similar to the cutaneous rabbit,'' Am. J. Psychol. **90**, 645–654.

Buell, T. N., and Hafter, E. R. (**1988**). ''Discrimination of interaural differences of time in the envelopes of high-frequency signals: Integration times,'' J. Acoust. Soc. Am. **84**, 2063–2066.

Deatherage, B. H., and Hirsh, I. J. (**1959**). ''Auditory localization of clicks,'' J. Acoust. Soc. Am. **31**, 486–492.

Dennett, D. C. (**1991**). *Consciousness Explained* (Little, Brown, Boston).

Domnitz, R. H., and Colburn, H. S. (**1977**). ''Lateral position and interaural discrimination,'' J. Acoust. Soc. Am. **61**, 1586–1598.

Farmer, M. E., and Klein, R. M. (**1995**). ''The evidence for a temporal processing deficit linked to dyslexia: A review,'' Psychonomic Bull. Rev. **2**, 460–493.

Gaik, W. (**1993**). ''Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling,'' J. Acoust. Soc. Am. **94**, 98–110.

Geldard, F. A. (**1975**). *Sensory Saltation: Metastability in the Perceptual World* (Erlbaum, Hillsdale, NJ).

Geldard, F. A. (**1976**). ''The saltatory effect in vision,'' Sens Processes **1**, 77–86.

Geldard, F. A., and Sherrick, C. E. (**1972**). ''The cutaneous rabbit: A perceptual illusion,'' Science **178**, 178–179.

Geldard, F. A., and Sherrick, C. E. (**1986**). ''Space, time and touch,'' Sci. Am. **255**, 90–95.

Greenberg, S. (**1999**). ''Speaking in shorthand—A syllable-centric perspective for understanding pronunciation variation,'' Speech Commun. **29**, 159–176.

Hafter, E. R., and Buell, T. N. (**1990**). ''Restarting the adapted binaural system,'' J. Acoust. Soc. Am. **88**, 806–812.

Hafter, E. R., and Dye, Jr., R. H. (**1983**). ''Detection of interaural differences of time in trains of high-frequency clicks as a function of interclick interval and number,'' J. Acoust. Soc. Am. **73**, 644–651.

Hari, R. (**1995**). ''Illusory directional hearing in humans,'' Neurosci. Lett. **189**, 29–30.

Hari, R., and Kiesilä, P. (**1996**). ''Deficit of temporal auditory processing in dyslexic adults,'' Neurosci. Lett. **205**, 138–140.

Harris, G. G. (**1960**). ''Binaural interactions of impulsive stimuli and pure tones,'' J. Acoust. Soc. Am. **32**, 685–692.

Keen, K. (**1972**). ''Preservation of constant loudness with interaural amplitude asymmetry,'' J. Acoust. Soc. Am. **52**, 1193–1196.

Lockhead, G. R., Johnson, R. C., and Gold, F. M. (**1980**). ''Saltation through the blind spot,'' Percept. Psychophys. **27**, 545–549.

Middlebrooks, J. C. (**1992**). ''Narrow-band sound localization related to external ear acoustics,'' J. Acoust. Soc. Am. **92**, 2607–2674.

Middlebrooks, J. C., and Pettigrew, J. D. (**1981**). ''Functional classes of neurons in primary auditory cortex of the cat distinguished by sensitivity to sound location,'' J. Neurosci. **1**, 107–120.

Mossop, J. E., and Culling, J. F. (**1998**). ''Lateralization of large interaural delays,'' J. Acoust. Soc. Am. **104**, 1574–1579.

Nagarajan, S., Mahncke, H., Salz, T. , Tallal, P., Roberts, T., and Merzenich, M. M. (**1999**). ''Cortical auditory signal processing in poor readers,'' Proc. Natl. Acad. Sci. U.S.A. **96**, 6483–6488.

Phillips, D. P. (**1999**). ''Auditory gap detection, perceptual channels, and temporal resolution in speech perception,'' J. Am. Acad. Audiol. **10**, 343–354.

Phillips, D. P., and Brugge, J. F. (**1985**). ''Progress in neurophysiology of sound localization,'' Annu. Rev. Psychol. **36**, 245–274.

Phillips, D. P., Hall, S. E., and Rutherford, L. E. D. (**2001**). ''An illusion of auditory motion in the free field, and its relation to sound localization mechanisms,'' Assoc. Res. Otolaryngol. Abs. **24**, 260–261.

Shore, D. I., Hall, S. E., and Klein, R. M. (**1998**). ''Auditory saltation: A new measure for an old illusion,'' J. Acoust. Soc. Am. **103**, 3730–3733.

Slattery, W. H., and Middlebrooks, J. C. (**1994**). ''Monaural sound localization: Acute versus chronic unilateral impairment,'' Hear. Res. **75**, 38–46.

Tallal, P., Miller, S., and Fitch, R. H. (**1993**). ''Neurobiological basis of speech: A case for the preeminence of temporal processing,'' Ann. N.Y. Acad. Sci. **682**, 27–47.

Thurlow, W. R., and Oneson, R. E. (**1984**). ''On the trail of an auditory rabbit,'' Bull. Psychon. Soc. **22**, 538–540.

Yabe, H., Tervaniemi, M., Sinkkonen, J., Huotilainen, M., Ilmoniemi, R. J., and Näätänen, R. (**1998**). ''Temporal window of integration of auditory information in the human brain,'' Psychophysiology **35**, 615–619.

Estimation of viscoelastic shear properties of vocal-fold tissues based on time–temperature superposition[a)]

Roger W. Chan[b)]
Vocal Fold Physiology and Biomechanics Laboratory, Department of Audiology and Speech Sciences, Purdue University, 1353 Heavilon Hall, West Lafayette, Indiana 47907-1353

(Received 16 December 2000; revised 9 April 2001; accepted 29 May 2001)

Empirical data on the viscoelastic shear properties of human vocal-fold mucosa (cover) were recently reported at relatively low frequency (0.01–15 Hz). For the data to become relevant to voice production, attempts have been made to parametrize and extrapolate the data to higher frequencies using constitutive modeling [Chan and Titze, J. Acoust. Soc. Am. **107**, 565–580 (2000)]. This study investigated the feasibility of an alternative approach for data extrapolation, namely the principle of time–temperature superposition (TTS). TTS is a hybrid theoretical–empirical approach widely used by rheologists to estimate the viscoelastic properties of polymeric systems at time or frequency scales not readily accessible experimentally. It is based on the observation that for many polymers, the molecular configurational changes that occur in a given time scale at a low temperature correspond to those that occur in a shorter time scale at a higher temperature. Using a rotational rheometer, the elastic shear modulus (G') and viscous shear modulus (G'') of vocal-fold cover (superficial layer of lamina propria) tissue samples were measured at 0.01–15 Hz at relatively low temperatures (5°–37 °C). Data were empirically shifted according to TTS, yielding composite ''master curves'' for predicting the magnitude of the shear moduli at higher frequencies at 37 °C. Results showed that TTS may be a feasible approach for estimating the viscoelastic shear properties of vocal-fold tissues at frequencies of phonation (on the order of 100–1000 Hz). © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1387094]

PACS numbers: 43.70.Aj, 43.70.Bk, 43.35.Mr [AL]

I. INTRODUCTION

Biomechanical properties of vocal-fold tissues are critical information for the theoretical and clinical studies of voice production. For example, data on viscoelastic shear properties of the vocal-fold mucosa are important for (1) computer modeling of vocal fold vibration (Alipour *et al.*, 2000); (2) theoretical estimations of phonation threshold pressure, an important objective measure of vocal effort and the energy required for sustaining phonation (Titze, 1992); and (3) predictions of clinical outcomes of vocal-fold surgery using implantable materials of different biomechanical properties (Chan and Titze, 1999a, 1999b). We have previously measured the linear viscoelastic shear properties of the human vocal-fold cover (epithelium and the superficial layer of lamina propria) as well as surgical biomaterials such as fat and hyaluronic acid, obtaining data on elastic shear modulus (G' or μ), viscous shear modulus (G'' or $\omega\eta$), dynamic viscosity (η' or η), and damping ratio (ζ) (Chan and Titze, 1999a, 1999b). Because the dynamic shear data were empirically obtained at a relatively low frequency ($\leqslant$15 Hz), we had also attempted to extrapolate the data to higher frequencies typical of phonation (usually $>$100 Hz) using two constitutive models, namely a quasilinear viscoelastic model and a statistical network model (Chan and Titze, 2000). Results of the constitutive modeling showed that the theoretical predictions matched the empirical data reasonably well, allowing for parametric descriptions of the data and their extrapolations to higher frequencies. However, it was not entirely clear from the results of that study which of the two constitutive models used would be more appropriate for the parametric descriptions of tissue shear properties of the vocal-fold cover (Chan and Titze, 2000). Certainly, further empirical measurement at higher frequencies is the ideal way for obtaining valid data and for validating the models, but technological limitations on most existing rheometers temporarily preclude us from obtaining valid empirical data at high-enough frequency (i.e., frequency $>$15 Hz, preferably $>$100 Hz).

This paper attempted to investigate the feasibility or applicability of time–temperature superposition (TTS) as an alternative approach for theory-based data extrapolation and prediction of vocal-fold tissue shear properties at higher frequencies. The principle of TTS, discussed exhaustively by Ferry (1980), has been widely used in the field of rheology to expand the effective time or frequency scale of linear and nonlinear viscoelastic measurements. Many rheologists routinely apply TTS to estimate the viscoelastic properties of polymer melts and polymer solutions at time or frequency scales not readily accessible experimentally. TTS has also been known as the method of ''reduced variables'' or the method of viscoelastic corresponding states. The basic notion of TTS is that the viscoelastic behavior of many polymeric systems is a function of two principal variables: time (or frequency) and temperature (Bird *et al.*, 1987; Ferry, 1980). The effect of change in temperature on the viscoelastic behavior of a polymeric system is often equivalent to the effect

[a)]A previous version of this paper was presented at the 138th meeting of the Acoustical Society of America, Columbus, Ohio, 1–5 November 1999.
[b)]Electronic mail: rchan@purdue.edu

of change in the time scale of deformation, such that the two variables time and temperature can be used interchangeably, or effectively ''reduced'' into one (hence the name reduced variables). Specifically, it has been observed that for many polymeric systems the molecular configurational changes that occur in a given time scale at a low temperature correspond to those that occur in a shorter time scale at a higher temperature (Bird *et al.*, 1987; Ferry, 1980), i.e., the viscoelastic functions measured at a given frequency at a low temperature would be equivalent to those measured at a higher frequency at a higher temperature. These viscoelastic effects have been observed for a variety of homogeneous polymers in nonpolar solvents at temperatures above the glass transition temperature (i.e., in the transition zone, where there is no change in the thermodynamic state of the polymer) (Ferry, 1980). For many biological tissues, including vocal-fold tissues, an analogous range of temperature may be located at 5°–37 °C, in which no structural changes of tissues have been reported in the literature (e.g., no denaturing of proteins) and there is certainly no phase change of the aqueous components of the tissues (no freezing or formation of ice crystals) (Chan, 1998; Peters *et al.*, 1997). The principle of TTS is briefly summarized next, including an outline of the procedure used to shift the empirical curves of various viscoelastic functions.

II. THE PRINCIPLE OF TIME–TEMPERATURE SUPERPOSITION

Historically, TTS was first established as a principle for empirically shifting the curves of viscoelastic functions obtained at different temperatures along a time or frequency axis, prior to the development of theories that supported its validity. It has since been found to be applicable to a variety of undiluted and diluted polymer systems whose time- and temperature-dependent viscoelastic behaviors could be described by constitutive equations from different molecular theories (e.g., bead–spring, bead–rod, and network theories). The following is a simplified overview of TTS based on linear flexible chain theories (i.e., bead–spring models, in particular the Rouse and Zimm models), in order to illustrate the basic theoretical validity and the specific empirical procedures of TTS. A more detailed illustration as well as discussions based on other molecular models can be found in Bird *et al.* (1987) and Ferry (1980).

Consider a dilute polymer solution consisting of linear molecules with no intermolecular interactions (entanglements), each of which can be modeled by N beads connected by $N-1$ Hookean springs. When the solution is at rest, there is a state of equilibrium associated with each molecule in which the orientations of chemical bond vectors form a Gaussian distribution, resulting in minimal free energy. When the solution is subjected to an external shear stress, each molecule is distorted by the flowing solvent (acting on the beads) such that orientations of the chemical bond vectors are changed and elastic energy is stored in the Hookean ''entropy springs.'' In other words, there is free-energy storage associated with an entropy decrease due to the perturbed Gaussian distribution of equilibrium. Elastic recovery is represented by Brownian reorientation of the bond vectors back to the equilibrium Gaussian distribution. The linear viscoelastic functions of the system, which depend on the interaction between the perturbation and the restoration of the equilibrium Gaussian distribution, can be explicitly expressed in terms of a series of relaxation time constants τ_i associated with the ith bead (Ferry, 1980). Regardless of the degree of hydrodynamic interaction and the specific constitutive equation used (e.g., Rouse's versus Zimm's), τ_i for the different beads have the same temperature dependence, provided that there is no phase change of the polymer within the temperature range under consideration. Hence, τ_i at different temperatures are related to one another

$$\tau_{iT} = a_T \tau_{iTo}, \tag{1}$$

where τ_{iT} is the relaxation time at temperature T (in absolute temperature), and τ_{iTo} is the relaxation time at temperature T_o, an arbitrary reference temperature. The constant a_T quantifies the temperature dependence of the system's viscoelastic behavior and is called the *shift factor*, which is a very important empirical parameter for TTS, as we shall see below. τ_i decreases with an increase in T, because there is more free energy at higher temperatures, and molecular configurational changes (Gaussian distribution changes) occur more rapidly. Hence, according to Eq. (1), $a_T<1$ if $T>T_o$, while $a_T>1$ if $T<T_o$.

The elastic shear modulus G' (identical to μ in our previous publications, i.e., Chan and Titze, 1999a, 1999b, 2000) of the linear flexible chain system can be expressed as a summation of the contributions of individual relaxation times, according to a version of the generalized Maxwell model of viscoelasticity (Ferry, 1980)

$$G' = \frac{cRT}{M} \sum_i^N \frac{\omega^2 \tau_i^2}{1+\omega^2 \tau_i^2}, \tag{2}$$

where c is molar concentration of the solute molecules, R is molar gas constant, M is the polymer molecular weight, and ω is angular frequency of oscillation. A change of temperature from T_o to T would effectively change *either* the molecular relaxation time τ_i *or* the angular frequency ω by a factor of a_T, because of the form of Eq. (2) (the product relationship between τ_i and ω)

$$\begin{aligned} &\text{Either } \tau_{iTo} \rightarrow \tau_{iT} \quad [\text{i.e., } \tau_i \rightarrow a_T \tau_i, \text{ see Eq. (1)}], \\ &\text{or } \quad \omega \rightarrow a_T \omega \quad \text{when } T_o \rightarrow T. \end{aligned} \tag{3}$$

This is virtually a merge of time and frequency into one single variable (reduced variables). With such changes in Eq. (2), it can be seen that G' measured at a frequency of ω at T is equivalent to $G' c_o T_o / cT$ measured at a frequency of $a_T \omega$ at T_o. The factor c_o/c is due to the very small change in concentration ($c_o \rightarrow c$) as a result of thermal expansion, which is usually negligible unless $T-T_o$ is very large. Likewise, any density changes due to thermal expansion are negligible for small temperature changes, which is true for biological tissues in the small temperature range involved in the present study (5°–37 °C). Recall that $a_T>1$ when $T<T_o$, i.e., G' measured at a low frequency (ω) at a low temperature (T) is equivalent to $G' T_o/T$ measured at a higher frequency ($a_T \omega$) at a higher temperature (T_o). In other words,

the empirical curves of G' obtained at a relatively low temperature T can be shifted along the y axis by a factor of T_o/T (an upward shift) and along the x axis by the shift factor a_T (a shift to the right) to predict the magnitude of G' at T_o. Exactly the same principles are applicable to shift the empirical curves of other viscoelastic functions obtained at different temperatures, most notably the viscous shear modulus G'' (Ferry, 1980) (G'' is identical to $\omega\eta$ in the previous publications by Chan and Titze). In the present investigation, both G' and G'' were empirically obtained for vocal-fold tissues and shifted according to TTS.

The shift factor a_T quantifies the effect of temperature on the system's viscoelastic behavior. It is itself a function of temperature and can be determined from an equation that has been shown to be widely applicable in many homogeneous nonaqueous polymeric systems

$$\log a_T = \frac{-c_1^o(T-T_o)}{c_2^o+T-T_o}, \tag{4}$$

where c_1^o and c_2^o are constants, to be found empirically. This is commonly called the WLF equation, as it was first introduced by Williams, Landel, and Ferry in 1955. The following procedure can be used to determine c_1^o and c_2^o (Ferry, 1980): (1) The linear viscoelastic functions G' and G'' empirically obtained at different temperatures are plotted as functions of frequency (ω) graphically; (2) Initial estimates of the shift factor a_T are made empirically by measuring the horizontal distances between adjacent curves of G' or G'', which are estimates of $\Delta \log a_T$ at different temperatures; (3) Using these tentative estimates of $\log a_T$, $-(T-T_o)/\log a_T$ are plotted against $-(T-T_o)$ to yield a straight line, based on a rearranged WLF equation

$$\frac{-(T-T_o)}{\log a_T} = \frac{-(T-T_o)}{-c_1^o} + \frac{c_2^o}{c_1^o}, \tag{5}$$

which has a slope of $-1/c_1^o$ and a y intercept of c_2^o/c_1^o. The constants c_1^o and c_2^o are given by

$$c_1^o = -1/m \quad \text{and} \quad c_2^o = -b/m, \tag{6}$$

where m is the slope and b is the y intercept of the linear regression plot.

Having found c_1^o and c_2^o, final values of the shift factor a_T can be calculated as a function of temperature using the WLF equation [Eq. (4)]. The empirical curves of G' and G'' obtained at different temperatures can then be shifted horizontally (as well as vertically) to form composite curves commonly called ''master curves'' that allow predictions of the magnitude of G' and G'' at higher frequency values at the reference temperature T_o.

It is widely accepted that the principle of TTS is a valid approach for extrapolating the viscoelastic properties of a wide variety of relatively simple polymeric systems: many homogeneous polymer melts and dilute polymer solutions in nonpolar solvents, to time or frequency scales beyond what is experimentally accessible. For more complex systems, such as polymers in polar solvents, aqueous polymer solutions, and nonhomogeneous systems including all biological tissues, the applicability of TTS may be limited because its theoretical assumptions of linear flexible chain models with no polymer entanglements are incompatible with the intermolecular (hydrophobic and hydrophilic) interactions present in those systems. In addition, the distributions of the relaxation time constants τ_i and their relationships with specific viscoelastic functions in nonhomogeneous systems can be expected to be much more complicated than that described by Eq. (2) or other similar simple viscoelastic models. Hence, there have been few reports in the literature documenting the application of TTS for these more complex polymeric systems and biological tissues in general. A careful search of the literature, however, found that two recent studies have been targeting precisely such novel and apparently incompatible applications of TTS. Labropoulos *et al.* (2000) investigated the effect of gel concentration on the rheological properties of an aqueous homogeneous polymeric system, namely agar gels which were prepared by cooling (gelation) of aqueous solutions of agar (a mixture of polysaccharides). They measured the storage modulus (elastic shear modulus G') of the gels at temperatures in the gelation range (25°–40 °C) at a frequency range of 0.016–16 Hz and were able to apply TTS to obtain master curves at a reference temperature of 30 °C over a frequency range of five decades (0.0016–160 Hz). Peters *et al.* (1997) measured the elastic shear modulus, loss angle (phase shift), and stress relaxation modulus of bovine brain tissues *in vitro* at 7°–37 °C over a frequency range of 0.16–16 Hz (over a time scale of around 0.4–400 s for the relaxation modulus). They found similar values of the shift factor a_T for both their dynamic shear data and stress relaxation data, and showed that within the temperature range studied it may be feasible to apply TTS to extrapolate their dynamic shear data to much higher frequency (up to 1.6 MHz) and their relaxation modulus data to a much shorter time (down to 0.1 msec) at 37 °C. These reports provided some of the very first evidence supporting such nontraditional applications of TTS for aqueous and nonhomogeneous polymeric systems.

III. METHOD

Viscoelastic shear properties of vocal-fold tissues at different temperatures were empirically measured using our previously reported methodology (Chan and Titze, 1999a). Briefly, vocal-fold tissue samples were subject to small-amplitude sinusoidal oscillatory shear in a rotational rheometer. A Bohlin CS-50 controlled stress rheometer with a parallel-plate (plate-on-plate) geometry was used, where a tissue sample was sandwiched between two rigid circular plates (diameter=20 mm, gap size=0.3 mm) with a stationary lower plate and a rotating upper plate. The sample was excited by a precisely controlled sinusoidal stress (torque) over a frequency range of three decades (0.01–15 Hz). At each frequency, the angular displacement (as well as angular velocity) of the upper plate resulting from the oscillation was monitored by a sensitive optical transducer as a function of time. Based on the prescribed torque and the measured angular velocity (their amplitudes and phase difference), the shear stress, shear strain, and strain rate associated with the oscillatory deformation were determined. Linear viscoelastic

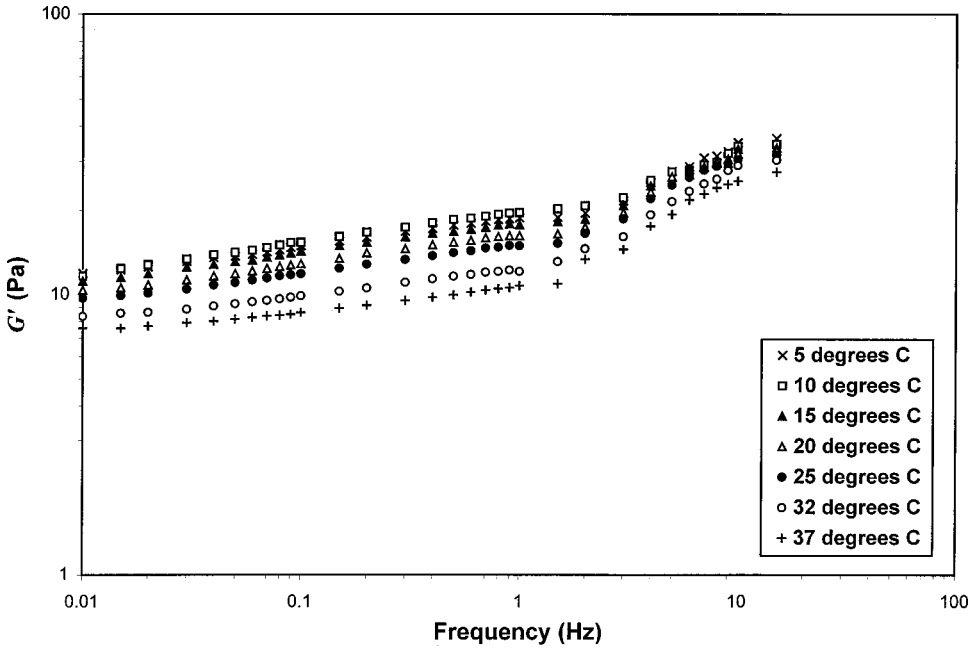


FIG. 1. Elastic shear modulus (G') of human vocal-fold cover (59-year-old male) as a function of frequency and temperature.

shear properties (G' and G'') of the sample can then be computed for each frequency.

Vocal-fold tissue samples were harvested from two adult human cadavers from autopsy, including a 59-year-old male and a 60-year-old male who had neither laryngeal pathologies nor head and neck disease (one specimen from the larynx of each subject). The specimens obtained were believed to be fresh enough (within 24 h post-mortem) to yield biomechanical data representative of the *in vivo* state, as a previous study has shown that post-mortem changes up to 24 h did not significantly affect the viscoelastic shear properties of the vocal-fold lamina propria (Chan and Titze, 2001). Tissue samples of the vocal-fold cover (epithelium and superficial layer of lamina propria) were dissected and mounted on the rheometer for rheometric measurements, as described previously (Chan and Titze, 1999a). Rheometric measurements for each sample were made at the temperatures of 5°, 10°, 15°, 20°, 25°, 32°, and 37 °C, which were maintained to an accuracy of ± 0.1 °C by a temperature control unit that circulated distilled water into the lower plate mounting of the rheometer. The reference temperature (T_o) chosen for the purpose of this study was 37 °C, because it is the body temperature or target temperature at which tissue viscoelastic properties become relevant to vocal-fold vibration and voice production. Tissue samples were allowed to equilibrate to the correct temperature for $\frac{1}{2}$ h in physiological (0.9%) saline solution before the beginning of each trial of measurements.

IV. RESULTS AND DISCUSSION

Following the procedure described in Sec. II, the shift factor a_T as a function of temperature was determined from the WLF equation for each cadaveric subject: (1) Empirical

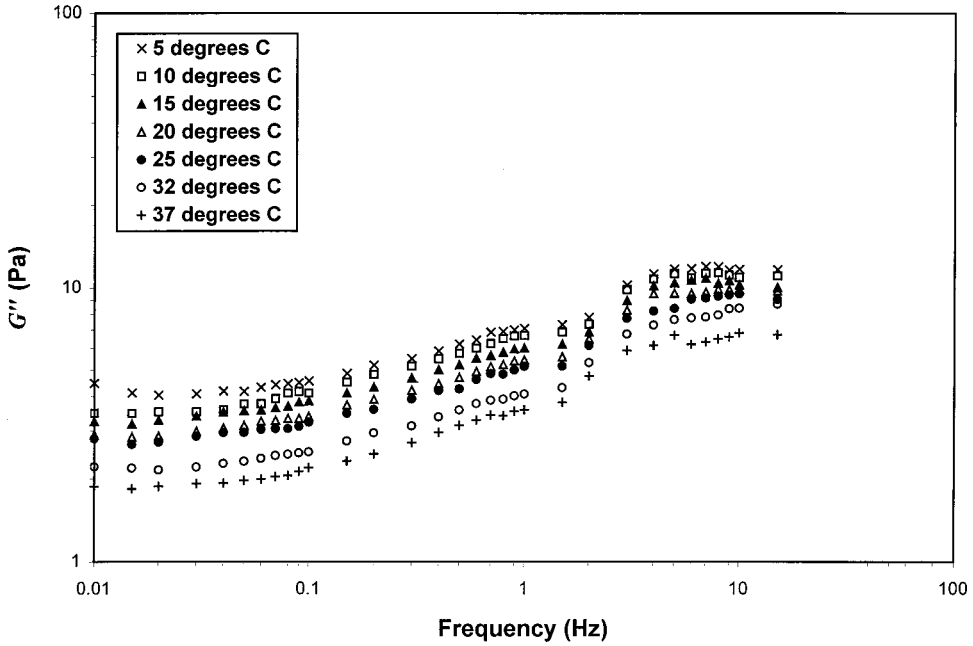


FIG. 2. Viscous shear modulus (G'') of human vocal-fold cover (59-year-old male) as a function of frequency and temperature.

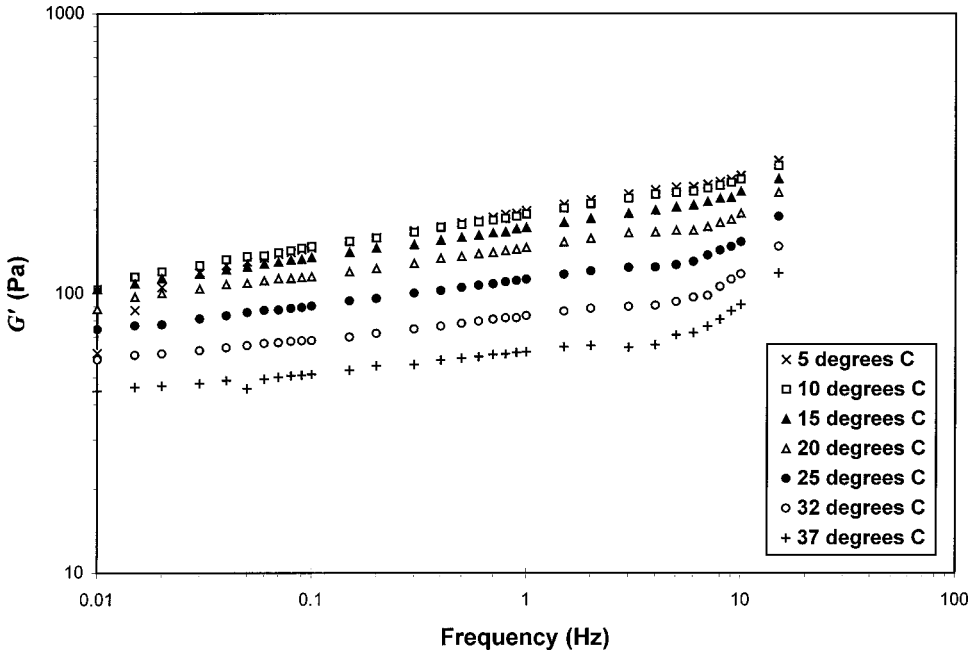


FIG. 3. Elastic shear modulus (G') of human vocal-fold cover (60-year-old male) as a function of frequency and temperature.

data of the elastic shear modulus (G') and the viscous shear modulus (G'') of the vocal-fold cover measured at different temperatures for the 59-year-old male and the 60-year-old male were plotted in Figs. 1–4. Note that there were significant inter-subject differences in the magnitude of G' and G'', which were consistent with previous findings and may reflect possible individual differences in tissue molecular constituents (Chan and Titze, 1999a). (2) Tentative values of $\log a_T$ at different temperatures were estimated by measuring the horizontal distances between adjacent G'' curves (Fig. 2 for the 59-year-old male, Fig. 4 for the 60-year-old male) (see Tables I and II, columns 3 and 4). (3) The WLF equation was rearranged [Eq. (5)] and $-(T-T_o)/\log a_T$ was plotted against $-(T-T_o)$ to yield a linear regression curve (straight line) for determining the constants c_1^o and c_2^o (Fig. 5 for the 59-year-old male, Fig. 6 for the 60-year-old male). (4) Final a_T values at different temperatures were computed using the WLF equation [Eq. (4)] once c_1^o and c_2^o were determined (Tables I and II, columns 7–8). Figures 7 and 8 show the temperature dependence of a_T for the 59-year-old male and the 60-year-old male, respectively, based on the known values of c_1^o and c_2^o in the WLF equation. Both sets of shift factors displayed a temperature dependence roughly consistent with those of many homogeneous polymeric systems in nonpolar solvents, i.e., a monotonic decrease of $\log a_T$ with temperature (Ferry, 1980). Note that the a_T values were very different for the two subjects, consistent with the significant individual differences in G' and G'' and suggesting that the relaxation behaviors of apparently similar nonhomogeneous systems could be very different, depending on the exact molecular make-up (for instance, varying concentrations of different vocal-fold constituents).

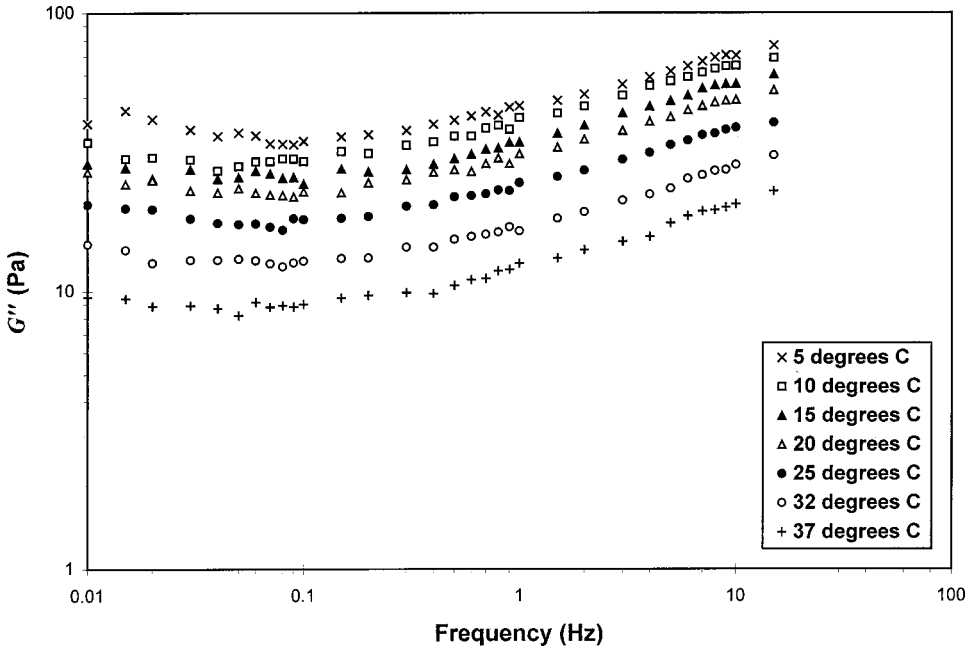


FIG. 4. Viscous shear modulus (G'') of human vocal-fold cover (60-year-old male) as a function of frequency and temperature.

TABLE I. Calculation of the shift factor a_T for the 59-year-old male based on the WLF equation [Eq. (4)]. Note: $c_1^o = -1/m = -1/0.2175 = -4.5972$, and $c_2^o = -b/m = -14.5530/0.2175 = -66.9031$ (from Fig. 5).

Temperature (°C)	Absolute temperature (K)	$\Delta \log a_T$ (from G'' data)	Tentative $\log a_T$ (cumulative)	$-(T-T_o)$ (T_o =310 K)	$-(T-T_o)/(\log a_T)$	$\log a_T$ (from WLF Eq.)	a_T
5	278	n.a.[a]	1.490	32	21.4765	1.4874	30.7203
10	283	0.160	1.330	27	20.3008	1.3218	20.9817
15	288	0.230	1.100	22	20.0000	1.1376	13.7287
20	293	0.200	0.900	17	18.8889	0.9315	8.5401
25	298	0.100	0.800	12	15.0000	0.6992	5.0023
32	305	0.500	0.300	5	16.6667	0.3197	2.0878
37	310	0.300	0	0	n.a.[a]	0	1

[a]n.a.=not applicable.

TABLE II. Calculation of the shift factor a_T for the 60-year-old male based on the WLF equation [Eq. (4)]. Note: $c_1^o = -1/m = -1/0.1765 = -5.6666$, and $c_2^o = -b/m = -7.2430/0.1765 = -41.0432$ (from Fig. 6).

Temperature (°C)	Absolute temperature (K)	$\Delta \log a_T$ (from G'' data)	Tentative $\log a_T$ (cumulative)	$-(T-T_o)$ (T_o =310 K)	$-(T-T_o)/(\log a_T)$	$\log a_T$ (from WLF Eq.)	a_T
5	278	n.a.[a]	2.465	32	12.9817	2.4825	303.745
10	283	0.200	2.265	27	11.9205	2.2485	177.229
15	288	0.300	1.965	22	11.1959	1.9774	94.9396
20	293	0.250	1.715	17	9.91254	1.6597	45.6726
25	298	0.475	1.240	12	9.6774	1.2820	19.1405
32	305	0.620	0.620	5	8.0645	0.6154	4.1243
37	310	0.620	0	0	n.a.[a]	0	1

[a]n.a.=not applicable.

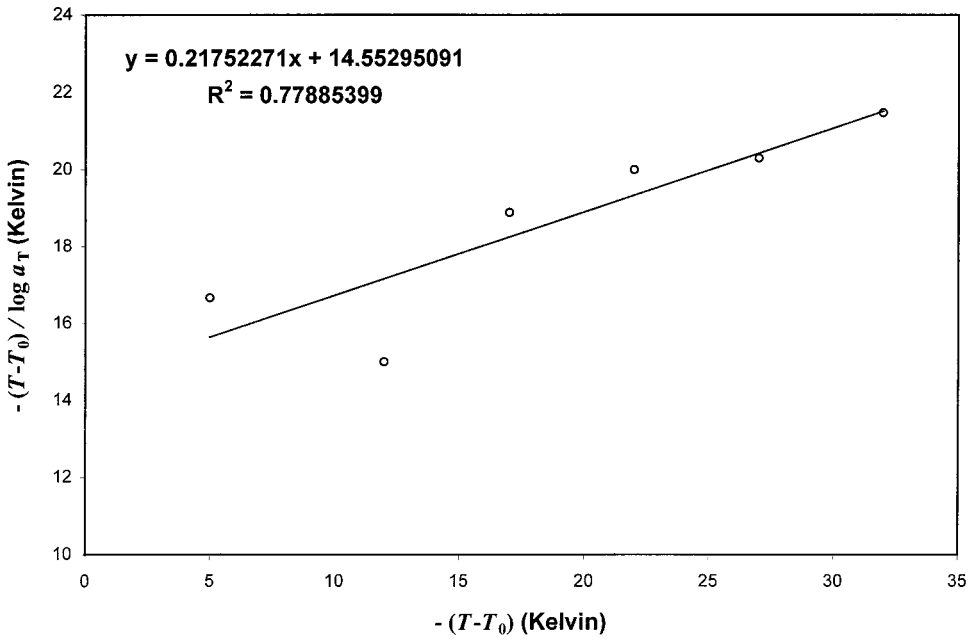


FIG. 5. Linear regression based on Eq. (5) for determining the constants c_1^o and c_2^o of the WLF equation for the 59-year-old male.

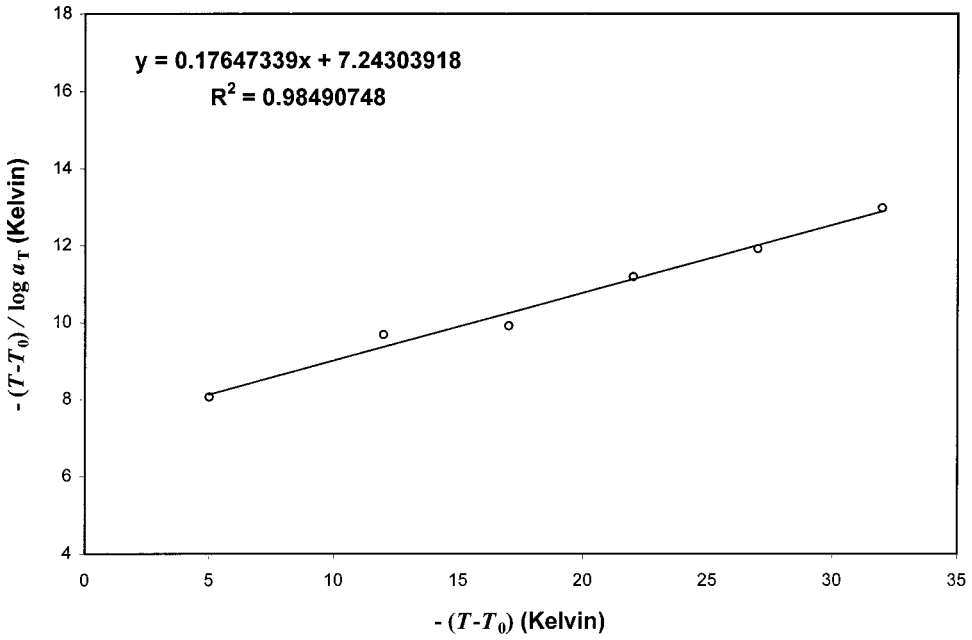


FIG. 6. Linear regression based on Eq. (5) for determining the constants c_1^o and c_2^o of the WLF equation for the 60-year-old male.

Next, each empirical curve of viscoelastic data (G' or G'') obtained at a specific temperature (5°–32 °C) was shifted to the right horizontally by the value of a_T at that particular temperature. Each curve was also shifted upward vertically by a factor of T_o/T, according to the principle of TTS presented in Sec. II. Figures 9–12 each displays a set of shifted curves which tend to superpose or overlap with one another to form a composite curve or ''master curve'' in each figure. The master curve is thus a representation of the estimated magnitude of G' or G'' at the reference temperature T_o (37 °C) (Ferry, 1980). Figures 9 and 10 show the master curves of G' and G'' for the 59-year-old male, while Figs. 11 and 12 show the master curves of G' and G'' for the 60-year-old male, respectively. The master curves for the 59-year-old male (Figs. 9 and 10) show a substantial overlap between the individual shifted curves, especially for G'', except at lower frequencies (<1 Hz). The degree of overlap among the individual curves was not as high for the master curves for the 60-year-old male (Figs. 11 and 12), but still there was excellent superposition at higher frequencies (>500 Hz for G', >100 Hz for G''). It was apparent that the ''goodness of fit'' of the superposition was dependent on the changing slope of the individual empirical curves, with a more positive slope at relatively high frequency leading to a better fit of the master curves, and a less positive (or even slightly negative) slope at lower frequencies leading to gross deviations from the master curves.

The variability of the predicted frequency values on the master curves (the data points) can be expressed in terms of the 95%-confidence intervals of a_T at different temperatures (Tables III and IV), which were estimated based on the standard errors of estimate (S_{YX}) of the linear regression curves

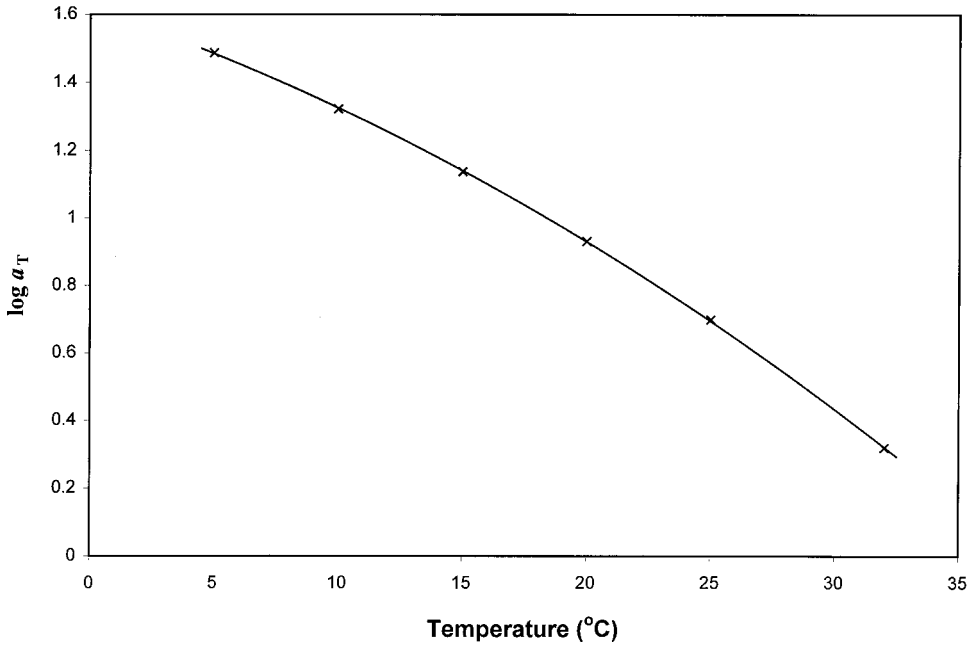


FIG. 7. Temperature dependence of the shift factor a_T for the 59-year-old male.

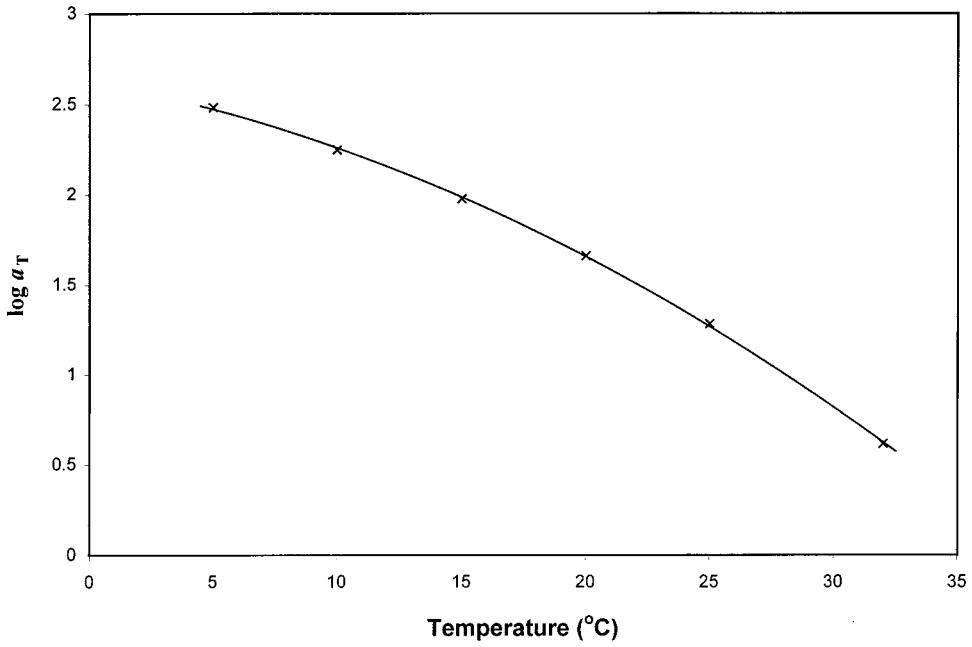


FIG. 8. Temperature dependence of the shift factor a_T for the 60-year-old male.

in Figs. 5 and 6. As can be seen in Figs. 5 and 6, the linear regression curve for the determination of the constants c_1^o and c_2^o for the 59-year-old male was associated with more scattered data points and hence a higher S_{YX} than that of the 60-year-old male (around 5 times higher, see Tables III and IV). As a result, the 95%-confidence intervals of the linear regression's *y*-intercept *b* and slope *m*, the constants c_1^o and c_2^o, as well as ultimately the confidence intervals of a_T, were all higher for the 59-year-old male. In fact, they were all more than 50% of the a_T values obtained for this subject, making the predicted frequency values of the master curves somewhat unreliable with such a magnitude of statistical error, especially for those predictions based on measurements made at higher temperatures (Table III). On the other hand, the 95%-confidence intervals of a_T for the 60-year-old male were all reasonably small, except at the temperature of 32 °C, where it was 30% of the a_T value (Table IV). Fortunately, the large variability at this particular temperature should not affect the overall reliability of the master curves for this subject because the high-frequency portion of the master curves was predicted based on a_T values obtained at lower temperatures (toward 5 °C). These differences in the variability and reliability of the master curves obtained between the two subjects suggested that TTS may only be truly applicable and meaningful for subjects or systems with a small statistical error for the determination of the shift factor a_T.

By TTS, the effective frequency range of the viscoelastic data for the 59-year-old male has been expanded to as high as around 500 Hz (Figs. 9 and 10), while the effective frequency range of the viscoelastic data for the 60-year-old male was expanded to close to 5000 Hz (Figs. 11 and 12). In fact, the maximum extent of frequency range expansion or

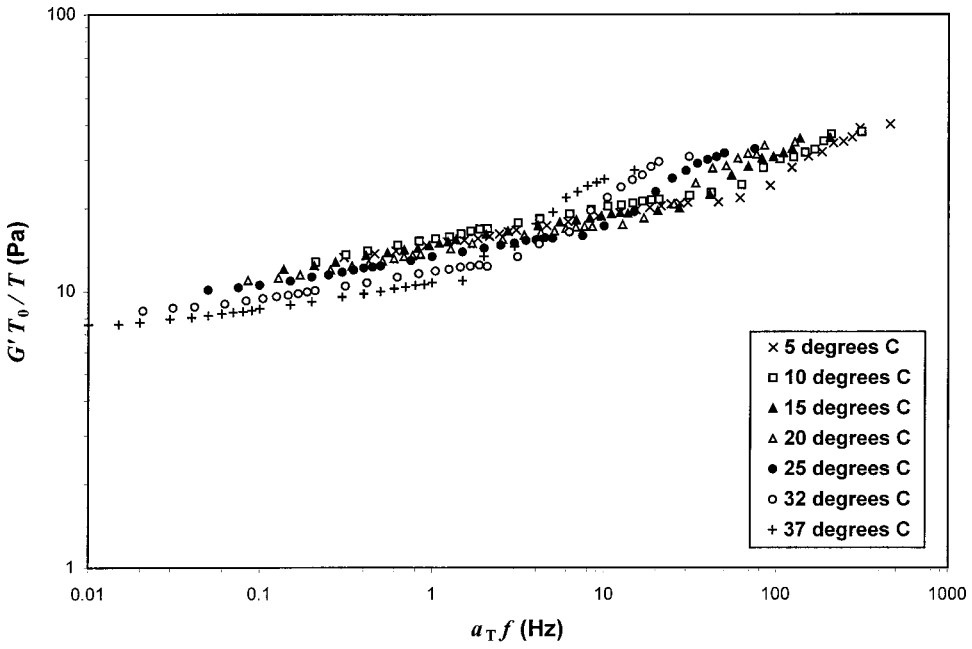


FIG. 9. Elastic shear modulus (G') of human vocal-fold cover (59-year-old male) based on time–temperature superposition (master curve with reference temperature at 37 °C).

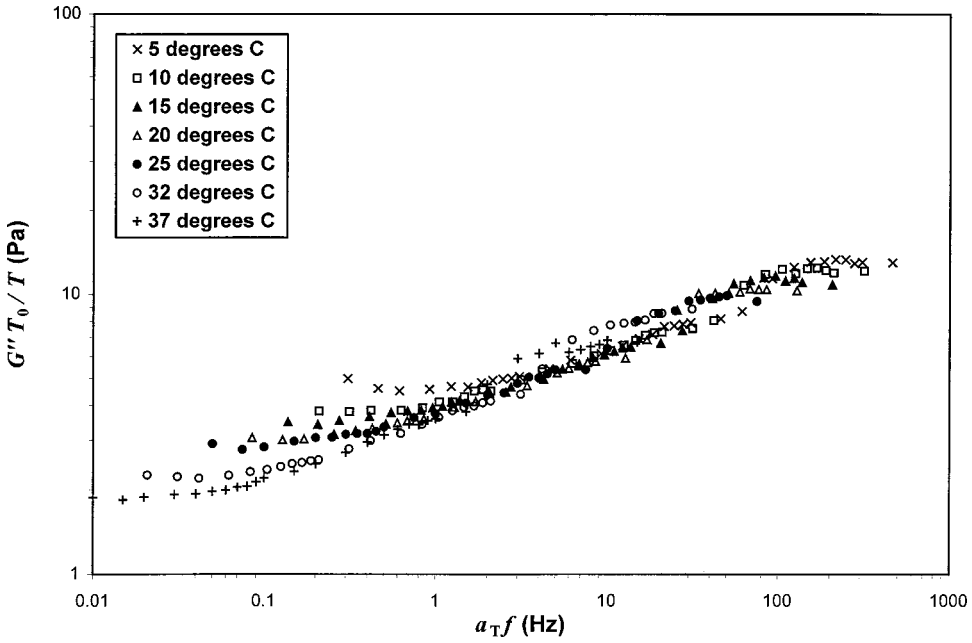


FIG. 10. Viscous shear modulus (G'') of human vocal fold cover (59-year-old male) based on time–temperature superposition (master curve with reference temperature at 37 °C).

data extrapolation was dictated by the maximum value of the shift factor a_T, which always occurred at the lowest temperature (5 °C). For the 59-year-old male, the maximum value of a_T was around 30 (Table I), suggesting that the highest frequency reached by the master curve was around 450 Hz (15 Hz×30). For the 60-year-old male, it was approximately 300 (Table II); i.e., the highest frequency of the master curve was approximately 4500 Hz (15 Hz×300).

Ferry (1980) presented some qualitative criteria for determining the applicability of TTS to a particular set of viscoelastic data or a particular polymeric system, including (1) the shapes of the empirical viscoelastic curves at different temperatures must match over a substantial range of frequencies; (2) the constants c_1^o and c_2^o in the WLF equation should have magnitudes in accord with experience; (3) the same values of a_T must be able to superpose different viscoelastic functions; and (4) the temperature dependence of a_T must have a reasonable form consistent with experience. In the present study, it seemed that all of these criteria have been met, as (1) the shapes of the G' and G'' curves obtained at different temperatures mostly matched with one another over the entire empirical frequency range (0.01–15 Hz) (Figs. 1–4); (2) the values of c_1^o and c_2^o as shown in Tables I and II were around the same orders of magnitude as those tabulated in Ferry (1980) for various homogeneous nonpolar polymeric systems (but note that they were negative values because of the form of the rearranged WLF equation [Eq. (5)] we used); (3) the same set of shift factors (a_T) was indeed used to generate both the G' and the G'' master curves for each subject (values of a_T in Table I for the 59-year-old male, and those in Table II for the 60-year-old male); and (4) Figures 7 and 8 show that the temperature dependence of a_T

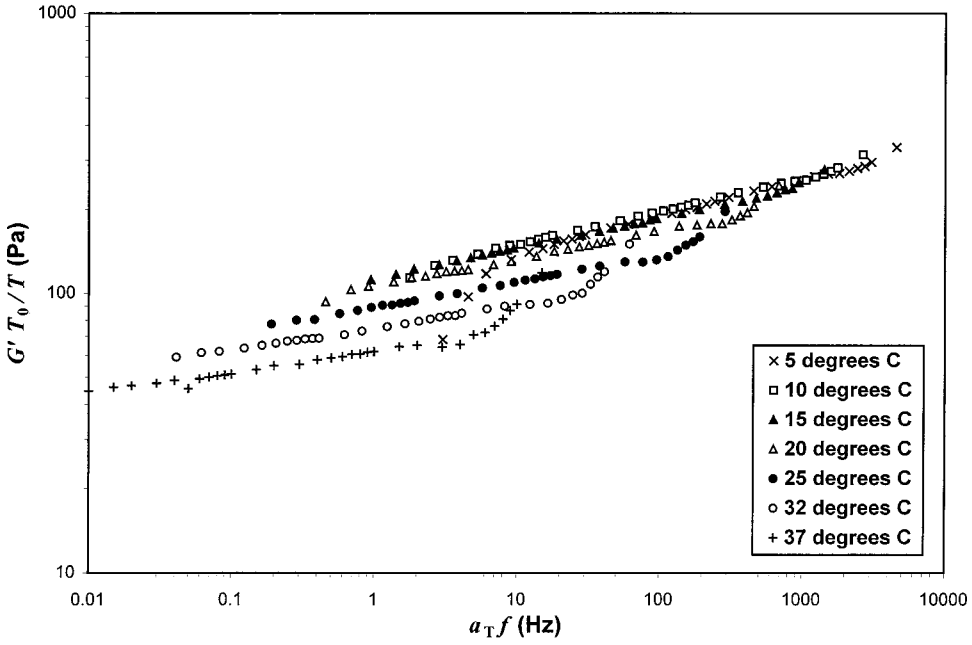


FIG. 11. Elastic shear modulus (G') of human vocal-fold cover (60-year-old male) based on time–temperature superposition (master curve with reference temperature at 37 °C).

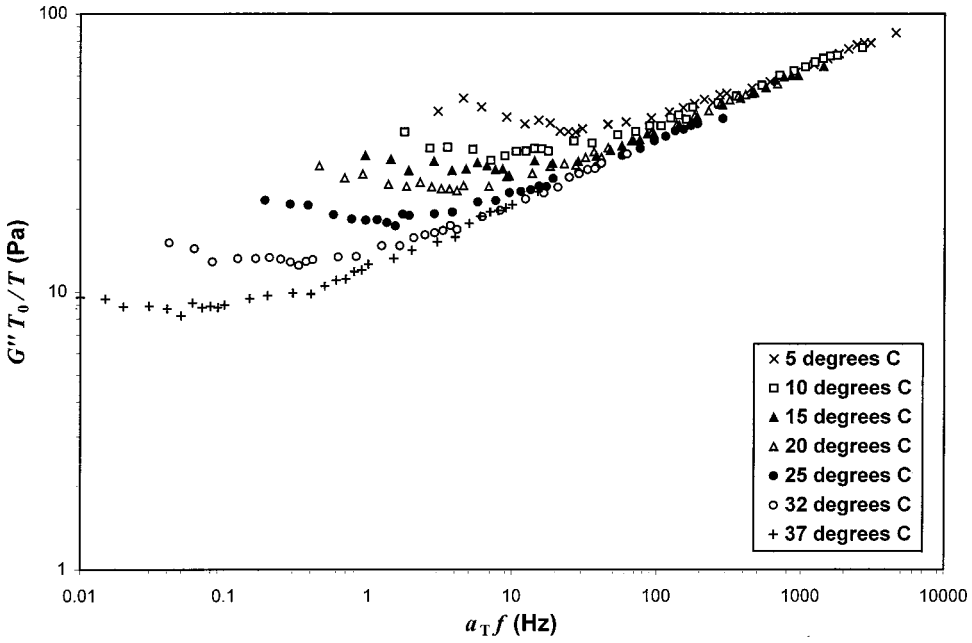


FIG. 12. Viscous shear modulus (G'') of human vocal-fold cover (60-year-old male) based on time–temperature superposition (master curve with reference temperature at 37 °C).

TABLE III. Statistical variability of the linear regression results and the shift factor a_T for the 59-year-old male.

95% Confidence interval	y-intercept b (from S_{YX})[a]	Slope m (from S_{YX},[a] max. and min. m)	c_1^o (from max. and min. c_1^o)	c_2^o (from confidence intervals of b and c_1^o)	$\log a_T$ (from confidence intervals of c_1^o and c_2^o)	a_T	$\%a_T$
(Temperature-independent)	2.2509	0.1407	5.1110	11.5043			
at 5 °C					1.2993	19.9211	64.8467
at 10 °C					1.1535	14.2393	67.8656
at 15 °C					0.9916	9.8083	71.4437
at 20 °C					0.8108	6.4688	75.7462
at 25 °C					0.6077	4.0522	81.0075
at 32 °C					0.2771	1.8930	90.6710

[a]Standard error of estimate of the linear regression in Fig. 5 (S_{YX}=s.d.$\times\sqrt{(1-r^2)}$=1.1484).

TABLE IV. Statistical variability of the linear regression results and the shift factor a_T for the 60-year-old male.

95% Confidence interval	y-intercept b (from S_{YX})[a]	Slope m (from S_{YX},[a] max. and min. m)	c_1^o (from max. and min. c_1^o)	c_2^o (from confidence intervals of b and c_1^o)	$\log a_T$ (from confidence intervals of c_1^o and c_2^o)	a_T	$\%a_T$
(Temperature-independent)	0.4242	0.0265	0.8710	0.3695			
at 5 °C					0.3709	2.3492	0.7734
at 10 °C					0.3352	2.1639	1.2210
at 15 °C					0.2941	1.9683	2.0732
at 20 °C					0.2461	1.7624	3.8589
at 25 °C					0.1894	1.5468	8.0815
at 32 °C					0.0904	1.2313	29.8553

[a]Standard error of estimate of the linear regression in Fig. 6 (S_{YX}=s.d.$\times\sqrt{(1-r^2)}$=0.2164).

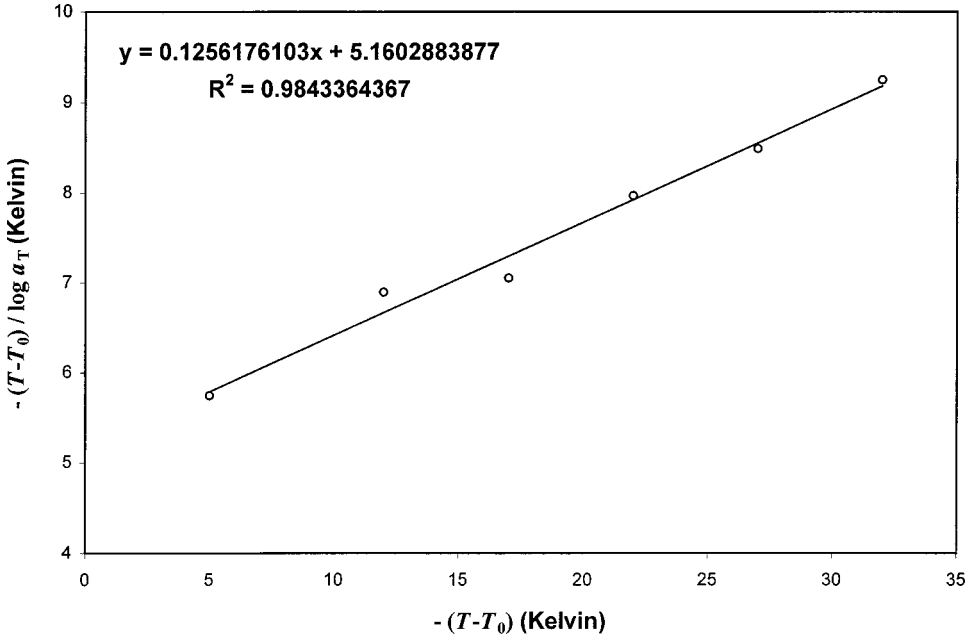


FIG. 13. Linear regression for determining the constants c_1^o and c_2^o for the 60-year-old male (validation analysis of TTS based on empirical data obtained at 0.1–1.0 Hz).

for both subjects was grossly similar to that of common homogeneous nonpolar systems (Ferry, 1980), i.e., a monotonic decrease of a_T with temperature. However, unlike nonpolar systems, the slope of the temperature dependence decreased gradually with temperature (i.e., a_T becoming more sensitive to temperature changes at higher temperatures), which may be related to the aqueous or the nonhomogeneous nature of the system (vocal-fold tissues).

In addition to fulfilling these qualitative criteria, a quantitative validation analysis was performed to further assess the applicability and validity of TTS for vocal-fold tissues. Because empirical data of vocal-fold tissue shear properties were available over a frequency range of three decades (0.01–15 Hz), a ''small-scale'' TTS was performed to shift the viscoelastic data obtained at 5° to 32 °C from 0.1–1.0 Hz to higher frequencies, establishing master curves for G' and G'' at 37 °C which may then be directly compared with the experimental data actually obtained at 37 °C from 1.0–15 Hz. This was the best approach available in the current study for a quantitative evaluation of the applicability of TTS for vocal-fold tissues, because empirical data at frequencies of phonation ($>$ 100 Hz) were unavailable for comparison with the master curves.

Table V shows this validation analysis with an estimation of the values of the shift factor a_T at different temperatures, following the same procedure as described before (cf. Tables I and II). The analysis was done only for the 60-year-old male because of the much smaller statistical errors of the master curves for this subject (Table IV). Figure 13 shows the linear regression line for the determination of the constants c_1^o and c_2^o of the WLF equation, and Fig. 14 shows the temperature dependence of a_T obtained from the WLF equa-

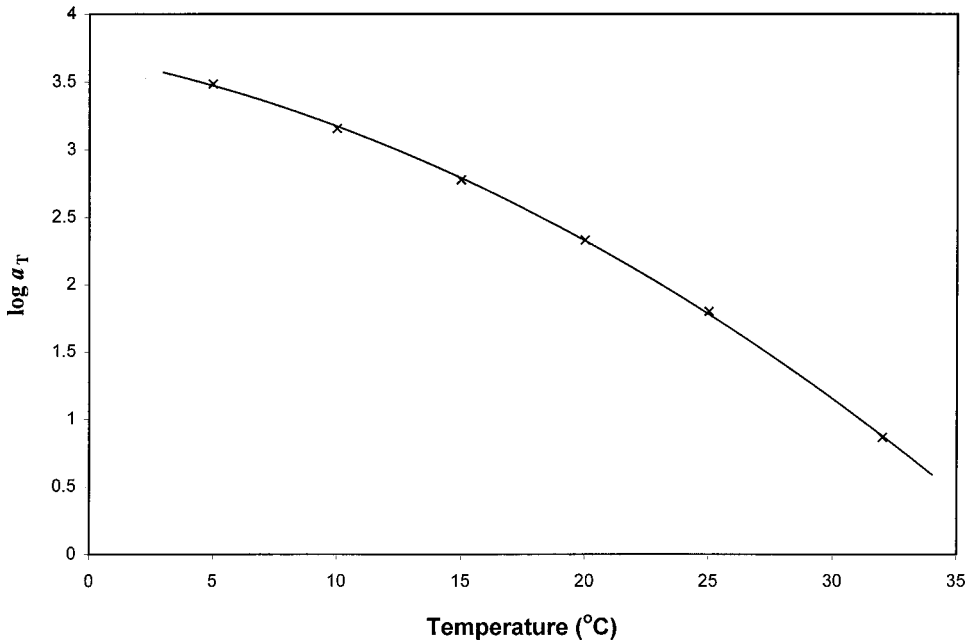


FIG. 14. Temperature dependence of the shift factor a_T for the 60-year-old male (validation analysis of TTS based on empirical data obtained at 0.1–1.0 Hz).

TABLE V. Calculation of the shift factor a_T for the 60-year-old male based on empirical data at 0.1–1.0 Hz (validation analysis for evaluating the applicability of TTS). Note: $c_1^o = -1/m = -1/0.1256 = -7.9607$, and $c_2^o = -b/m = -5.1603/0.1256 = -41.0793$ (from Fig. 13).

Temperature (°C)	Absolute temperature (K)	$\Delta \log a_T$ (from G'' data)	Tentative $\log a_T$ (cumulative)	$-(T-T_o)$ (T_o =310 K)	$-(T-T_o)/(\log a_T)$	$\log a_T$ (from WLF Eq.)	a_T
5	278	n.a.[a]	3.460	32	9.2486	3.4858	3060.69
10	283	0.280	3.180	27	8.4906	3.1572	1436.05
15	288	0.420	2.760	22	7.9710	2.7764	597.612
20	293	0.350	2.410	17	7.0539	2.3301	213.851
25	298	0.670	1.740	12	6.8966	1.7997	63.0552
32	305	0.870	0.870	5	5.7471	0.8638	7.3080
37	310	0.870	0	0	n.a.[a]	0	1

[a]n.a.=not applicable.

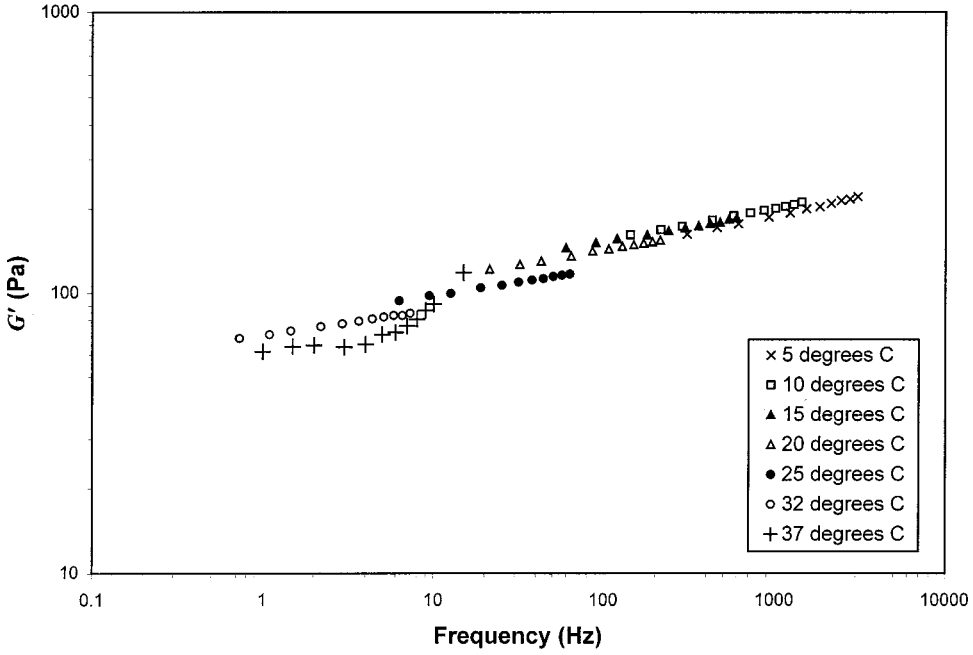


FIG. 15. Master curve for elastic shear modulus (G') of human vocal-fold cover superposed with empirical data obtained at 37 °C (60-year-old male) for evaluating the applicability of time–temperature superposition.

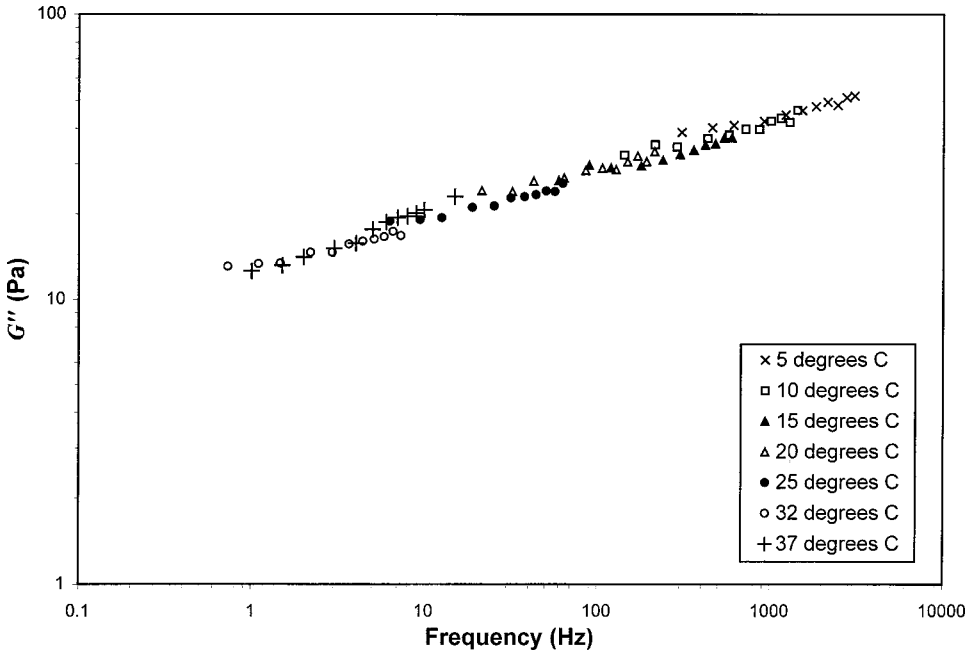


FIG. 16. Master curve for viscous shear modulus (G'') of human vocal-fold cover superposed with empirical data obtained at 37 °C (60-year-old male) for evaluating the applicability of time–temperature superposition.

TABLE VI. Statistical variability of the linear regression results and the shift factor a_T for the validation analysis for evaluating the applicability of TTS.

95% Confidence interval	y-intercept b (from S_{YX})[a]	Slope m (from S_{YX},[a] max. and min. m)	c_1^o (from max. and min. c_1^o)	c_2^o (from confidence intervals of b and c_1^o)	log a_T (from confidence intervals of c_1^o and c_2^o)	a_T	$\%a_T$
(Temperature-independent)	0.3077	0.0192	1.2481	0.3841			
at 5 °C					0.5310	3.3961	0.1110
at 10 °C					0.4799	3.0191	0.2102
at 15 °C					0.4209	2.6360	0.4411
at 20 °C					0.3522	2.2503	1.0523
at 25 °C					0.2711	1.8668	2.9606
at 32 °C					0.1293	1.3468	18.4292

[a]Standard error of estimate of the linear regression in Fig. 13 (S_{YX}=s.d.$\times\sqrt{(1-r^2)}$=0.1570).

tion based on the values of c_1^o and c_2^o. Master curves for G' and G'' were constructed based on these a_T values (Table V). Figures 15 and 16 show the master curves superposed with the empirical data of G' and G'' obtained at relatively high frequency (1.0–15 Hz) at 37 °C, respectively, enabling a direct comparison between experimental data and predictions based on TTS. Table VI shows the variability of the master curves in terms of the 95%-confidence intervals of a_T at different temperatures, based on the standard errors of estimate of the linear regression in Fig. 13. Similar to the results of the ''full-scale'' TTS for the 60-year-old male (Table IV), the 95%-confidence intervals of a_T for the validation analysis were small except at the temperature of 32 °C.

The ''goodness of fit'' between the master curves and the empirical data points at 37 °C (shown as crosses) over the frequency range of overlap (1.0–15 Hz) was assessed statistically by the following correlation coefficient:

$$r=\sqrt{1-\frac{\Sigma_i(x_i-y_i)^2}{\Sigma_i(x_i-\overline{x})^2}}, \tag{7}$$

where x_i are the empirical data values, y_i are the predictions based on TTS, and $\overline{x}$ is the average of x_i. Results showed that the master curves agreed with the experimental data reasonably well, particularly for G'' ($r=0.7296$ for G', 0.9735 for G''). As shown in Figs. 15 and 16, the master curve for G' slightly overpredicted the empirical data at lower frequencies (Fig. 15), but there was excellent agreement between the master curve for G'' and the empirical data (Fig. 16). These findings suggested that TTS may be applicable to vocal-fold tissues for at least one of the subjects examined in the present study.

In the rheology literature (see, e.g., Bird *et al.*, 1987; Ferry, 1980), TTS has been well established as a feasible approach for the extrapolation of empirical viscoelastic data of a wide variety of homogeneous polymeric systems in nonpolar solvents. For polymeric systems in polar solvents or aqueous solutions and for nonhomogeneous systems (including all biological tissues), however, there have been few studies reporting the applicability of TTS until recently. As discussed in Sec. II, Labropoulos *et al.* (2000) showed that it may be feasible to apply TTS to obtain master curves for aqueous polymeric systems, while Peters *et al.* (1997) showed that TTS may be applicable for biological soft tissues. The results of the present study suggested that as long as the statistical errors associated with the determination of the shift factor a_T (the linear regression procedure) are reasonably small, TTS may be a feasible approach for extrapolating the viscoelastic data of vocal-fold tissues to frequencies of phonation, supporting the applicability of TTS to nonhomogeneous aqueous systems despite its theoretical assumptions that are essentially incompatible with such systems. Our reasonably successful application of TTS to vocal-fold tissues for at least one subject (the 60-year-old male) was consistent with the results of Peters *et al.* (1997) and Labropoulos *et al.* (2000). Together with their reports, this study has provided some of the very first evidence supporting the use of TTS for systems other than homogeneous polymers in nonpolar solvents. Further studies on other polar/aqueous polymeric systems and biological tissues are needed to better establish such novel applications of the principle of TTS.

V. CONCLUSION

The purpose of this study was to establish the feasibility of the principle of time–temperature superposition (TTS) as a data extrapolation approach for the estimation of vocal-fold tissue shear properties at frequencies of phonation, a relatively high frequency range not yet accessible experimentally (on the order of 100 Hz). The elastic shear modulus and viscous shear modulus of vocal-fold cover specimens were empirically measured at a range of relatively low frequencies (0.01–15 Hz) and low temperatures (5°–37 °C) and data were shifted to generate master curves of the viscoelastic functions at 37 °C, covering much higher frequencies on the order of 100–1000 Hz. Despite the theoretical assumptions of TTS that are incompatible with complex polymeric systems such as biological tissues, our results showed that for one of the two subjects examined TTS may be a feasible approach for extrapolating the low-frequency viscoelastic data of vocal-fold tissues to higher frequencies. Further studies involving more subjects are needed, but the feasibility of TTS for vocal-fold tissues can only be truly validated when empirical viscoelastic data at high frequencies are available for comparison with predictions from TTS. Empirical mea-

surements involving the use of a rheometer capable of reaching higher frequencies are currently underway.

ACKNOWLEDGMENTS

A portion of this work (empirical data collection) was done while the author was affiliated with the National Center for Voice and Speech (NCVS) at the University of Iowa, with support from NIH (NIDCD Grant No. P60 DC 00976). The author gratefully acknowledges the generous support of Professor Ingo R. Titze, Director of NCVS, and an anonymous reviewer for constructive and helpful comments which significantly strengthened this paper.

Alipour, F., Berry, D. A., and Titze, I. R. (**2000**). ''A finite element model of vocal fold vibration,'' J. Acoust. Soc. Am. **108**, 3003–3012.

Bird, R. B., Armstrong, R. C., and Hassager, O. (**1987**). *Dynamics of Polymeric Liquids. Vol. 1: Fluid Mechanics* (Wiley, New York), pp. 99–166.

Chan, R. W. (**1998**). ''Shear properties of vocal fold mucosal tissues and their effect on vocal fold oscillation,'' unpublished Ph.D. dissertation, Iowa City, The University of Iowa.

Chan, R. W., and Titze, I. R. (**1999a**). ''Viscoelastic shear properties of human vocal fold mucosa: Measurement methodology and empirical results,'' J. Acoust. Soc. Am. **106**, 2008–2021.

Chan, R. W., and Titze, I. R. (**1999b**). ''Hyaluronic acid (with fibronectin) as a bioimplant for the vocal fold mucosa,'' Laryngoscope **109**, 1142–1149.

Chan, R. W., and Titze, I. R. (**2000**). ''Viscoelastic shear properties of human vocal fold mucosa: Theoretical characterization based on constitutive modeling,'' J. Acoust. Soc. Am. **107**, 565–580.

Chan, R. W., and Titze, I. R. (**2001**). ''Effects of post-mortem changes and freezing storage on the viscoelastic properties of vocal fold tissues'' (to be published).

Ferry, J. D. (**1980**). *Viscoelastic Properties of Polymers*, 3rd ed. (Wiley, New York), pp. 264–320.

Labropoulos, K. C., Rangarajan, S., Niesz, D. E., and Danforth, S. C. (**2000**). ''Correlation of the dynamic rheological behavior and microstructure of agar gel based aqueous binders for powder injection molding,'' Ceram. Trans. **115**, 161–172.

Peters, G. W. M., Meulman, J. H., and Sauren, A. A. H. J. (**1997**). ''The applicability of the time/temperature superposition principle to brain tissue,'' Biorheology **34**, 127–138.

Titze, I. R. (**1992**). ''Phonation threshold pressure: A missing link in glottal aerodynamics,'' J. Acoust. Soc. Am. **91**, 2926–2935.

Effects of oscillation of a mechanical hemilarynx model on mean transglottal pressures and flows

Fariborz Alipour[a)]
Department of Speech Pathology and Audiology, The University of Iowa, 334 WJSHC, Iowa City, Iowa 52242-1012

Ronald C. Scherer
Department of Speech Pathology and Audiology, The University of Iowa, Iowa City, Iowa 52242-1012 and Department of Communication Disorders, Bowling Green State University, Bowling Green, Ohio 43403

(Received 4 October 2000; revised 18 April 2001; accepted 29 July 2001)

This study introduces a mechanical model of the larynx for investigating dynamic aerodynamic effects of phonation. The model mimics the hemilarynx. The tracheal inlet section was rectangular (25-mm width, 20-mm height). The vocal fold was fabricated with precision machinery from hard plastic with an attached oscillating plunger. A speaker assembly and audio amplifier drove the plunger, mimicking one-dimensional vocal-fold motion toward a flat wall. The glottal shape was rectangular. The glottal diameter was well specified or dynamically followed with a laser system. The air was sucked through the channel using a vacuum with controlled speed. Frequency and amplitude of the glottis were varied. The mean pressure and mean flow data were recorded. For steady-flow conditions, the glottal gap ranged from 0.39 to 2.58 mm. The pressure coefficient for steady flow had a range of 3.1 to 1.3 for Reynolds numbers between 300 and 9000. For oscillation conditions (a) the frequency was varied from 75 to 150 Hz while the amplitude was held relatively constant, and (b) the amplitude was varied to 0.3 mm for a fixed frequency of 100 Hz. The results indicate that the hemilarynx model provides mean pressure-flow data similar in form to other models with two vocal folds. Furthermore, the dimensional and non-dimensional pressure coefficient is sensitive to variations in glottal gap and glottal amplitude, but relatively insensitive to the frequency of oscillation. [DOI: 10.1121/1.1396334]

PACS numbers: 43.70.Bk [AL]

I. INTRODUCTION

The purpose of this study was to introduce a physical model of the hemilarynx in which the vocal fold could be driven sinusoidally with various amplitudes and frequencies to determine mean transglottal pressure-flow relationships for different glottal gaps, frequencies, and amplitudes. This information should enhance our understanding of hemilarynx function and unsteadiness in phonation.

Physical models of laryngeal function can be made with precise geometry of the vocal folds and the glottis, and are used to control transglottal pressure and motion of the vocal folds (e.g., Alipour, Scherer, and Knowles, 1996; Alipour and Scherer, 2000a; Scherer and Guo, 1990; Mongeau *et al.*, 1997). Physical models therefore can be used to study mean flows, flow resistance, velocity profiles, intraglottal pressures, and acoustic generation, all being dependent upon the glottal geometry and transglottal pressure. Such models, however, may also be limited in their similarity to actual phonation, and such limitations need to be determined.

Physical models of the larynx with nonmoving vocal folds have been used to establish equations that relate glottal geometry, transglottal air pressure, and glottal flow (van den Berg, Zantema, and Doornenbal, 1957; Ishizaka and Flanagan, 1972; Binh and Gauffin, 1983; Scherer, Titze, and Curtis, 1983; Scherer and Guo, 1990). The motivation for these studies has been to create equations that can be used in computer models of phonation. The applicability of these studies has been based on the assumption that phonation is approximately a quasisteady phenomenon so that the pressure-flow-geometry results hold for both the constant conditions as well as for ''snapshots'' during the oscillating condition during phonation. Although this assumption has been questioned (Teager and Teager, 1983; McGowan, 1993), the dynamic physical model research of Mongeau *et al.* (1997) found that a forced-oscillation model of phonation, in which the two vocal folds (forming a convergent glottal shape) were forced to oscillate toward and away from each other, supported the quasisteady assumption for the most part. The physical model introduced here may add important information to further examine the quasisteady assumption.

The model used in this study was of the vibrating hemilarynx type, that is, with one vocal fold opposite a flat wall. The most direct human similarity is with the hemilarynx created by surgery to remove one diseased vocal fold, leaving one vocal fold to vibrate against the opposite reconstructed airway wall (Hirano, Kurita, and Matsuoka, 1987). It is also directly related to excised animal hemilarynx studies in which one-half of the larynx is removed and bench-mounted with a flat plate opposite the vocal fold so that, for example, contact pressures (Jiang and Titze, 1993, 1994) and dynamic pressures (Alipour and Scherer, 2000b) measured at the flat plate can be obtained.

[a)]Author to whom correspondence should be addressed; electronic mail: alipour@shc.uiowa.edu

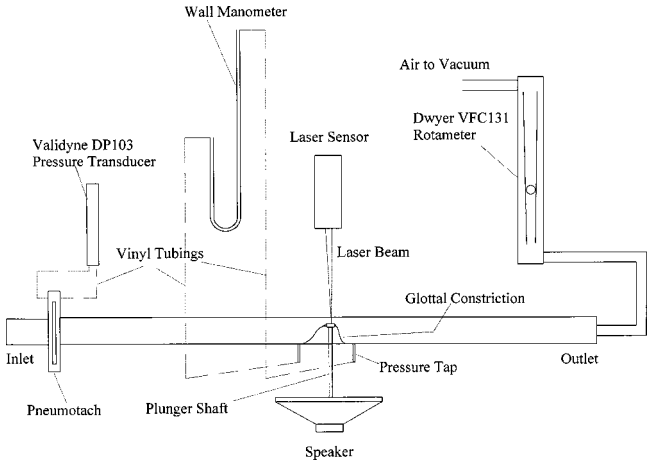


FIG. 1. Schematic of the mechanical hemilarynx model and experimental setup.

The Plexiglas model and setup to be described below was similar to that produced by Ikeda, Matsuzaki, and Sasaki (1994). Their model also incorporated a single piston-like vocal fold that could oscillate near a flat wall. Their study, however, differed significantly from the present one because they were interested in flow reattachment downstream of the glottis and reported empirical results for steady flow and for only a 4-Hz oscillation frequency. Shadle, Barney, and Thomas (1991) also built a mechanical model with driven vocal folds. Their goal also was different from the current one, in that they were interested in describing the jet flow from the glottis using flow visualization.

II. METHODS

The Plexiglas model was designed with three sections, the inlet tunnel, the glottis, and the outlet tunnel (Fig. 1). The inlet section (with pneumotach attached) was 98 cm long and rectangular (25 mm wide, 20 mm high). Straws were placed at the entrance of the inlet section to smooth and laminarize the flow. The outlet section was also rectangular with the same cross section, and 60 cm long (111 cm to a downstream flow valve). The end was connected to a flowmeter (Dwyer model VFC131) by means of 22 cm of flexible tubing with a 2.5-cm inside diameter. A vacuum source drew the air through the model.

The vocal-fold section of the model was fabricated out of hard plastic using precision machinery. The shape of the vocal fold included a sinusoidal convergence from the tunnel wall up to the entrance of the glottis. The exit also had a sinusoidal shape with a steeper descent to the tunnel wall. The medial surface of the vocal fold and the opposite Plexiglas wall formed a rectangular glottal duct. Within the glottal section, a rectangular piece (7 mm long and 25 mm wide) was attached to a speaker cone by means of a small hollow shaft. A latex sheet was stretched over the vocal-fold piece, covering the movable rectangular portion. The sheet was taped to the vocal-fold piece near the tunnel wall by thin plastic tape. The latex sheet formed a smooth transition between the moving rectangular section of the glottis and the neighboring glottal portions (glottal entrance and exit). The rectangular section of the vocal-fold model acted as a ''plunger'' because of the sinusoidal movement of the speaker system (6-in. speaker cone, audio amplifier, and function generator).

To measure the glottal gap between the surface of the latex and the top Plexiglas plate, an analog laser system (Balluff, model BOD-26K) was used. The laser signal was shone through the Plexiglas perpendicular to the surface of the vocal fold. The laser beam fell onto a small silver spot painted on the top of the latex covering the vocal fold. The reference position for determining subsequent glottal gaps was determined by moving the vocal-fold plunger vertically so that the vocal fold touched the Plexiglas plate. The glottal gap was measured (± 0.005 mm) by subtracting the signal at a given location from the reference signal. The gap size was corrected for the approximate thickness (0.02 mm) of the silver mark on the latex sheet. When the vocal fold was oscillated sinusoidally, the changing position of the vocal fold was monitored by the laser system. Setting the peak-to-peak voltage from the laser system, while changing the audio amplifier gain level, controlled the amplitude. The levels were read from a Data 6000 digital oscilloscope (Data Precision, Analogic Corporation). Amplitudes ranged from 0 to 0.3 mm. It is noted that the amplitude of 0.3 mm is about one-third of what is expected in normal phonation in adult males (Hirano *et al.*, 1981), so this study would correspond to smaller amplitudes of phonation without glottal closure. During oscillation conditions, the mean glottal gap was obtained by reading the mean laser voltage.

The airflow was measured with a pneumotach calibrated against a midrange flowmeter (Gilmont, model J197) and a high-range flow meter (Dwyer model VFC131). The pneumotach was placed at the entrance of the inlet tunnel (Fig. 1). Its frequency response, with tubing and pressure transducer (Validyne DP103) was approximately 100 Hz (following charts in Edmonds, Lilly, and Hardy, 1971). The mean pressure drop across the glottis was measured by a well-type manometer (Dywer), using pressure taps 30 mm upstream of the moving vocal-fold section and 23 mm downstream of the moving vocal-fold section. The time-varying pressure drop across the glottis was measured with a Validyne DP45 pressure transducer from the same pressure taps. The pressure taps were located on the floor of the tunnel, opposite the side near which the vocal fold vibrated. The mean values of pressure and flow were used for both the constant glottal gap and the oscillating glottal gap conditions.

III. RESULTS

A. Mean pressure-flow relationships for constant glottal gap

The glottal gap of the hemilarynx mechanical model was set at different values to obtain basic pressure-flow-geometry relationships for the nonoscillating glottis. Results for five gaps are shown in Fig. 2. The curve fit for each gap is quadratic, consistent with other studies of glottal pressure-flow results in physical laryngeal models. Figure 3 shows the transglottal pressure coefficient versus the Reynolds number. The pressure coefficient was defined as

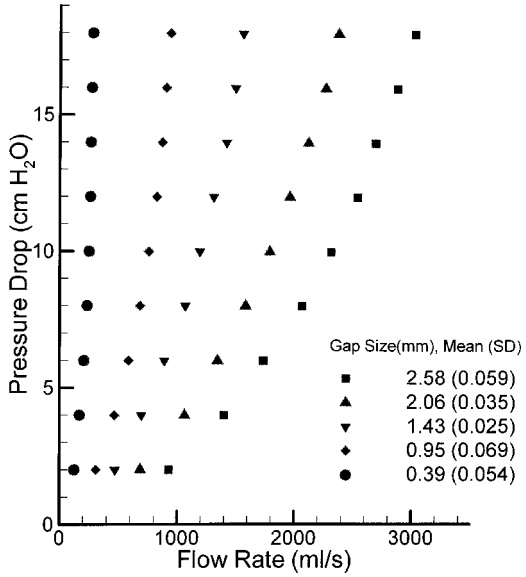


FIG. 2. Mean pressure-flow relationship for various constant gap sizes.

$$P^* = \frac{P}{0.5\rho V^2}, \tag{1}$$

where $\rho = 1.1767$ g/cm^3 is the air density, and V is the mean air velocity in the glottal duct, obtained by dividing the volume velocity by the glottal area (the glottal gap times the length of the vocal folds). The Reynolds number was defined as

$$\mathrm{Re} = \frac{VG}{\nu}, \tag{2}$$

where G is the glottal gap and $\nu = 0.15$ cm^2/s is the kinematic viscosity of air. The data of Fig. 3 essentially collapse onto one curve, suggesting that the steady-flow data for the rectangular glottis of this hemilarynx configuration was insensitive to gap size when expressed in nondimensional terms. Both the dimensional and nondimensional figures are consistent with similar figures for the rectangular-glottis model results presented in Scherer, Titze, and Curtis (1983).

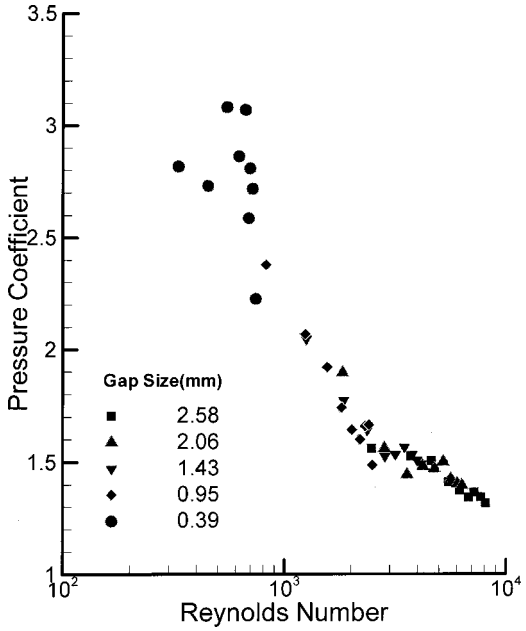


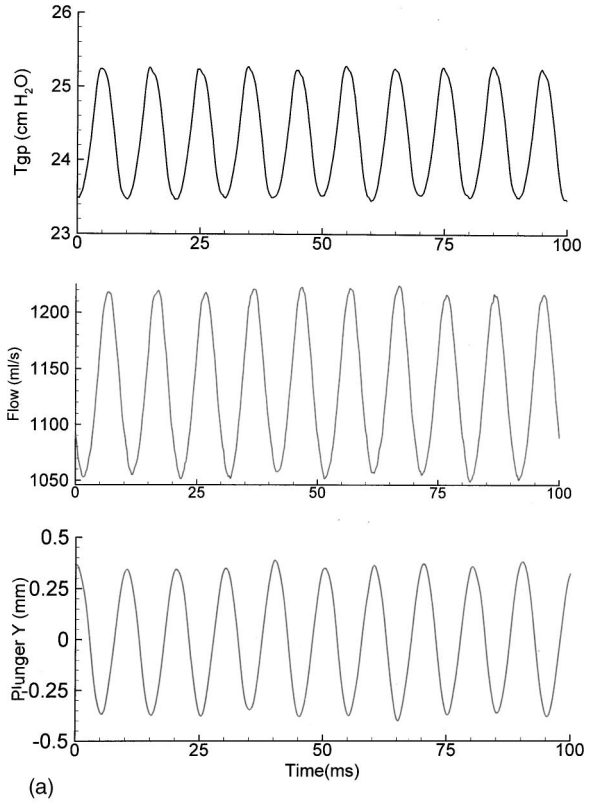
FIG. 3. Mean pressure coefficient as a function of Reynolds number for various constant gap sizes.

B. Effect on the mean flow due to pulsation of the flow

The oscillating vocal fold created flow oscillation ''ripples'' on the constant flow. The data reported in this study are the mean transglottal pressures and the mean flows. The flow oscillations may increase the mean flow measurements according to Doebelin (1990, p. 569) when the transglottal pressure is a function of the flow squared, as it is in this study. That is, the average transglottal pressure is a function not only of the mean flow squared, but also the square of the amplitude of the flow ripple: $P_{ta} = k(Q_{av}^2 + Q_p^2/2)$, where P_{ta} is the average transglottal pressure, Q_{av} is the actual average flow, and the Q_p is the amplitude of the flow ripple. The pneumotach recorded the oscillating flows at the entrance of the flow tunnel. Figure 4(a) shows oscillations of transglottal pressure, volume flow rate, as well as the displacement of the plunger for an oscillation fundamental frequency of 100 Hz, a glottal gap of 1.2 mm, amplitude of motion of 0.378 mm, and a mean transglottal pressure of 24.2-cm H_2O. Figure 4(b) shows three curves of flow as a function of transglottal pressure, the maximum flow, the average flow, and the minimum flow for each pressure condition. This case is a ''worst case'' relative to the potential influence of the flow amplitude on the average flow calculation for this study. The frequency of oscillation was below the measured first subglottal resonance (approximately 195 Hz). The average flow ripple amplitude in Fig. 4(b) was 7.7% of the mean flow. Even when using twice the amplitude ($2Q_p$) in the above expression, the difference between the measured mean flow and the ''actual'' mean flow in the expression above was less than 0.9%. This suggests that for the research reported here, the pulsatile nature of the flow did not yield mean flows that were inaccurate due to the pulsatility. The mean pressure and mean flow data should reflect the effects of gap size, amplitude of vocal-fold motion, and oscillation frequency, and not be significantly affected by the fluctuations of the signals.

C. Amplitude effects

The hemilarynx was oscillated at three amplitudes, 0.1, 0.2, and 0.3 mm, for mean nominal glottal gaps of 1.0 and 1.5 mm and a frequency of 100 Hz. For each condition, a number of mean transglottal pressures were used. Figure 5 shows the results for the nominal 0.1-mm (solid symbols) and 1.5-mm (hollow symbols) glottal gaps. The curves range from no amplitude (the steady-flow condition discussed above) to 0.3 mm. The latter, for example, means that the vocal-fold plunger varied 0.3 mm above and below the mean gap (1.0 or 1.5 mm). The figure indicates that for the relatively small glottal gap of 1.0 mm, the greater the oscillation amplitude, the less was the mean flow for any given value of transglottal pressure. For example, for the 1-mm gap at 10-cm H_2O pressure, the flow for the no-oscillation condition



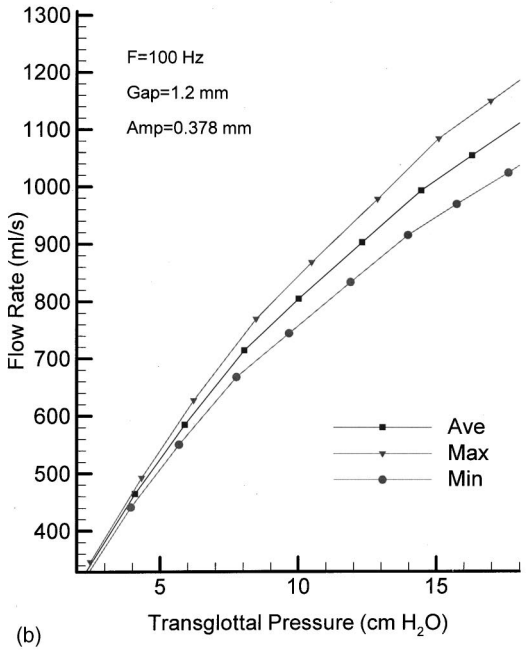


FIG. 4. (a) Waveforms of transglottal pressure, flow rate, and plunger displacement oscillating at a frequency of 100 Hz and amplitude of 0.378 mm for a mean glottal gap of 1.2 mm. (b) Maximum, mean, and minimum flow rates of oscillation shown in (a) as a function of transglottal pressure.

was 758 ml/s, and 624 ml/s for the 0.3-mm amplitude, a drop of 18%. The corresponding flow resistance (transglottal pressure divided by the flow) was 21%. The findings are a little different for the larger glottal gap of 1.5 mm. Here, the data for amplitudes of 0.1 and 0.2 mm are virtually the same as for the no-oscillation condition. The largest oscillation amplitude, 0.3 mm, created greater flow resistances (at 10-cm H_2O, the change in flow resistance was 11%). The results from Fig. 5 therefore suggest that for this hemilarynx model, the pressure-flow relationships are more similar to the steady-flow results for greater gaps and smaller amplitudes, and the effect of amplitude appears to be present for all transglottal pressures, especially for smaller gaps.

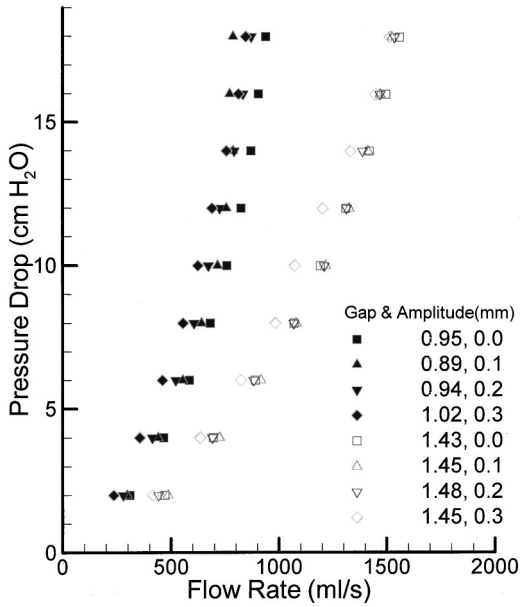


FIG. 5. Mean pressure-flow relationships for various glottal amplitudes for a frequency of 100 Hz. Solid symbols represent data for the smaller glottal gap group and hollow symbols for larger glottal gap group.

The pressure coefficient versus Reynolds number for the amplitude data is shown in Fig. 6 with solid symbols for the nominal 1.0-mm gap and hollow symbols for the 1.5-mm gap. For the 1.0-mm gap, the increase of the amplitude of oscillation had an increasing effect on the pressure coeffi-

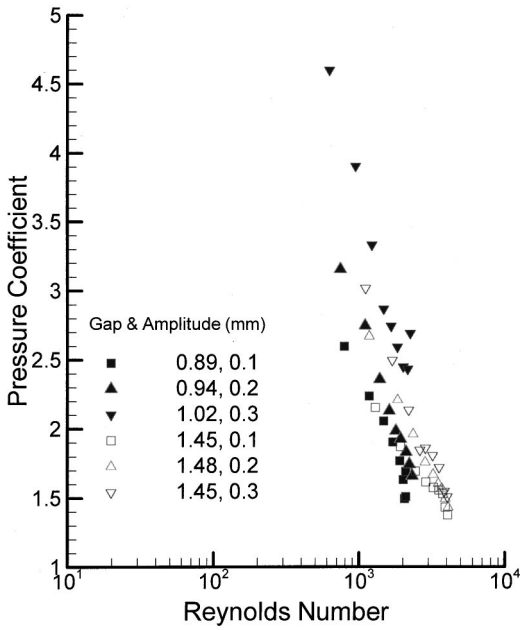


FIG. 6. Mean pressure coefficient as a function of Reynolds number for various amplitudes for a frequency of 100 Hz, corresponding to the data of Fig. 5.

cients. Unlike the steady-flow case, these data points do not converge into one curve. Although the different amplitudes separate the data, the amplitude data are similar across the two glottal gaps; that is, the data for the same amplitude for the two glottal gaps nearly collapse on the same curve, but the curves separate out according to amplitude. Thus, the dimensional display of the data (Fig. 5) separates the data strongly relative to the gap size (and flow resistance), whereas the nondimensional display (Fig. 6) separates the data more on the basis of amplitude of the oscillation.

D. Frequency effects

The results for varying fundamental frequency (with a constant amplitude of 0.2 mm) are shown in Fig. 7(a). As was suggested in Fig. 5, there is a strong effect due to the gap. Using the data of Fig. 5 for the amplitude of 0.2 mm [the amplitude used in Fig. 7(a)], the following relationship was found:

$$\frac{dF}{dG} = 286.1 P_t^{0.5}, \tag{3}$$

where the left-hand side is the change in flow per mm gap change, and P_t is the transglottal pressure. Using that finding to "correct" the data of Fig. 7(a) yields Fig. 7(b), which shows that the effect of frequency change is essentially negligible (within reasonable experimental error). This is in agreement with Mongeau *et al.* (1997). The pressure coefficient versus Reynolds number for the gap-corrected data, shown in Fig. 8, converge essentially onto one curve as was the case for zero-frequency (steady-flow) data. (The uncorrected data tend to coalesce into two groups based on the gap size, having higher-pressure coefficient values for the 1.5-mm gap size.)

E. Strouhal number

The Strouhal number (St) is a nondimensional term that often can give insight into the unsteadiness of the phenomenon being observed, in this case, the oscillation of the hemilarynx model. The Strouhal number is defined as

$$\mathrm{St} = \frac{fG}{V}, \tag{4}$$

where f is the frequency of oscillation, G is the glottal gap, and V is the average velocity (volume flow divided by the cross-sectional area of the glottis).

The next figures show the Strouhal number for gap and amplitude with frequency held constant at 100 Hz (Fig. 9) and for gap and frequency with the amplitude held constant at 0.2 mm (Fig. 10). Figure 9 shows a strong separation of the data between the smaller and larger gap sizes (average of 0.95 and 1.46 mm, respectively), with higher values of the Strouhal number for the same pressure coefficient values for the larger gaps. In addition, the amplitude of oscillation data separates out better for the smaller gap sizes than for the larger gap sizes; this suggests that the sensitivity to amplitude is greater for smaller gaps. The larger pressure coefficient values for larger amplitudes were consistent with the data of Fig. 6. For the larger gaps of Fig. 9, the data overlap

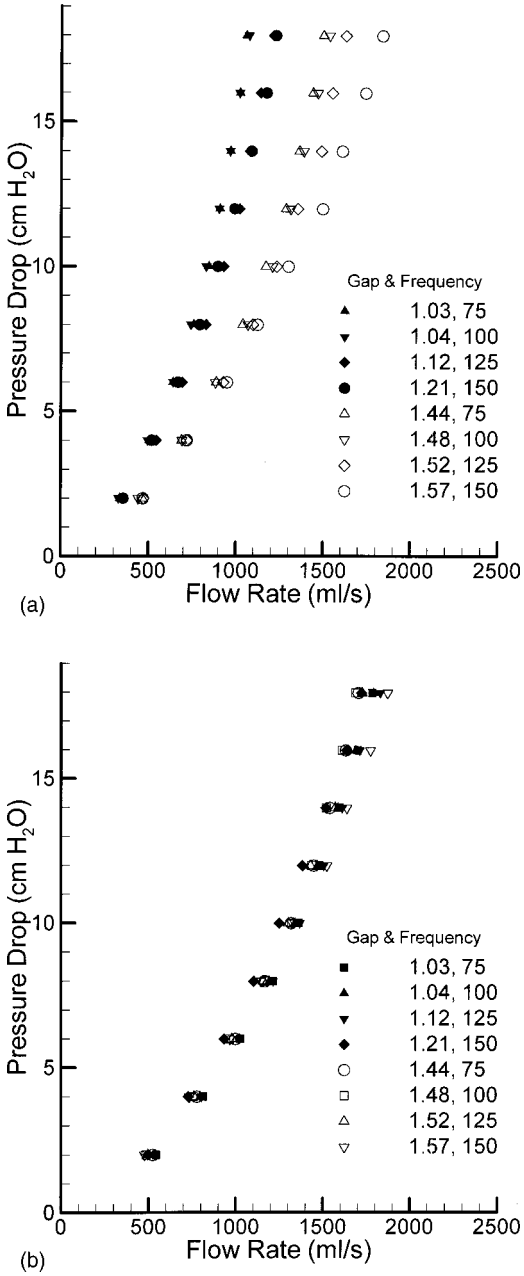


FIG. 7. (a) Mean pressure-flow relationship for various frequencies for an amplitude of 0.2 mm. Solid symbols represent data for the smaller glottal gap group and hollow symbols for larger glottal gap group. (b) Mean pressure-flow relationship for various frequencies for an amplitude of 0.2 mm. Flow rate has been corrected for the effect of gap size.

for lower St values, and begin to separate relative to amplitude at an St value of about 0.004. The sensitivity to both glottal gap size and amplitude of oscillation is greater in the pressure coefficient versus Strouhal number data (Fig. 9) than it is in the pressure coefficient versus Reynolds number data (Fig. 6) because the Strouhal number takes the frequency of oscillation into account as well as velocity and glottal gap. It is noted that the operating range of the Strouhal and Reynolds numbers of the mechanical larynx model is similar to that found in full excised larynges (Alipour, Scherer, and Patel, 1995). The range in the present study starts at lower values than found in Alipour *et al.* (1995) due to the absence of a threshold pressure required for self-oscillations in excised larynges.

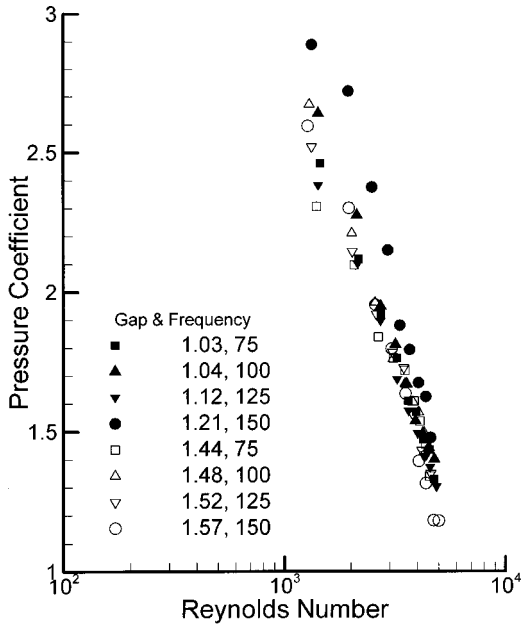


FIG. 8. Mean pressure coefficient as a function of Reynolds number for various frequencies for a glottal gap of 0.2 mm, corresponding to the data of 7(b).

Figure 10 shows the pressure coefficient versus Strouhal number data for gap and frequency with the amplitude of oscillation held constant at 0.2 mm. There is an overlap of data between the smaller and larger gap sizes. However, the Strouhal number is always larger for the larger gap at the same pressure coefficient and frequency. The data are well separated by frequency for the larger gap group. This is less obvious for the smaller gap group, but most likely would have also occurred if the size of the gap for the 150-Hz condition had a gap closer to the values of the other gaps (a target of 1.0 mm, instead of its value of 1.212 mm). Again, it is noted that the pressure coefficient versus Strouhal number is more sensitive to frequency than its counterpart figure comparing the pressure coefficient with the Reynolds number (Fig. 8).

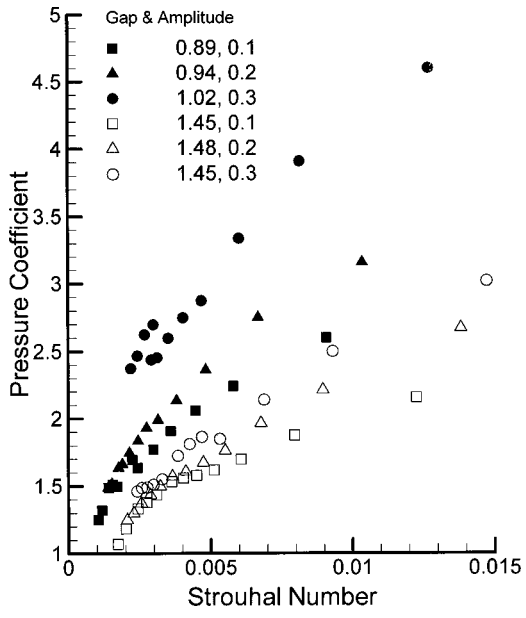


FIG. 9. Mean pressure coefficient as a function of Strouhal number for various amplitudes and glottal gaps (frequency held constant at 100 Hz).

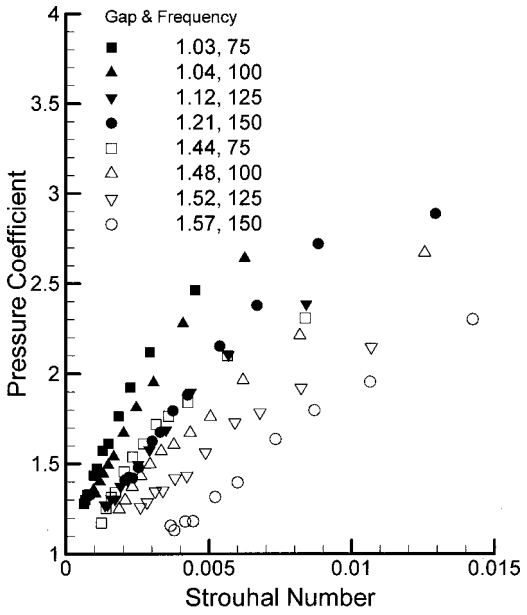


FIG. 10. Mean pressure coefficient as a function of Strouhal number for various frequencies and glottal gaps (amplitude held constant at 0.2 mm).

IV. DISCUSSION

The vibration of this hemilarynx model is different from normal phonation in a number of ways. The glottis remains rectangular through the oscillations, rather than changing from convergent to divergent; the motion does not include collision of the vocal fold at the opposite flat plate, and the motion is driven, rather than flow induced. The results here, therefore, speak more to the dynamics of the pressure-flow characteristics of a driven glottal wall. Apparently no other studies, however, have controlled the oscillation frequency, amplitude, and transglottal pressure with values similar to human phonation as this study has. It is expected, therefore, that this study should yield insights into certain aspects of the dynamics of glottal aerodynamic behavior, which will now be discussed.

The dimensional data of Figs. 5 and 7 suggest that the average glottal gap size strongly affects the flow resistance (or flow rate given a transglottal pressure drop), and that a gap difference of about 0.5 mm (approximately 1.0 to 1.5 mm) yields larger differences in average flow or flow resistance than the variation due to amplitudes up to 0.3 mm or variation due to frequencies up to 150 Hz.

However, the effects of amplitude of oscillation were not negligible for the range of values used in this study. Flow resistance increased by approximately 21% as the amplitude of motion increased from zero to 0.3-mm amplitude for the smaller gap group (approximately 1.0-mm gap; Fig. 5). The effect of amplitude, however, decreased as the gap became larger (to approximately 11% increase), suggesting that at large glottal gaps, perhaps in breathy voice, the amplitude of motion may not have a significant effect on the flow resistance.

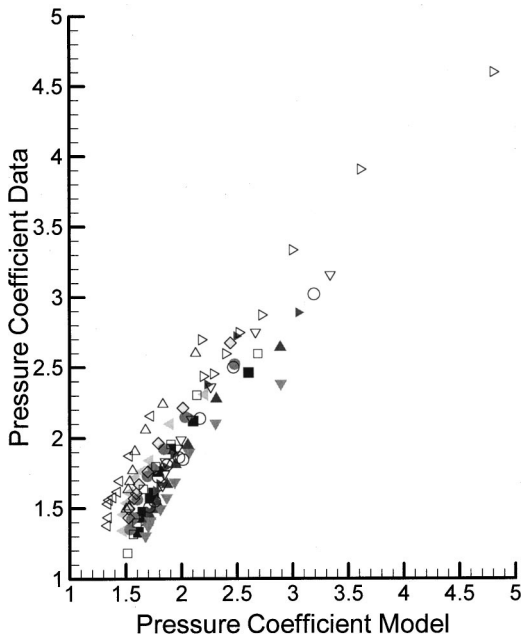


FIG. 11. Mean pressure coefficient as a function of a complex term that includes five other nondimensional terms. The data include all the amplitude and frequency data for the study. The correlation was $r=0.934$.

The nondimensional rendering of the data in Figs. 5 and 7 was given in Figs. 9 and 10. The pressure coefficient was displayed relative to the Strouhal number for variation in the amplitude (Fig. 9) and in frequency (Fig. 10). The separation of the data in these figures suggests that a greater sensitivity can be obtained to the amplitude and frequency variations studied here. Typically, the use of dimensionless variables of interest follows from a dimensional analysis of the situation being studied (Streeter, 1973), and can render a more concise presentation of the variables as they inter-relate to one another. Such an analysis yielded six dimensionless terms, the pressure coefficient (P^*), the Strouhal number (St), the Reynolds number (Re), and three distance variables, $G^*=G/Lg$, where G is the glottal gap and Lg is the longitudinal length of the glottis, $T^*=T/Lg$, where T is the axial length of the glottis, and $W^*=W/Lg$, where W is the amplitude of motion.

The dimensional analysis resulted in the following empirical equation relating the pressure coefficient to the other dimensionless terms:

$$P^*=\frac{10\ 570\mathrm{St}^{0.5}W^*T^*}{\mathrm{Re}^{0.5}G^*}+1.143 \quad (r=0.934). \tag{5}$$

Figure 11 shows this relation and collapses all of the data involving amplitude and frequency variation from this study. The expression defining the x axis values becomes a predictor term for variation in the pressure coefficient values. For example, it suggests that the pressure coefficient would rise with an increase in the Strouhal number (that is, with frequency), amplitude of motion of the glottis, or length of the axial duct, and would lower with an increase in the Reynolds number or average glottal gap. The prediction of the mean pressure drop would be obtained by multiplying both sides of the equation by $0.5\rho V^2$, and solving for values of the variables making up the complex expression in Eq. (5). It should be noted that this expression is for a hemilarynx model with a vibrating uniform glottal duct, and may differ for other models of phonation.

V. CONCLUSION

A hemilarynx model of phonation, where the vocal fold was driven sinusoidally toward and away from a flat plate, was used to examine mean transglottal pressure and mean flow characteristics as a function of amplitude and frequency of motion. The steady-flow results were similar to published work by others. The flow resistance through the model was strongly affected by the glottal gap size, being the primary determinant of resistance change in this study. The flow resistance was more affected by the amplitude of motion of the vocal fold at smaller glottal gaps. Flow resistance was not significantly affected by the frequency of oscillation. A dimensional analysis of the variables resulted in an empirical predictive equation that related the transglottal pressure coefficients to the other dimensionless terms, which included the Reynolds number, Strouhal number, and three other length terms involving the glottal gap, glottal amplitude, and axial length of the glottis.

ACKNOWLEDGMENTS

This work was supported by NIDCD Grant No. DC03566-03 from the National Institutes of Health. We would also like to thank the reviewers for their helpful comments and suggestions.

Alipour, F., Scherer, R. C., and Patel, V. C. (**1995**). ''An experimental study of pulsatile flow in canine larynges,'' J. Fluids Eng. **117**, 577–581.

Alipour, F., Scherer, R. C., and Knowles, J. (**1996**). ''Velocity distributions in glottal models,'' J. Voice **10**(1), 50–58.

Alipour, F., and Scherer, R. C. (**2000a**). ''Pressure-Flow Relationship in a Hemilarynx Mechanical Model,'' in *Proceedings of 5th Seminar on Speech Production: Models and Data* (Kloster Seeon, Bavaria, Germany), May 2000, pp. 189–192.

Alipour, F., and Scherer, R. C. (**2000b**). ''Dynamic glottal pressures in an excised hemilarynx model,'' J. Voice **14**, 443–454.

Binh, N., and Gauffin, J. (**1983**). ''Aerodynamic measurements in an enlarged static laryngeal model,'' *Technical Phoniatrics*, STL-QPSR, Department of Gas Dynamics, Royal Institute of Technology, Stockholm.

Doebelin, E. O. (**1990**). *Measurement Systems, Application and Design*, 4th ed. (McGraw-Hill, New York).

Edmonds, T. D., Lilly, D. J., and Hardy, J. C. (**1971**). ''Dynamic characteristics of air-pressure measuring systems used in speech research,'' J. Acoust. Soc. Am. **50**, 1051–1057.

Hirano, M., Kurita, S., and Matsuoka, H. (**1987**). ''Vocal function following hemilaryngectomy,'' Ann. Otol. Rhinol. Laryngol. **96**, 586–589.

Hirano, M., Kakita, Y., Kawasaki, H., Gould, W. J., and Lambiase, A. (**1981**). ''Data from high-speed motion picture studies,'' in *Vocal Fold Physiology*, edited by K. N. Stevens and M. Hirano (University of Tokyo Press, Tokyo, Japan), pp. 85–93.

Ikeda, T., Matsuzaki, Y., and Sasaki, T. (**1994**). ''Separated flow in a channel with an oscillating constriction (Numerical analysis and experiment),'' J. Jpn. Soc. Mech. Eng. (Japanese) **93**, 28–35.

Ishizaka, K., and Flanagan, J. L. (**1972**). ''Synthesis of voiced sounds from a two-mass model of the vocal cords,'' Bell Syst. Tech. J. **51**(6), 1233–1268.

Jiang, J. J., and Titze, I. R. (**1993**). ''A methodological study of hemilaryngeal phonation,'' Laryngoscope **103**, 872–882.

Jiang, J. J., and Titze, I. R. (**1994**). ''Measurement of vocal fold intraglottal pressure and impact stress,'' J. Voice **8**, 132–144.

McGowan, R. S. (**1993**). ''The quasisteady approximation in speech production,'' J. Acoust. Soc. Am. **94**, 3011–3013.
Mongeau, L., Franchek, N., Coker, C. H., and Kubli, R. A. (**1997**). ''Characteristics of a pulsating jet through a small modulated orifice, with application to voice production,'' J. Acoust. Soc. Am. **102**, 1121–1133.
Scherer, R. C., and Guo, C. G. (**1990**). ''Laryngeal modeling: Translaryngeal pressure for a model with many glottal shapes,'' ICSLP Proceedings, *1990 International Conference on Spoken Language* [J. Acoust. Soc. Jpn. **1**, 3.1.1–3.1.4 (1990)].
Scherer, R. C., Titze, I. R., and Curtis, J. F. (**1983**). ''Pressure-flow relationships in two models of the larynx having rectangular glottal shapes,'' J. Acoust. Soc. Am. **73**, 668–676.
Shadle, C. H., Barney, A. M., and Thomas, D. W. (**1991**). ''An investigation into the acoustics and aerodynamics of the larynx,'' in *Vocal Fold Physiology*, edited by J. Gauffin and B. Hammarberg (Singular, San Diego), pp. 73–81.
Streeter, V. L. (**1973**). *Fluid Mechanics*, 4th ed. (McGraw-Hill, Tokyo).
Teager, H. M., and Teager, S. M. (**1983**). ''The effects of separated air flow on vocalization,'' in *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, edited by D. M. Bless and J. H. Abbs (College Hill, Boston), pp. 124–143.
van den Berg, J. W., Zantema, J. T., and Doornenbal, Jr., P. (**1957**). ''On the air and the Bernoulli effect of the human larynx,'' J. Acoust. Soc. Am. **29**, 626–631.

An inverse dynamics approach to face animation

Michel Pitermann[a)]
Department of Psychology, Queen's University, Kingston, Ontario K7L 3N6, Canada

Kevin G. Munhall[b)]
Department of Psychology and Department of Otolaryngology, Queen's University, Kingston, Ontario K7L 3N6, Canada

(Received 31 August 2000; accepted for publication 3 May 2001)

Muscle-based models of the human face produce high quality animation but rely on recorded muscle activity signals or synthetic muscle signals that are often derived by trial and error. This paper presents a dynamic inversion of a muscle-based model (Lucero and Munhall, 1999) that permits the animation to be created from kinematic recordings of facial movements. Using a nonlinear optimizer (Powell's algorithm), the inversion produces a muscle activity set for seven muscles in the lower face that minimize the root mean square error between kinematic data recorded with OPTOTRAK and the corresponding nodes of the modeled facial mesh. This inverted muscle activity is then used to animate the facial model. In three tests of the inversion, strong correlations were observed for kinematics produced from synthetic muscle activity, for OPTOTRAK kinematics recorded from a talker for whom the facial model is morphologically adapted and finally for another talker with the model morphology adapted to a different individual. The correspondence between the animation kinematics and the three-dimensional OPTOTRAK data are very good and the animation is of high quality. Because the kinematic to electromyography (EMG) inversion is ill posed, there is no relation between the actual EMG and the inverted EMG. The overall redundancy of the motor system means that many different EMG patterns can produce the same kinematic output. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1391240]

PACS numbers: 43.70.Bk, 43.70.Jt, 43.71.Ma [AL]

I. INTRODUCTION

In recent years, there has been considerable commercial interest in face modeling for producing realistic animation in the motion picture and computer games industries, as well as for teleconferencing and multimedia educational purposes (see Parke and Waters, 1996, for a review). At the same time there has also been interest in facial animation as a research tool. In the study of speech motor control, models that take into account the geometrical, physiological and biomechanical characteristics of the face and vocal tract permit tests of the form and complexity of neural control signals (Laboissière *et al.* 1996). In speech perception research, facial animation has been used for audiovisual stimulus generation (e.g., Cohen and Massaro, 1990, 1993). In this application, animation provides visual stimulus control that cannot be achieved with human actors.

Available animation techniques cover a broad spectrum including key framing, performance animation, physically based animation, and parametrized geometrical models (Parke and Waters, 1996). Each approach involves a trade-off between computational cost and realism. For example, simple key framing involves the interpolation between key poses or postures, and this requires far less computation than a muscle-based, physical model that includes representations of the tissue biomechanics and muscle physiology. However, the physical models may offer greater dynamic realism, because the biomechanics of skin tissue is simulated. Hence, subtle deformations and motions of the facial surface may be more accurately reproduced.

We have been pursuing a muscle approach (e.g., Lucero and Munhall, 1999) following the work of Waters and Terzopoulos (Terzopoulos and Waters, 1990; Waters and Terzopoulos, 1991). The model is composed of three components that are incorporated in a 3-D rendering of an individual talker's morphology. (1) A jaw that is modeled as a single degree of freedom hinge joint. The jaw is kinematically controlled from recorded data as in performance animation. (2) A muscle module that represents a subset of the facial musculature, including their geometry and physiology. The muscles are modeled using a standard Hill-type formulation that contains force generation due to the contractile element (a force depending on muscle length variation and velocity) and a static dependence of force on muscle length (Zajac, 1989; Winters, 1990). (3) A skin component that represents multiple layers of soft tissue with a deformable multilayered mesh.

The model is controlled through activations of the modeled muscles that generate forces deforming the attached modeled tissue. In a test of this physical modeling, Lucero and Munhall (1999) drove the animation with recorded intramuscular electromyographic (EMG) signals. The animation produced by these EMG signals was highly realistic and corresponded well with 3-D kinematic data recorded from the talker at the same time as EMG data acquisition (Lucero and Munhall, 1999).

a)Electronic mail: mpiter@psyc.queensu.ca
b)Electronic mail: munhallk@psyc.queensu.ca

While these results are promising, the use of the model for stimulus generation in audiovisual perception experiments is limited by its reliance on EMG signals. Recording high quality facial EMG signals requires invasive intramuscular techniques and complicated experimental procedures. Acquiring good signals from all of the many muscles of the face would be difficult if not impossible. Further, intramuscular EMG recordings such as the ones used in Lucero and Munhall (1999) are far from perfect measures of the full muscle activation and force generation. Such problems as recording noise in the signals, movement artifact, interdigitation of the muscles fibers potentially leading to recordings from multiple muscles at any single recording site (Blair and Smith, 1986), and nonlinearities between EMG and force generation can potentially corrupt the measured muscle activation patterns. Thus, in the long run it seems impractical to depend on recorded EMG signals as the basis for animation control.

Two alternative control schemes can be considered for our muscle-based face model. First, a higher-order command-level ''language'' could be developed that maps actions at a task level onto the muscle level (Saltzman, 1979). There are a number of complexities involved with accomplishing this and few formal attempts have been made to do this for speech motor control. Saltzman and Munhall (1989) proposed a task-level scheme for the control of constrictions in a midsagittal vocal tract, however, this was a purely kinematic model with no mass or physiology modeled for the articulators. Ostry and his colleagues, on the other hand, have implemented a version of the equilibrium point model for the jaw (Laboissière *et al.*, 1996) and the tongue–jaw complex (Sanguineti *et al.*, 1998). In these models, commands at the level of the degrees of freedom of the articulator produce activation patterns across a set of modeled muscles that result in the desired kinematic patterns. To produce fluent speech both approaches would require the development of an additional level that encodes the sequential dynamics of articulation. Implementing such a scheme for a 3-D facial model with dozens of muscles, however, would be a daunting task. A second alternative is to drive the model kinematically by inverting the motion of a talker's face and computing the EMG signal and forces required by the model to produce this motion. It is this inversion approach that is the focus of this paper.

In human speech motor control, there is redundancy in both the articulatory and the neuromuscular systems, which means that there are many potential motor solutions for a given intention. This redundancy gives rise to a range of ill-posed problems for which it is difficult to arrive at unique solutions. Inversions (kinematic, dynamic, etc.) fall into this class of ill-posed problems and there is little agreement on how or whether the nervous system performs these inversions.

For example, Flash (1990) has suggested that a form of equilibrium control obviates the need for the nervous system to invert the planned trajectory. On the other hand, there are a number of proposals in the robotics and motor control literature for constraining inversions and thus making them computationally tractable [e.g., use of an objective function or performance index such as smoothness, use of a hierarchical control strategy, etc. See Kawato (1996) and Jordan and Rosenbaum (1989) for reviews].

In animation work, several kinematic-to-muscle inversions have been tested. For example, a static inversion was implemented in Terzopoulos and Waters (1993) to estimate muscle activity from single video frames. Energy minimizing splines (snakes; Kass *et al.*, 1987) were used to track features of the face, then muscle activity corresponding to the facial contours tracked by the snakes was found. Although the method produced interesting results, the snake technique was essentially a static mapping between facial configurations in a single frame and a muscle activation equilibrium that could produce that configuration. Further, it relied on facial contour detection, which is noisy and may not optimally parametrize the face (see the discussion). Morishima *et al.* (1998) used a neural network approach to compute the correspondence between static expressions (speech and emotion) measured by optical flow or optically tracked markers attached to the face. As in Terzopoulos and Waters (1993) the inverted EMG was used to drive a physical model.

The approaches taken by both Morishima *et al.* (1998) and Terzopoulos and Waters (1993) share common challenges. The head motion and 3-D kinematics of the face are only approximately corrected for. This can lead to aberrant face movements stemming from head motion accounted for face movements and from inaccurate input used in the inversion. Further, both approaches do not take advantage of the inherent dynamics of facial motion. A more comprehensive approach to mapping muscle activity to kinematics has been carried out by researchers at ATR Laboratories (Kyoto, Japan; e.g., Yehia *et al.*, 1998; Kuratate *et al.*, 1999).

As part of a general research program to study the relation between various correlates of speech production (acoustic, EMG, facial kinematics, head motion), the linear and nonlinear mappings between pairs of variables have been studied (Kuratate *et al.*, 1999). The estimation of 3-D facial motion components from EMG was good for both linear and nonlinear approaches, although the stability of the nonlinear approach over time was an issue. In addition, the animation model driven by these mappings was purely statistical and contained no physiological constraints.

Our approach shares the physical modeling of Terzopoulos and Waters (1993) and Morishima *et al.* (1998) and uses precise 3-D tracking of the face and head, as in work by the ATR group. In comparison to Morishima *et al.* (1998), our approach uses a classical nonlinear optimizer that does not need a training phase. In addition, our approach is truly a dynamic inversion and is thus constrained by motions and forces generated in the model over time. Our aim in the present research is to create realistic animation based on 3-D kinematic recordings. The use of muscle-based animation is preferred for its realism, however, this choice necessitates EMG signals as input. The current inversion permits the creation of naturalistic animation sequences from a non invasive kinematic recording procedure.

We present in this article a dynamic inversion based on a classical nonlinear optimizer called Powell's algorithm (Press *et al.*, 1992, Sec. 10.5). The optimizer looks for a set of modeled muscle activities minimizing the Euclidean distance between three-dimensional positions of markers attached to a talker's face and the corresponding nodes of the face model. Three experiments were carried out to evaluate the inversion. The goal of the first experiment was to test the model with controlled synthetic data. The modeled muscles were activated by a sawtooth EMG signal, then the resulting face movements were used to estimate a new set of muscle activities by means of the inversion. Standard Pearson correlations between inverted and synthetic muscle activity were analyzed. The inverted muscle activity was thereafter used to synthesize a second set of face movements, and the two animations were compared by means of correlation analyses. Our purpose for the second experiment was to test the inversion on real speech production. EMG and facial kinematic data from Lucero and Munhall (1999) were used. OPTOTRAK markers on a talker's face and EMG data were simultaneously collected while the talker produced an English sentence; the OPTOTRAK is an electronic movement tracking device, its stated 3-D resolution at 2.5 m distance is 0.01 mm (Vatikiotis-Bateson *et al.*, 1993). The motion measured by this tracking system was used in the inversion (estimate inverted muscle activity), and standard Pearson correlations between the recorded EMG and inverted muscle activity were computed. An animation was produced from the inverted muscle activity, and the motions of the OPTOTRAK markers and the corresponding nodes of the face model were compared by means of correlation analyses. A third experiment was carried out to test if the face model could be driven by a different talker's face motion without any face morphology model adaptation other than a global head-size scaling. The kinematics of OPTOTRAK markers attached to a new talker's face were tracked over time during syllable production. An inversion was carried out from the OPTOTRAK marker motions, and animation was produced from the inverted muscle activity. The correlations between the motions of the OPTOTRAK markers and the corresponding nodes of the face model were analyzed. Finally, the root mean square (rms) distance between the OPTOTRAK markers and corresponding nodes of the model was compared to the standard deviation of node positions around their mean position.

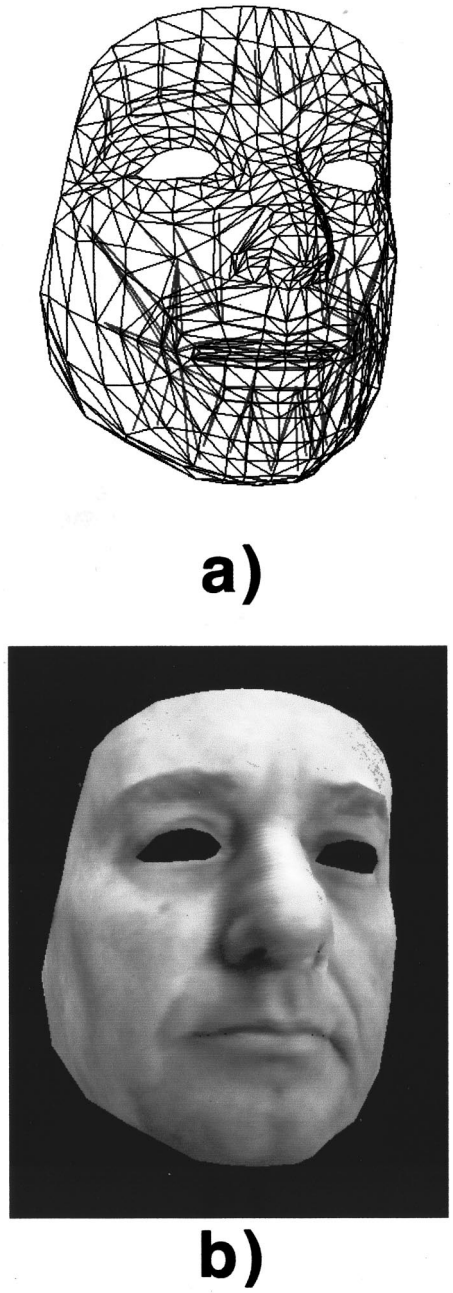


FIG. 1. (a) The epidermal mesh in black thin lines and the lines of action of the muscles in gray thick lines. (b) The superimposed texture map. The face model was adapted to the same speaker for both parts of the figure.

II. METHOD

A. The model

As noted above, the facial tissue is modeled as a multilayered mesh with isotropic mechanical characteristics. The nodes in the mesh are point masses, and each segment connecting a pair of nodes is a damped spring. The nodes are arranged in three layers representing the structure of facial tissues. The top layer corresponds to the epidermis, the middle layer represents the fascia, and the bottom layer models the skull surface. The elements between the top and middle layers represent the dermal-fatty tissues, and the elements between the middle and bottom layer represent the muscle tissues. The skull nodes are fixed in the three-dimensional space. A piecewise linear, biphasic approximation is used for the dermalfatty spring force elongation, and a linear approximation is used for all other spring force elongation. A nonlinear approximation is used for spring force compression to provide an infinite growth of the force as a spring length tends to zero. Figure 1(a) shows the mesh adapted to a talker's morphology and indicates the lines of action of the muscles. Figure 1(b) shows the superimposed texture map for the talker.

The generation of muscle force was computed by using rectified and integrated EMG as a measure of activity. A graded force development of the muscle force M was simu-

lated by a second-order low-pass filtering of this EMG signal, according to the equation

$$\tau^2 \ddot{M} + 2\tau \dot{M} + M = \overline{M}, \tag{1}$$

where $\tau = 15$ ms and $\overline{M}$ is the integrated EMG (Laboissière *et al.*, 1996). We will use *filtered EMG* to refer to the filtered, rectified, and integrated EMG in the rest of the article.

The equation of motion of each node i of the model had the general expression (Terzopoulos and Waters, 1990; Lee *et al.*, 1995; Lucero and Munhall, 1999):

$$m \frac{d^2 x_i}{dt^2} + r \sum_j \left(\frac{dx_i}{dt} - \frac{dx_j}{dt} \right) + \sum_j g_{ij} + \sum_e q_i^e + s_i + h_i = F_i, \tag{2}$$

where x_i was the current position of node i, m was the node mass equal to 0.000 23 kg for all nodes, the second term was the total damping force acting on the node i (x_j represented the nodes connected to node i and r was a constant equal to 0.050 kg/s), g_{ij} was the spring force applied by node j on node i, the fourth term modeled the skin incompressibility (q_i^e represented the triangular prism elements containing node i), the fifth term s_i was the skull reaction to the force applied by the fascia nodes, the sixth term h_i was a nodal restoration force applied to the fascia nodes connected to the skull, F_i was the total muscle force applied to node i.

B. Physiological measurements and model commands

The common characteristics of the three experiments are described here while the unique aspects of the experiments will be outlined in separate sections.

The face model had been adapted to a single subject's morphology for Lucero and Munhall (1999) using data from a Cyberware laser scanner (Lee *et al.*, 1993, 1995). This morphology was used in our three experiments in order to ease the comparison between results and in order to use the physiological data collected for Lucero and Munhall (1999).

In order to use EMG data collected for Lucero and Munhall (1999), and in order to compare the results of our three experiments, the face model was also controlled in the present work in the same manner as in Lucero and Munhall (1999). The model was symmetrically controlled by eight pairs of muscles, one muscle of each pair on each side of the face. They were the levator labii superior, levator anguli oris, zygomatic major, depressor anguli oris, depressor labii inferior, mentalis, orbicularis oris superior, and orbicularis oris inferior. The pair levator anguli oris/zygomatic major could not be reliably distinguished for the EMG measurements, hence these muscles were driven in the model by the same activation. This left seven degrees of freedom in the control space. The black circles of Fig. 2 show the approximate positions of the seven EMG electrode insertion points.

The sampling rate of the OPTOTRAK data used in Experiments 2 and 3 was 60 Hz. The facial movement data in both experiments were corrected for motion of the head by transforming the data to a coordinate system in which the origin is the incisor cusp and the horizontal and protrusion axes lie along the bite surface (Ramsay *et al.*, 1996). The number of markers and their positions on the talkers' faces were not the same for all experiments, and they will be described in the following individual methods.

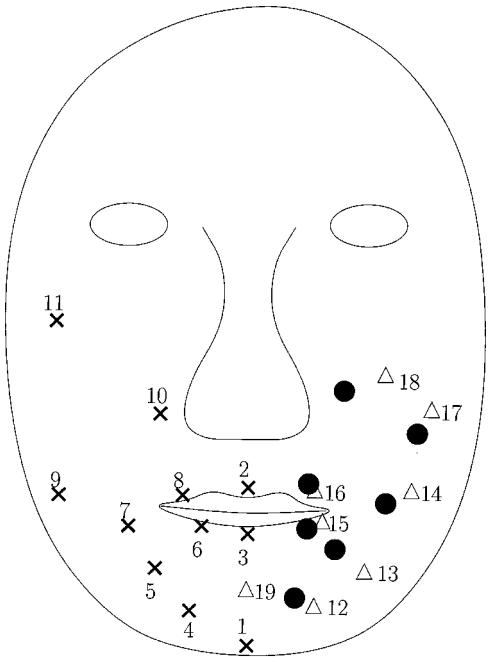


FIG. 2. Positions of OPTOTRAK markers (crosses), EMG electrode insertion points (filled circle), and face model nodes (triangle).

C. Inversion technique

The principle of the inversion is to continuously update the muscle activity to produce a face movement following a given face trajectory. To follow the face trajectory, the inversion minimizes the Euclidean distance between OPTOTRAK markers and the corresponding nodes of the model. Given the mass positions and velocities and given the muscle activity that brought the face model into a state corresponding to a frame, the inversion finds a new muscle activity for which the solution of the differential equation (2) brings the masses in one 1/60th of second to the position corresponding to the next frame. All calculations are based on the physics of the model, including the node masses, velocities, and muscle forces. As a consequence, modeled skin inertia partly determines how muscle activity has to be modified to bring the face model from one position to the next one. This is fundamentally different from inversion techniques matching each OPTOTRAK position to a facial expression at equilibrium. Our inversion is a truly dynamic inversion matching the dynamics of the face model to a kinematic pattern.

As in all nonlinear iterative algorithms, the inversion needs a starting set of values for the muscle activity for each frame, then updates the set until convergence is achieved. The muscle activity estimated for a frame was used as the seed of the next one. The resting position (no muscle activity) was always used as the seed of the first frame. When the inversion had been carried out for all frames, the inverted muscle activity was used to generate an animation.

A conventional nonlinear optimizer minimizing a cost function was selected to implement the inversion. The cost function E was the sum of the squares of the Euclidean distances between the OPTOTRAK markers and the corresponding face model nodes:

$$E=\sum_{i=1}^{N} |m_i-n_i|^2, \tag{3}$$

where m_i and n_i are the 3-D positions of the ith OPTOTRAK marker and face model node, respectively, N is the number of nodes used in the inversion, and $|\ |^2$ is the vectorial magnitude square operator, i.e., the sum of the squares of each coordinate of the vector. The optimizer minimizing the cost function was Powell's algorithm (Press *et al.*, 1992, Sec. 10.5) The algorithm searched a set of seven special orthogonal directions in the seven-dimensional control parameter space driving 16 muscles. The constraint in the selection of those directions was that a minimization along each of them would not influence the minimizations carried out along the six other directions. As a consequence, once the set had been found a simple one-dimensional minimization algorithm could be used sequentially along each direction.

The inversion could produce different muscle activity patterns, depending on the initial conditions. Constraints may be added to the inversion to limit the number of solutions. In all analyses, the inversion was carried out without constraints; then with the constraint that the inverted filtered EMG values had to be positive. The new positive constraint cost function E' was redefined in the second case by

$$E'=\begin{cases} \Sigma_{i=1}^{N}|m_i-n_i|^2, & \text{if all filtered EMG}>0, \\ 10^6(1+|\Sigma\text{EMG0}|), & \text{if at least one filtered EMG}<0, \end{cases} \tag{4}$$

where m_i and n_i are the 3-D positions of the ith OPTOTRAK marker and face model node, respectively, N is the number of nodes used in the inversion, and EMG0 is the set of negative muscle activity levels. The constraint that all filtered EMG had to be greater than zero will be called the *positive constraint* in the rest of this article.

For all inversions, $\sqrt{E/N}$ and $\sqrt{E'/N}$ were calculated over time to estimate for each frame the rms of the distances between the OPTOTRAK markers and their corresponding nodes.

To compare the 3-D time series of the OPTOTRAK markers and of the face model nodes, we generalized a few 1-D statistical features to three dimensions. The mean position of a 3-D node trajectory v composed of n samples (x_i,y_i,z_i) was its centroid, i.e., a point μ_v for which each coordinate was the arithmetic mean of the corresponding coordinate values of all samples of the time series:

$$\mu_v=\left(\frac{1}{n}\sum_{i=1}^{n} x_i, \frac{1}{n}\sum_{i=1}^{n} y_i, \frac{1}{n}\sum_{i=1}^{n} z_i\right). \tag{5}$$

The standard deviation σ_v of a 3-D node trajectory v was estimated by

$$\sigma_v=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n} |v_i-\mu_v|^2}, \tag{6}$$

where $|\ |^2$ is the vectorial square magnitude operator. The 3-D correlation ρ_{vw} between two node trajectories v and w composed of n samples v_i and n samples w_i was

$$\rho_{vw}=\frac{1/n\ \Sigma_{i=1}^{n} v_i\cdot w_i-\mu_v\cdot\mu_w}{\sigma_v\sigma_w}, \tag{7}$$

where $v_1\cdot v_2$ is the dot product between vectors v_1 and v_2. Like a 1-D correlation, ρ_{vw} always belongs to interval $[-1, 1]$.

III. EXPERIMENT 1

Two spaces are involved in our inversion: the kinematic space and the muscle activity space. The purpose of the first experiment was to analyze how consistently we could go and come back from one space to the other. We therefore worked with synthetic data.

A. Method

The 16 selected muscles were synchronously activated by a wave of triangles composed of the sequence (0, 1/6, 2/6, 3/6, 4/6, 5/6, 6/6, 5/6, 4/6, 3/6, 2/6, 1/6) repeated 3 times to create a 36-sample time series made of 3 identical triangular shapes. Then the same 11 nodes used in Lucero and Munhall (1999) were tracked over time. Their approximate positions are shown by the 11 crosses of Fig. 2. The 3-D time series of those 11 nodes were used to carry out the dynamic inversion. Then, standard Pearson correlations between the original and inverted muscle activity were calculated to compare the original and inverted EMG patterns.

Next, the inverted muscle activity was used to calculate a new animation. The 3-D standard Pearson correlations between the 11 nodes tracked during the first and the second animation were computed by means of Eq. (7) to compare the two kinematics.

Finally, we also calculated 3-D standard correlations between the two animations for eight nodes that were not used in the inversion. Their approximate positions are shown by the white triangles in Fig. 2. The correlation indicate how accurately a face movement can be reproduced using controlled simulated signals.

B. Results and discussion

Table I presents standard Pearson correlations between synthetic and inverted muscle activity. The first line of the table shows that the two types of muscle activity were poorly correlated. Only three standard Pearson correlations out of seven were significantly different from zero. This means that we cannot loop with forward calculations and inversion and find the same muscle activity. The second line of the same table shows that adding the positive constraint did not change fundamentally the results. A one-way analysis of variance of the correlations using the constraint as the factor was not significant at the 0.05 level $[F(1,12)=0.012;\ p=0.913]$, i.e., no significant difference in average correlation was found whether the positive constraint was used or not.

TABLE I. Correlations between synthetic and inverted muscle activity. The column labels ''DAO'' to ''OOS'' stand for Depressor Anguli Oris, Depressor Labii inferior, Levator Anguli Oris/zygomatic major, Levator Labii superior, Mentalis, Orbicularis Oris Inferior, and Orbicularis Oris Superior. The row labels ''No const.'' and ''EMG>0'' mean that no constraints or the positive one were used in the inversion, respectively. The correlations printed in bold were significantly different from 0 at 0.05 level according to a two-tail standard Pearson sample correlation test with 34 degrees of freedom.

	DAO	DL	LAO	LL	M	OOI	OOS
No const.	0.28	**0.43**	−0.02	0.14	**0.38**	**0.65**	0.14
EMG>0	**0.34**	**0.46**	−0.18	**0.34**	**0.41**	**0.54**	0.19

Table II shows 3-D correlations [Eq. (7)] between animation mesh node movements resulting from synthetic muscle activations and node movements resulting from inverted muscle activations. The top half of the table shows correlations for the 11 nodes used in the inversion, the bottom half of the table contains the results for eight nodes unused in the inversion. The correlations were always greater than 0.7 in all but one case out of 38, greater than 0.8 in 33 cases out of 38, and greater than 0.9 in 16 cases out of 38. This shows a very good match between the two animations.

To analyze if the movements of the nodes used in the inversion were better reconstructed than the movements of the other nodes, and to test whether using the positive constraint in the inversion led to different results, a two-way analysis of variance was carried out. The two factors were ''node used or unused in the inversion'' and ''positive constraint used or not in the inversion.'' The two factors and their interaction were not significant [$F(1,34)=2.57$, $p=0.118$ for the node factor; $F(1,34)=1.13$, $p=0.296$ for the constraint factor; and $F(1,34)=0.053$, $p=0.820$ for the interaction]. This demonstrates that the movements of parts of the face that were not used in the inversion were as well reconstructed as those used in the inversion.

The 11 nodes used for the inversion belonged to the right-hand side of the face model. The eight nodes used to test the face reconstruction belonged to the other side. Despite symmetric control of the face muscles, the animations were slightly asymmetric because the talker's face and the adapted mesh were asymmetric. The kinematics of the left half of the face were, nevertheless, as well reproduced as the kinematics of the right half of the face, even though the left half was not used in the inversion. This is an important result since it suggests that the physiological constraints of the model are accurately mimicking facial tissue dynamics.

Table III summarizes the distribution of the rms distance between the OPTOTRAK markers and the corresponding nodes for the stimuli used in the three experiments. The data give an indication of the average error made by the method reconstructing a node position. For each node time series, the table presents its minimum, its first quartile, its median, its third quartile, and its maximum. The 3-D standard deviation of each node trajectory was computed by means of Eq. (6). The last column of Table III contains the double of the rms of the standard deviations of the nodes used in the inversion for each stimulus, i.e., the nodes used to compute the other columns of the table. This is an estimate of the average movement amplitude of the nodes. This can be compared to the average error made by the method reconstructing a node movement.

The top part of Table III shows that the rms distance between the OPTOTRAK markers and the corresponding nodes of the face model was small for the synthetic stimuli. The rms distance was generally close to 0.3 mm and never reached 0.5 or 0.7 mm during the whole simulation when no constraints were used or when the positive constraint was added to the inversion, respectively. As can be seen, the reconstruction error was always smaller than the average movement amplitude of the nodes for the synthetic data.

To summarize the results so far, carrying out an inversion without constraints will not lead to the original set of muscle activities because many possible muscle activity patterns can lead to the same kinematics. Adding the positive constraint does not lead to the original EMG dataset either. Conversely, a face movement generated by the model can easily be reproduced by means of our ''inversion-resynthesis'' method when data for only a small set of nodes are available (e.g., a set of 11 nodes covering only half of the face). The next question is ''Would it be possible to replicate face movements produced by a real talker?''

TABLE II. The 3-D correlations [Eq. (7)] between node movements resulting from synthetic and inverted modeled muscle activity. The approximate positions of the nodes on the face can be seen in Fig. 2. The row labels ''No const.'' and ''EMG>0'' stand for ''No constraints used in the inversion'' and ''positive constraint used in the inversion.''

Nodes used in the inversion											
Node #	1	2	3	4	5	6	7	8	9	10	11
No const.	0.92	0.90	0.95	0.87	0.88	0.92	0.93	0.84	0.91	0.82	0.90
EMG>0	0.92	0.87	0.95	0.88	0.89	0.93	0.92	0.61	0.90	0.75	0.89

Nodes unused in the inversion								
Node #	12	13	14	15	16	17	18	19
No const.	0.84	0.73	0.85	0.87	0.81	0.91	0.93	0.87
EMG>0	0.87	0.82	0.83	0.85	0.76	0.90	0.90	0.74

TABLE III. A summary of the distribution functions of the rms Euclidean distances between the OPTOTRAK markers and the corresponding node positions for all stimuli used in the experiments. The last column of the table also contains an estimate of the average movement amplitude of the nodes. This was estimated by the double of the rms of the 3-D standard deviation [Eq. (6)] of the node trajectories. This must be compared to the ''Median'' column of the table. The results are given in mm. The top part of the table is related to the synthetic stimuli of the first experiment, the middle part to the natural sentence ''Where are you going?'' of the second experiment, and the bottom part to the four monosyllables used in the third experiment. The column labels ''Min.'' to ''Mov.'' stand for ''Minimum,'' ''First Quartile,'' ''Median,'' ''Third Quartile,'' ''Maximum,'' and ''Movement mean.'' The row labels ''No const.'' and ''EMG>0'' mean that no constraints or the positive one were used in the inversion, respectively.

	Min.	1st quantile	Median	3rd quantile	Max.	Mov.
			Synthetic data			
No const.	0.079	0.213	0.293	0.347	0.490	2.024
EMG>0	0.131	0.262	0.339	0.456	0.663	2.024
			Measurements with face adaptation			
No const.	0.034	0.753	1.126	1.714	2.289	5.748
EMG>0	0.035	0.903	1.867	2.643	4.020	5.748
			Measurements without face adaptation			
[bæb]	0.008	0.569	0.922	1.712	3.176	6.580
[bɛb]	0.008	0.462	0.747	1.409	2.466	4.876
[dæd]	0.008	0.523	0.933	1.734	2.658	6.060
[dɛd]	0.008	0.453	0.772	1.601	2.664	5.784

IV. EXPERIMENT 2

The second step in our series of experiments was to test the method using recorded data. We had two goals in mind with this experiment. First, we wanted to know if recorded EMG data could be estimated from facial kinematics using the dynamic inversion. Second, we wanted to know how accurately the OPTOTRAK marker kinematics could be reproduced after an inversion-synthesis operation.

A. Method

EMG and OPTOTRAK data collected for Lucero and Munhall (1999) were used in this test. A native American English talker produced the sentence ''Where are you going?'' Muscle activity from the left part of the talker's face and 3-D positions of 11 OPTOTRAK markers attached on the right side of the talker's face were recorded simultaneously along with the speech signal. The black spots and the black crosses of Fig. 2 show the approximate positions of the 7 EMG electrode insertion points and of the 11 OPTOTRAK markers, respectively.

The OPTOTRAK data were used to carry out a dynamic inversion, and the standard Pearson correlations between the inverted muscle activity and the EMG data were calculated. Next, the inverted muscle activity was used to synthesize a new animation, and the 3-D standard Pearson correlations between the OPTOTRAK marker positions and the corresponding nodes of the face model were calculated by means of Eq. (7).

B. Results and discussion

Table IV shows correlations between EMG measurements and muscle activity inverted from OPTOTRAK data. None of these correlations were significantly different from zero at the 0.05 level according to a two-tail standard Pearson sample correlation test, even though the degrees of freedom were relatively large (133). This confirms that the inversion with or without the positive constraint could not be used to determine which muscle activity pattern lay behind the face movements.

Table V presents 3-D correlations [Eq. (7)] between the 11 OPTOTRAK markers and the corresponding node movements of the face model. As in the first experiment, the 3-D correlations were high, except for nodes 2 and 8 (upper lip) when the positive constraint was used in the inversion. A one-way (positive or no constraints in the inversion) analysis of variance of the 3-D correlations showed that the difference was not significant at the 0.05 level [$F(1,20)=1.62$; $p=0.217$].

Synthetic facial movements such as those in the first experiment might be more accurately reproduced than natural movements of a human talker. Possible recording errors in the real kinematics, movement limitations in the model, or differences in the power spectra in the model and face could all contribute to lower correlations between the resynthesized recorded movements and the actual kinematics. In the synthetic facial movements, the motions are obviously realizable by the model, but real articulation may not be to the same

TABLE IV. Correlations between EMG measurements and muscle activity estimated from OPTOTRAK data. The column labels ''DAO'' to ''OOS'' stand for Depressor Anguli Oris, Depressor Labii inferior, Levator Anguli Oris/zygomatic major, Levator Labii superior, Mentalis, Orbicularis Oris Inferior, and Orbicularis Oris Superior. The insertion points of the electrode used to measure the EMG data can be seen in Fig. 2. The row labels ''No const.'' and ''EMG>0'' mean that no constraints or the positive one were used in the inversion, respectively.

	DAO	DL	LAO	LL	M	OOI	OOS
No const.	−0.06	−0.16	−0.06	0.11	−0.01	−0.04	−0.04
EMG>0	0.08	−0.11	−0.03	0.08	−0.07	0.04	0.03

TABLE V. The 3-D correlations [Eq. (7)] between 11 OPTOTRAK marker trajectories and the corresponding node movements of the face model resulting from inverted EMG for the sentence ''Where are you going?'' The approximate positions of the nodes can be seen in Fig. 2.

	1	2	3	4	5	6	7	8	9	10	11
No const.	0.93	0.72	0.96	0.95	0.95	0.95	0.91	0.78	0.67	0.61	0.78
EMG>0	0.89	−0.13	0.92	0.87	0.92	0.86	0.80	0.50	0.61	0.78	0.80

extent. To examine this issue, we compared the 3-D correlations of Tables II and V for the nodes numbered from 1 to 11. Those nodes were the same in both experiments. A two-way (''synthetic versus OPTOTRAK data'' and ''positive constraint versus no constraints'' analysis of variance of the correlations did not reveal any significant difference at the 0.05 level [$F(1,40)=3.99$, $p=0.053$ for ''synthetic versus OPTOTRAK data;'' $F(1,40)=2.20$, $p=0.146$ for ''constraint presence;'' and $F(1,40)=0.886$, $p=0.352$ for the interaction]. This suggests that replicating a natural face movement with the face model using real OPTOTRAK measurements may be as precise as replicating a face movement originally produced by the face model.

To summarize, the results showed that the OPTOTRAK kinematics could be recovered by the inversion-synthesis procedure with very good accuracy. EMG measurements, however, could not be recovered from OPTOTRAK measurements by means of the inversion.

V. EXPERIMENT 3

For practical reasons (e.g., not all laboratories have access to a laser range finder such as Cyberware) and theoretical concerns (the study of facial motion independent of morphology), we wanted to invert the facial motion of one talker and animate the morphology of another talker. To be practical, the adaptation of the talker's morphology to the face model had to be simple and achievable with a 2-D image. In this test, we simply aligned key features of the face and model, and linearly scaled the model to the *x*, *y*, and *z* dimensions of the talker. To be theoretically interesting, the animation must preserve the talker's 3-D kinematics on the new facial morphology. This is tested with correlation analysis and estimation error analysis.

A. Method

A native Canadian English talker produced the monosyllables /bæb/, /bɛb/, /dæd/, and /dɛd/. He was asked to begin each utterance from the same closed mouth initial position. The 3-D positions of 22 OPTOTRAK markers were recorded during his speech production [Fig. 3(a)].

To determine to which node of the face model each OPTOTRAK marker corresponded, a picture of the surface layer of the mesh and a picture of the talker were overlaid in Photoshop (Adobe, San Jose, CA). The width and height of the mesh picture were manually adjusted to obtain a good match between the talker's face and mesh [Fig. 3(b)]. The closest node to the center of each OPTOTRAK marker was then selected for use in the inversion. The dots superimposed on the mesh of Fig. 3(c) show the selected nodes.

For each stimulus, the face model was roughly adapted to the dimensions of the talker's head at the stimulus beginning. The difference in position (do_x, do_y, do_z) between two OPTOTRAK markers was computed for the first frame of the stimulus, i.e., in the resting position of the talker. The same two OPTOTRAK markers were used for all adaptations [see the '*' signs in Fig. 3(c) showing the corresponding model nodes]. They were manually selected only on the basis of having a large distance between them for each dimension. The difference (dn_x, dn_y, dn_z) between the two corresponding model nodes was calculated for the model in its resting position. For each dimension, the OPTOTRAK data was rescaled by the ratio between node and marker distance, e.g., dn_x/do_x. Each dimension was thereby linearly rescaled by a different factor. Finally, the coordinate system of each OPTOTRAK marker was shifted to match its position with the corresponding model node's for the resting position.

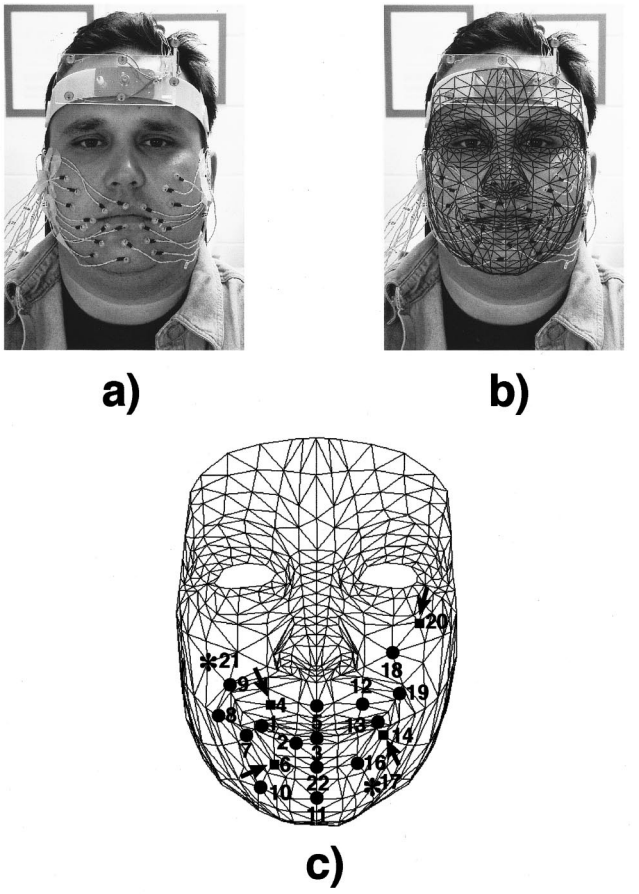


FIG. 3. (a) The talker in Experiment 3 with the 22 OPTOTRAK markers attached to his face. (b) Overlay of the talker's image and the mesh. The height and width of the mesh were manually adjusted to match the talker's head size. (c) The 22 nodes of the mesh selected to match the OPTOTRAK marker positions on the talker's face. The two '*' signs (17 and 21) were used to automatically adapt the mesh size to the talker's head size. The four squares (4, 6, 14, and 20) were not used in the inversion of Experiment 3, but were used to test the accuracy of the inversion-synthesis operation.

TABLE VI. The 3-D correlations between OPTOTRAK markers and corresponding node movements for syllables [bæb], [bɛb], [dæd], and [dɛd]. The approximate positions of the nodes can be seen in Fig. 3(c). The columns corresponding to the nodes numbered 4, 6, 14, and 20 are printed in bold because those nodes were not used in the dynamic inversion.

Node #	1	2	3	**4**	5	**6**	7	8	9	10	11
[bæb]	0.70	0.94	0.93	**0.78**	0.41	**0.89**	0.87	0.94	0.90	0.90	0.89
[bɛb]	0.66	0.93	0.92	**0.75**	0.35	**0.88**	0.87	0.89	0.86	0.86	0.87
[dæd]	0.66	0.96	0.97	**0.81**	−0.07	**0.82**	0.77	0.95	0.92	0.93	0.93
[dɛd]	0.62	0.96	0.96	**0.77**	0.02	**0.85**	0.76	0.94	0.93	0.92	0.93

Node #	12	13	**14**	15	16	17	18	19	**20**	21	22
[bæb]	0.87	0.81	**0.88**	0.94	0.95	0.86	0.79	0.87	**0.57**	0.81	0.90
[bɛb]	0.79	0.82	**0.85**	0.94	0.93	0.80	0.73	0.84	**0.53**	0.66	0.92
[dæd]	0.83	0.79	**0.75**	0.97	0.93	0.91	0.85	0.82	**0.66**	0.88	0.89
[dɛd]	0.83	0.74	**0.72**	0.97	0.92	0.92	0.84	0.83	**0.67**	0.93	0.89

Four OPTOTRAK markers [squares in Fig. 3(c)] were omitted from the inversion to serve as test data. The inversion was thereby carried out using the 3-D time series of 18 OPTOTRAK markers. Subsequently, an animation was produced from the inverted muscle activity. The 3-D standard Pearson correlations between OPTOTRAK marker measurements and face model movements were calculated by means of Eq. (7).

B. Results

The inversion process always diverged when no constraints were added. Several negative muscle activities were selected by the algorithm, then the method increased the absolute values of the negative levels of activity (they became more negative over time) until the cost function increased to an unacceptable level. As a consequence, only the results produced by the inversion using the positive constraint will be presented.

Table VI shows the 3-D correlations between the 22 OPTOTRAK markers and the corresponding node trajectories. All reconstructed node movements were highly correlated to the OPTOTRAK marker movements, with the exception of node 5 (the upper lip center). Note that the four nodes that were not used in the inversion (nodes 4, 6, 14, and 20) were also well correlated to their corresponding OPTOTRAK markers. The lower correlation for node 20 is likely related to its small movement amplitude. The average movement amplitude of that node was 1.204 mm, which was small in comparison to the average movement amplitude estimated across the whole face (see last column of Table III). Hence, the impact of node 20 on the animation was low.

To assess if the kinematics of the 18 nodes used in the inversion was more strongly related to the OPTOTRAK markers than the four reserved nodes, a two-way analysis of variance of the correlations was carried out. The two factors were ''node used versus unused in the inversion'' and ''stimulus'' ([bæb], [bɛb], [dæd], or [dɛd]). The null hypothesis was rejected at the 0.05 level in all cases [$F(1,80) = 1.89$, $p = 0.173$ for ''used versus unused;'' $F(3,30) = 0.140$, $p = 0.936$ for the stimulus and $F(3,80) = 0.006$, $p = 0.999$ for the interaction]. Thus, the reconstructed node movements were equally correlated for the four stimuli and the nodes unused in the inversion were as well correlated to the OPTOTRAK marker movements as the other nodes. This confirms the important results of the first experiment, suggesting that the whole facial surface can be synthesized from a sampling of position data.

In addition to the correlation analysis, the spatial error was estimated. As can be seen in the lower part of Table III, the rms distance between a node and the corresponding OPTOTRAK markers was usually smaller than 1 mm and always stayed below 3.2 mm. This indicates that the modeled movements were close to the real ones.

In this experiment, the morphology of the face model was not adapted to the talker's, unlike the previous experiments. In other words, one modeled face was driven by movements of another face. A one-way analysis of variance comparing the 22 correlations of the previous experiment (Table V) to the 88 correlations of the present experiment (Table VI) did not reveal any significant difference at the 0.05 level [$F(1,108) = 0.923$; $p = 0.339$]. Table III also shows that the rms distance between the OPTOTRAK markers and the corresponding nodes were not worse for the unadapted model than the adapted one.

VI. GENERAL DISCUSSION

In a series of tests, a dynamic inversion of facial kinematics has been successfully demonstrated. Using 3-D marker data as input, the inversion minimized the error between the model behavior and the recorded kinematics by varying activity in the modeled muscles of a physically based model of the face. Successful inversion-synthesis was demonstrated for synthetic model data, for EMG and kinematic data using a morphologically adapted animation model, and finally using kinematic data collected for a different subject than the facial model was morphologically adapted to. These accurate animations were achieved without reproducing the original EMG patterns. There was no correlation between the inverted and recorded EMG.

This inversion is important for use in perceptual research for a number of reasons. As demonstrated here, naturalistic animations can be produced by the approach and the facial kinematics in the animations are well characterized since they derive from actual kinematic data. As we have suggested before (e.g., Munhall and Tohkura, 1998), one of the current weaknesses in audiovisual speech research is that the visual stimuli are often poorly controlled and not well described. Since the animations in the present approach are

produced from kinematic data, a variety of experimental manipulations are feasible. Head motion and face motion are separated as part of the standard data processing and can be independently controlled in the animation (cf. Kuratate *et al.*, 1999). In addition, scalar manipulation of the kinematic amplitudes or time scales require only trivial manipulations of the kinematics prior to inversion. Finally, the ability to use the motions of one individual to drive the facial features of another individual permits a range of studies of identity and speech processing (cf. Nygaard and Pisoni, 1998).

When considered as a model of speech production, the inversion serves as a reminder of the computational complexity of motor control. Many muscle activity patterns can produce similar face movements, and the recorded EMG could not be estimated by means of our inversion procedure in its present state. This is not a surprising finding since, as noted above, the kinematic inversion is a mathematically ill-formed problem. To data, we have not explored the kinds of constraints that might make the problem tractable. Reducing the degrees of freedom (e.g., muscle synergy) and applying various cost functions (e.g., minimum jerk) are common suggestions in the motor control literature and these possibilities warrant further exploration in the context of this model.

One of the striking findings from the inversion was that the kinematics of markers that did not contribute to the inversion solution were reproduced as accurately as the marker data that served as input to the inversion. This suggests that the animation is spatially and temporally correct across a broad surface of the face, even when those regions of the face were not directly sampled in the inversion process. This behavior of the model is essential for its use in audiovisual perception research. A number of studies have indicated that when more of the face is shown, intelligibility increases (e.g., Le Goff *et al.*, 1997). Further, statistical studies of the relationship between facial kinematics and acoustics (Vatikiotis-Bateson *et al.*, 1996) have shown that even small motions on the periphery of the face contribute independent information about the acoustics.

In spite of its success, there are a number of issues about the inversion that will need to be the focus of ongoing research. The inversion constraint that was implemented (positive constraint) had little effect on the overall movement fit nor any effect on the correspondence of the inverted EMG to the synthetic or recorded EMG. However, as shown in Experiment 3, this constraint can be important in enabling the algorithm to reach a minimum. While the positive constraint is physiologically plausible, and perhaps more stable, there was evidence that the animation in some small regions of the face might be aided by negative EMG and the lengthening of the muscles that accompanies this signal. In the current lip muscle configuration, a protrusion of the most central upper lip node seems to have been reproduced better in the presence of negative EMG (see Tables II and V). At present, it is not clear whether modifications to the lip muscle geometry or the use of some other cost function in the inversion is the best solution to this effect. The movements of this particular portion of the lip are small so there was minimal influence on the overall animation.

The adaptation of the animation for use with a new talker (Exp. 3) and the matching of OPTOTRAK markers to nodes in the mesh in all of the experiments were simplistic, albeit effective. The influence of error in this phase of the inversion is, at present, unknown and will require 3-D imaging of the talkers with and without the OPTOTRAK markers attached. Also, it is unknown how a talker's and a face model's morphology can differ before the inversion diverges. We need a recording of more subjects to address this issue. In addition, it is unknown at this point, what the optimal number and placement of OPTOTRAK markers is. Resolving this problem will require a better understanding of the degrees of freedom of the face during speech production. Studies of the principal components of static lip shape (Linker, 1982) and the principal components of lip kinematics (Ramsay *et al.*, 1996) show a small number of modes of variation during speech. In Linker's data, the English vowels can be distinguished with a single measure, horizontal opening, while the Cantonese vowels required two factors and Finnish, Swedish, and French vowels three factors. Ramsay *et al.* (1996) calculated the principal components of lip motion in English for the 3-D motion of markers positioned around the oral aperture. In this data, the motion of any single position marker on the lip was strongly one-dimensional. When point-light facial displays are used to study audiovisual speech, the number and placement of lights is also an issue. Rosenblum *et al.* (1996) have manipulated the number and location of lights and shown enhancement of speech perception in noise with more lights. However, the necessary and sufficient number of markers needed to optimize point-light perception and the inversion is not known.

In spite of these unknowns, the success of the animation produced by the dynamic inversion is testament to the advantages of physically based animation. The underlying differential equations of the model provide a unitary description of the shape and motion of the human face and its gestures (Terzopoulos and Fleischer, 1988). The animation that is generated by the numerical solution of these equations is realistic across the full facial surface. The ability to drive the model with kinematic data that the current inversion provides makes this an attractive approach for stimulus generation.

ACKNOWLEDGMENTS

This research was supported by NIH Grant No. DC-00594 from the National Institute of Deafness and other Communication Disorders and NSERC. The authors thank D. J. K. Mewhort for the access to his computers funded by a NSERC equipment grant, and by an Academic Equipment Grant from Sun Microsystems of Canada. We are grateful to Anders Löfqvist for his helpful suggestions.

Blair, C., and Smith, A. (**1986**). "EMG recording in human lip muscles: can single muscles be isolated?," J. Speech Hear. Res. **29**, 256–266.

Cohen, M. M., and Massaro, D. W. (**1990**). "Synthesis of visible speech," Behav. Res. Methods Instrum. Comput. **22**, 260–263.

Cohen, M. M., and Massaro, D. W. (**1993**). "Modeling coarticulation in synthetic visual speech," in *Models and Techniques in Computer Animation*, edited by N. M. Thalmann and D. Thalmann (Springer-Verlag, Tokyo).

Flash, T. (**1990**). ''The organization of human arm trajectory control,'' in *Multiple Muscle Systems: Biomechanics and Movement Organization*, edited by J. M. Winters and S. L. Y. Woo (Springer-Verlag, New York).
Jordan, M. I., and Rosenbaum, D. A. (**1989**). ''Action,'' in *Foundations of Cognitive Science*, edited by M. I. Posner (MIT Press, Cambridge, MA), pp. 727–767.
Kass, M., Witkin, A., and Terzopoulos, D. (**1987**). ''Snakes: Active contour models,'' Int. J. Comput. Vis. **1**, 321–331.
Kawato, M. (**1996**). ''Trajectory formation in arm movements: minimization principles and procedures,'' in *Advances in Motor Learning and Control*, edited by H. N. Zelaznik (Human Kinetics, Illinois).
Kuratate, T., Munhall, K. G., Rubin, P. E., Vatikiotis-Bateson, E., and Yehia, H. (**1999**). ''Audio-visual synthesis of talking faces from speech production correlates,'' in *6th European Conference on Speech Communication and Technology (Eurospeech'99)* (International Speech Communication Association, Budapest, Hungary), Vol. 3, pp. 1279–1282.
Laboissière, R., Ostry, D. J., and Feldman, A. G. (**1996**). ''The control of multi-muscle systems: Human jaw and hyoid movements,'' Biol. Cybern. **74**, 373–384.
Le Goff, B., Guiard-Marigny, T., and Benoît, C. (**1997**). ''Analysis–synthesis and intelligibility of a talking face,'' in *Progress in Speech Synthesis*, edited by J. P. H. van Santen, R. Sproat, J. P. Olive, and J. Hirshberg (Springer-Verlag, New York).
Lee, Y., Terzopoulos, D., and Waters, K. (**1993**). ''Constructing physics-based facial models of individuals,'' in *Proceedings of Graphics Interface"93,"* Toronto, Ontario, Canada, pp. 1–8.
Lee, Y., Terzopoulos, D., and Waters, K. (**1995**). ''Realistic modeling for facial animation,'' in *Proceedings of SIGGRAPH'95*, Computer Graphics Proceedings, Annual Conference Series, Los Angeles, CA, pp. 55–62.
Linker, W. (**1982**). ''Articulatory and acoustic correlates of labial activity in vowels: A cross-linguistic study,'' UCLA Working Papers in Phonetics, Vol. 56, pp. 1–134.
Lucero, J. C., and Munhall, K. G. (**1999**). ''A model of facial biomechanics for speech production,'' J. Acoust. Soc. Am. **106**, 2834–2842.
Morishima, S., Ishikaw, T., and Terzopoulos, D. (**1998**). ''Model based 3D facial image reconstruction from frontal image using optical flow,'' in *ACM SIGGRAPH 98 Conference Abstracts and Applications*, Orlando, Florida, p. 258.
Munhall, K. G., and Tohkura, Y. (**1998**). ''Audiovisual gating and the time course of speech perception,'' J. Acoust. Soc. Am. **104**, 530–539.
Nygaard, L. C., and Pisoni, D. B. (**1998**). ''Talker-specific learning in speech perception,'' Percept. Psychophys. **60**, 355–376.
Parke, F. I., and Waters, K. (**1996**). *Computer Facial Animation* (A. K. Peter Ltd., Wellesley, MA).
Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (**1992**). *Numerical Recipes in C*, 2nd ed. (Cambridge University Press, Cambridge).
Ramsay, J., Munhall, K. G., Gracco, V., and Ostry, D. J. (**1996**). ''Functional data analysis of lip motion,'' J. Acoust. Soc. Am. **99**, 3718–3727.
Rosenblum, L. D., Johnson, J. A., and Saldaña, H. M. (**1996**). ''Point-light facial displays enhance comprehension of speech in noise,'' J. Speech Hear. Res. **39**, 1159–1170.
Saltzman, E. L. (**1979**). ''Levels of sensorimotor representation,'' J. Math. Psychol. **23**, 91–163.
Saltzman, E. L., and Munhall, K. G. (**1989**). ''A dynamical approach to gestural patterning in speech production,'' Ecological Psychol. **1**, 333–382.
Sanguineti, V., Laboissière, R., and Ostry, D. J. (**1998**). ''A dynamic biomechanical model for neural control of speech production,'' J. Acoust. Soc. Am. **103**, 1615–1627.
Terzopoulos, D., and Fleischer, K. (**1988**). ''Deformable models,'' Visual Comput. **4**, 306–331.
Terzopoulos, D., and Waters, K. (**1990**). ''Physically-based facial modeling, analysis, and animation,'' J. Vis. Comp. Anim. **1**, 73–80.
Terzopoulos, D., and Waters, K. (**1993**). ''Analysis and synthesis of facial image sequences using physical and anatomical models,'' IEEE Trans. Pattern Anal. Mach. Intell. **15**, 569–579.
Vatikiotis-Bateson, E., Munhall, K. G., Kasahara, Y., Garcia, F., and Yehia, H. (**1996**). ''Characterizing audiovisual information during speech,'' in *Proceedings of International Conference on Spoken Language Processing '96* (University of Delaware and A. I. duPont Institute, Philadelphia), Vol. 3, pp. 1485–1488.
Vatikiotis-Bateson, E., Munhall, K. G., and Ostry, D. J. (**1993**). ''Optoelectronic measurement of orofacial motions during speech production,'' in *Measuring Speech Production*, edited by M. Stone (Acoustical Society of America).
Waters, K., and Terzopoulos, D. (**1991**). ''Modeling and animating faces using scanned data,'' J. Vis. Comp. Anim. **2**, 123–128.
Winters, J. M. (**1990**). ''Hill-based muscle models: A system engineering perspective,'' in *Multiple Muscle Systems: Biomechanics and Movement Organization*, edited by J. M. Winters and S. Woo (Springer-Verlag, London), pp. 69–93.
Yehia, H., Rubin, P. E., and Vatikiotis-Bateson, E. (**1998**). ''Quantitative association of vocal tract and facial behavior,'' Speech Commun. **26**, 23–44.
Zajac, F. E. (**1989**). ''Muscle and tendon: properties, models, scaling, and application to biomechanics and motor control,'' CRC Crit. Rev. Biomed. Eng. **17**, 359–411.

The acoustic features of human laughter

Jo-Anne Bachorowski[a] and Moria J. Smoski
Department of Psychology, Wilson Hall, Vanderbilt University, Nashville, Tennessee 37203

Michael J. Owren[b]
Department of Psychology, Uris Hall, Cornell University, Ithaca, New York 14853

(Received 18 January 2001; accepted for publication 13 June 2001)

Remarkably little is known about the acoustic features of laughter. Here, acoustic outcomes are reported for 1024 naturally produced laugh bouts recorded from 97 young adults as they watched funny video clips. Analyses focused on temporal features, production modes, source- and filter-related effects, and indexical cues to laugher sex and individual identity. Although a number of researchers have previously emphasized stereotypy in laughter, its acoustics were found now to be variable and complex. Among the variety of findings reported, evident diversity in production modes, remarkable variability in fundamental frequency characteristics, and consistent lack of articulation effects in supralaryngeal filtering are of particular interest. In addition, formant-related filtering effects were found to be disproportionately important as acoustic correlates of laugher sex and individual identity. These outcomes are examined in light of existing data concerning laugh acoustics, as well as a number of hypotheses and conjectures previously advanced about this species-typical vocal signal. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1391244]

PACS numbers: 43.70.Gr [AL]

I. INTRODUCTION

Laughter plays a ubiquitous role in human vocal communication, being frequently produced in diverse social circumstances throughout life. Surprisingly, rather little is currently known about the acoustics of this species-typical vocal signal. Although there has been an enduring view that some variation may occur among the individual sounds that constitute laughter, these components are predominantly conceptualized as being vowel-like bursts (e.g., Darwin, 1872/1998; Hall and Allin, 1897; Mowrer, LaPointe, and Case, 1987; Ruch, 1993; Nwokah *et al.*, 1999; cf. Ruch and Ekman, 2001). While there is thus some information available about the mean fundamental frequency (F_0) of voiced laugh segments, reports have been markedly inconsistent. For example, the mean F_0 of male laughs has been reported to be as low as 126 Hz (Mowrer *et al.*, 1987; also see Bickley and Hunnicutt, 1992), but also as high as 424 Hz (Rothgänger *et al.*, 1998). Likewise, values for females have included an improbably low estimate of 160 Hz (Milford, 1980) and a high of 502 Hz (Provine and Yong, 1991).

Provine (1996, 2000; Provine and Yong, 1991) in particular has emphasized laughter's harmonically rich, vowel-like structure, further arguing that while vowel quality can show marked variation among laugh bouts, it is highly consistent within a series. In other words, with the possible exception of variation in the first or last sounds of a bout, Provine maintains that laughers routinely produce aspirated sequences of either ''ha,'' ''he,'' or ''ho'' sounds in discrete bouts (we infer the phonetic transcriptions of ''ha'' to be either /ɑ/, /ə/, or /ʌ/, and ''he'' and ''ho'' to be /i/, and /o/, respectively; cf. Edmonson, 1987). Provine also argues that the formant structure of laughter is less prominent than that of speech vowel sounds, although in neither case have quantitative formant measurements been provided in support of these claims. Given that formant structure is apparent in the spectrographic example shown in several publications (e.g., Provine, 1996, 2000; Provine and Yong, 1991) and several researchers have extracted formant values from at least a small number of laughs (Milford, 1980; Bickley and Hunicutt, 1992), this issue warrants closer scrutiny.

In contrast to Provine's emphasis on vowel-like laughter, Grammer and Eibl-Eibesfeldt (1990) drew a basic distinction between ''vocalized'' and ''unvocalized'' laughter. This contrast evidently referred to the presence or absence of voicing, and proved to be functionally important in their work. For example, individual males, after interacting with an unfamiliar female partner for a brief interval, were more interested in seeing her again if she produced voiced but not unvoiced laughter during the encounter. The importance of this basic distinction was subsequently confirmed in perceptual studies, which showed that voiced laughter induces significantly more positive emotional responses in listeners than do unvoiced laughs (Bachorowski and Owren, 2001). The latter is nonetheless a common element of laugh repertoires (Bachorowski, Smoski, and Owren, 2001), which raises the question of the relative prevalence of voiced and unvoiced laughter as a basic issue in laugh acoustics.

Other investigators have also considered laughter to be a variable signal, both in the kinds of sounds produced (Hall and Allin, 1897) and in its acoustic features (Rothgänger *et al.*, 1998). Variability of this sort is largely at odds with perspectives that treat laughter as a stereotyped vocalization. As exemplified by the work of Provine (e.g., Provine, 1996) and Grammer (1990; Grammer and Eibl-Eibesfeldt, 1990; see also Deacon, 1997), this approach proposes that laughter is—or at least resembles—a fixed action pattern (FAP) specialized for communication through an evolutionary process of ''ritualization.'' The expected outcome of this process is

[a]Electronic mail: j.a.bachorowski@vanderbilt.edu
[b]Electronic mail: mjo@cornell.edu

constancy in the rate, intensity, and most importantly in the form of signal production.

The goal of the current work was to further investigate each of these issues. In so doing, we sought to improve on the number of subjects recorded, the number of laugh exemplars included for each, and the methods used in acoustic analysis. Ultimately, we examined 1024 bouts of laughter, representing every analyzable laugh sound recorded from 97 adult males and females while they watched humorous video clips presented in a comfortable laboratory setting. The resulting sample was thus significantly larger than in previous studies, which have for instance included 3 bouts from each of 3 adult females (Nwokah *et al.*, 1999), a total of 15 bouts from 1 male and 1 female (Bickley and Hunnicutt, 1992), 5 bouts produced from each of 11 males (Mowrer *et al.*, 1987), and one bout from each of 23 males and 28 females (Provine and Yong, 1991). Acoustic measures were designed to characterize temporal properties, source-energy characteristics, and spectral features of every sound, with additional attention paid to sex differences in the use of laughter as well as indexical cueing of laugher sex and individual laugher identity.

II. METHOD

A. Subjects

One hundred thirty-nine students enrolled at Vanderbilt University were recorded as they watched funny video clips either alone or as part of a same- or other-sex friend or stranger dyad. Volunteers were primarily recruited from a General Psychology course and received research credit toward that course. Participants solicited by a friend were typically paid $10 for their involvement, but could instead receive research credit if enrolled in General Psychology. Before testing, subjects provided oral and written consent to the procedures. As individuals were recorded without knowing that laughter was specifically of interest, consent to use laughter data was obtained after testing was complete.

Data collected from ten subjects were excluded because of equipment failure ($n=2$), experimenter error ($n=2$), illnesses that might affect laugh acoustics (e.g., strep throat, $n=2$,) or use of mood-altering prescription drugs (e.g., serotonin reuptake inhibitors, $n=4$). In 11 cases, data were not used because the individual was not a native American-English speaker or was tested with a partner whose native language was not English. Finally, data from 21 subjects were excluded because the three or less laughs produced during the 3.95-min film clip period were deemed too few for statistical analysis. The final sample included 45 males and 52 females who had a mean age of 19.23 years (s.d. $=1.13$) and were primarily white ($n=87$). However, the sample also included six blacks, three Asian Americans, and one Native American. None reported any speech- or hearing-related problems. Of these 97 individuals, 11 were tested alone, 24 with a same-sex friend, 21 with an other-sex friend, 20 with a same-sex stranger, and 21 with an other-sex stranger. Results concerning the use of laughter in these various social contexts are discussed elsewhere (Bachorowski *et al.*, 2001).

B. Stimuli and apparatus

Subjects all watched a total of 11 emotion-inducing film clips, two of which were included specifically for their positive-emotion and laugh-inducing potential (other clips elicited either sad, fearful, disgusted, or neutral emotional responses). The first was the 1.42-min ''bring out your dead'' segment from *Monty Python and the Holy Grail*, and the second was the 2.53-min ''fake orgasm'' scene from *When Harry Met Sally* (total time $=3.95$ min). Film clips were presented using a Panasonic AG-5700 video cassette recorder (VCR) located on a shelf next to a 31-in. Panasonic CT 31G10 television monitor. Both the monitor and VCR were housed in a large media center. An experimenter operated the VCR from the adjacent control room via a Panasonic AG-A570 editing device attached through a wall conduit.

Recordings were made using Audio-Technica Pro 8 headworn microphones (Stow, OH), which were connected through the conduit to separate inputs of an Applied Research Technology 254 preamplifier (Rochester, NY) located in the control room. Each signal was amplified by 20 dB and then recorded on separate channels of a Panasonic Professional SV-4100 digital audiotape (DAT) recorder (Los Angeles, CA). Recordings were made using BASF digital audiotapes (Mount Olive, NJ). Tandy Optimus LV-20 headphones (Fort Worth, TX) connected to the DAT recorder were used to monitor participants throughout testing, and the experimenter communicated with participants as necessary through a Tandy 43-227 intercom.

C. Design and procedure

Participants were tested in a large laboratory room furnished to resemble a comfortable den. After providing informed consent, participants were told that they would be rating the emotion-inducing impact of each of a series of short film clips and that their evaluations would be used to select stimuli for upcoming studies of emotional response processes. Thus, subjects were unaware that their laughter was the focus of the research. After seating participants in futon chairs placed 3.3 m in front of the television monitor, the experimenter helped each individual position the microphone approximately 2.5 cm in front of the labiomental groove, and explained that the film-clip ratings (not relevant here) would be audio recorded. Next, input levels were adjusted, participants were given the opportunity to ask questions, and were informed that they would be left on their own and should treat the experience as if watching videos in their own living room. At the end of the viewing session, the experimenter returned to the testing room, debriefed participants as to the nature of the study, and obtained consent to use all data.

D. Laugh selection, classification, and acoustic analysis

Laughter was defined as being any perceptibly audible sound that an ordinary person would characterize as a laugh if heard under everyday circumstances. While inclusive, this broad criterion was considered reasonable on several grounds. First, these sounds were produced while subjects

watched film clips selected for their likelihood of eliciting positive affect. Indeed, the clips were rated as producing positive emotional responses by virtually all participants. Second, although no restrictions were placed on talking during the film clips, subjects almost never did—thereby making it unlikely that the sounds they were making represented either linguistic or paralinguistic events. Finally, each sound was routinely heard dozens of times during the course of acoustic analysis, and questionable ones were removed from further consideration.

Borrowing terminology from acoustic primatology (e.g., Struhsaker, 1967; Owren, Seyfarth, and Cheney, 1997), laughs were analyzed at ''bout,'' ''call,'' and ''segment'' levels. Bouts were entire laugh episodes that are typically produced during one exhalation. Although many bouts ended with audible inhalations or exhalations, these sounds were not included in bout-level characterizations unless they were deemed to be critical to the laugh itself. Calls were the discrete acoustic events that together constitute a bout, and have elsewhere been referred to as ''notes'' or laugh ''syllables.'' Isolated calls that were difficult to distinguish from sighs or other nonlaugh vocalizations were excluded from analysis. Overall, however, any sound deemed integral to a laugh bout was considered to be a call. Segments were defined as temporally delimited spectrogram components that either visibly or audibly reflected a clear change in production mode occurring during the course of an otherwise continuous call.

Laughs were digitized at 50 kHz using Kay Elemetric's COMPUTERIZED SPEECH LAB (CSL; Lincoln Park, NJ). Acoustic analyses were conducted using ESPS/WAVES+ 5.2 digital signal-processing software (Entropic Research Lab, Washington, DC) implemented on a Silicon Graphics O2 unix-based processor with the Irix 6.3 operating system (SGI; Mountain View, CA). Preprocessing of files included format conversions on a personal computer using custom-written software programs by Tice and Carrell (available at http://hush.unl.edu/LabResources.html). Files were then downsampled to 11.025 kHz and normalized to a common maximum-amplitude value.

In preparation for automatic extraction of various acoustic measurements using unix-csh-script routines, each file was first segmented with cursor-based onset and offset marks for every bout, call, and segment. Each of these levels was then categorized as to type. At the bout level, laughs were assigned to one of three mutually exclusive types. Bouts consisting primarily of voiced sounds were considered ''song-like,'' and included comparatively stereotyped episodes of multiple vowel-like sounds with evident F_0 modulation as well as sounds that might best be described as giggles and chuckles. Bouts largely comprised of unvoiced calls with perceptually salient nasal-cavity turbulence were labeled ''snort-like.'' Acoustically noisy bouts produced with turbulence evidently arising in either the laryngeal or oral cavities were called ''unvoiced grunt-like'' sounds, and included breathy pants and harsher cackles. To assess the reliability of bout-level categorizations, a second research assistant independently labeled each bout. The obtained kappa coefficient of 0.92, $p<0.001$, indicated a high level of inter-rater agreement in bout-level classification.

Both bouts and individual calls were identified as either ''voiced,'' ''unvoiced,'' or ''mixed,'' and segments were labeled as being either voiced or unvoiced. Calls were further labeled according to whether the sound was perceived as being produced with the mouth open or closed. Inter-rater reliability for mouth-position judgments was high: a kappa coefficient of 0.91, $p<0.001$, was obtained for 329 calls from 100 randomly selected bouts that were each coded independently by two raters. Finally, calls and segments that showed evidence of non-normative, atypical source energy were also noted. These events included vocal fry, in which individual glottal pulses are perceptually discernible, as well as a number of nonlinear types (i.e., glottal whistles, subharmonics, and biphonation; see Wilden *et al.*, 1998).

Acoustic measurements focused on durations, F_0-related features, and spectral characteristics of bouts, calls, and segments. Durations were readily extracted from onset and offset markers, but because F_0 is routinely much higher in laughter than in speech, pitch-tracking algorithms designed for the latter did not always perform well. These analyses were therefore conducted at the call level by first using the ESPS/WAVES+ pitch-tracking routine to extract an F_0 contour for each sound, and then overlaying the resulting plot on a corresponding narrow-band spectrogram. If the algorithm failed, the first harmonic was manually enclosed both in time and frequency using cursor settings, and its frequency contour was extracted as a series of maximum-amplitude points occurring one per column in the underlying spectrogram (Owren and Casale, 1994).

Spectral measurements focused on formant frequencies, which were derived from smooth spectral envelopes produced through linear predictive coding (LPC). The measurement procedure included first producing both a narrow-band, FFT-based (40-ms Hanning window, 0.94 preemphasis factor, 512-point FFT, 2-ms step size) and a wideband, LPC-based (fast modified Burg method, 40-ms rectangular window, 0.94 preemphasis factor, 10 coefficients, 2-ms step size) spectrogram of each sound. One location was then designated within each call or segment based on these displays, selected so as to provide clear outcomes that were also representative of the sound as a whole (see Fig. 1). Setting the cursor in this location produced a display of both underlying spectral slices, with the LPC envelope overlaid on the FFT-based representation. Formant-peak locations were located through visual inspection, marked on the LPC function by setting the cursor, and automatically recovered from the associated data record. Formant measurements were not taken from unvoiced, snort-like sounds. Although their resonances were often consistent with normative values from nasal speech sounds, many of these calls also seemed to be affected by noisiness resulting from airstream interactions with the microphone element.

Estimates of supralaryngeal vocal-tract length (VTL) were derived from formant frequencies using the following equation (adapted from Lieberman and Blumstein, 1993):

$$\mathrm{VTL}=\frac{(2k+1)c}{4F_{k+1}},$$

where $k=(0,1,2)$, F_{k+1} is the frequency of the formant of

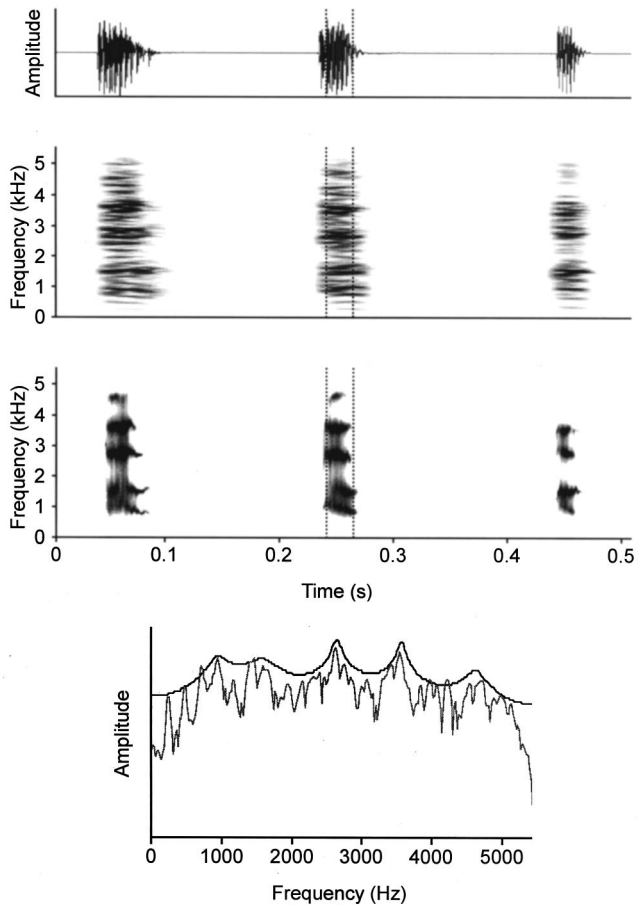


FIG. 1. Waveform (top) and corresponding narrow-band (second panel) and wideband (third panel) spectrograms of a voiced-laugh bout. Dotted vertical lines in the second of the three calls indicate the window from which spectral measurements were made. At the bottom, the smoothed LPC envelope is shown overlaid on the FFT-based representation.

interest, and c is the speed of sound (34 400 cm/s). Separate calculations were made for each of the five formants, and the mean of these estimates provided the VTL value used in classification analyses of laugher sex and individual identity.

III. RESULTS

A. Laugh types and durations

1. Bout-level descriptive outcomes

Descriptive outcomes associated with bout-level analyses are provided in Table Ia, and representative spectrograms of male and female voiced song-like, unvoiced grunt-like, and unvoiced snort-like bouts are shown in Fig. 2. Sample laughs can be heard at http://www.psy.vanderbilt.edu/faculty/bachorowski/laugh.htm. A total of 1024 laugh bouts was analyzed. Of these, 30% were predominantly voiced, 47.6% were mainly unvoiced, 21.8% were a mix of voiced and unvoiced components, and the remaining 0.7% were largely comprised of glottal whistles. Of the unvoiced bouts, 37.2% were grunt-like, whereas the remaining 62.8% were snort-like. This bout-level variability did not appear to be a matter of differences in individual laugher style. Many individuals (40.2%) produced all three of the most common bout types (i.e., voiced song-like, unvoiced snort-like, and unvoiced grunt-like), just as many produced two types (43.3%), while comparatively few (16.5%) produced just one types. Bouts that were either mixed or not readily classified were not included in further analysis of bout type.

Laugh bouts were highly variable in duration, with a standard deviation of 0.77 associated with the mean of 0.87 s. Outcomes of an analysis of variance (ANOVA) and Scheffé follow-up comparisons showed that a main effect of bout type, $F(2,933)=30.52$, $p<0.001$, was due to the shorter durations of snort-like rather than either song- or grunt-like bouts (see Table Ia). On average, males and females did not differ in the number of laughs produced, $F(1,96)=0.14$, ns. However, laugher sex did mediate the type of bout produced, $\chi^2(5)=137.26$, $p<0.001$. Follow-up binomial tests revealed that females produced significantly more voiced, song-like bouts than did males ($p<0.001$), whereas males produced significantly more unvoiced, grunt-like laughs than did females ($p<0.025$). There were no sex differences in the number of unvoiced, snort-like laughs produced. Laugher sex exerted a slight influence on bout duration, $F(1,935)=4.75$, $p<0.05$, with male laughs being a bit longer than female laughs.

2. Call-level descriptive outcomes

Descriptive outcomes associated with the corpus of 3479 calls are provided in Table Ib. On average, laugh bouts were comprised of 3.39 calls, but the associated standard deviation of 2.71 indicates that the number of calls per bout was highly variable. Most calls (45.2%) were unvoiced, but a notable proportion were either voiced (34.2%) or a mix of production modes (13.0%). In addition, 3.5% of the calls were essentially glottal pulses, 2.5% were produced in the fry register, and 1.6% were glottal whistles. On average, fewer than two call types were used in the course of bout production ($M=1.62$, s.d.$=0.84$), although some bouts consisted of as many as five types. Like bout durations, call durations were highly variable, with a standard deviation of 0.14 associated with the mean of 0.17 s. Call duration was strongly related to the type of call produced, $F(5,3473)=175.97$, $p<0.001$. Calls involving two or more production modes were the longest and, not surprisingly, glottal pulses were the shortest (see Table Ib).

The total number of calls produced did not differ by laugher sex, $F(1,96)=0.21$, ns. Consistent with their longer overall durations, male bouts contained somewhat more calls than did bouts produced by females, $F(1,1021)=6.90$, $p=0.01$ ($M_{\text{male}}=3.63$, s.d.$=2.86$; $M_{\text{female}}=3.18$, s.d.$=2.56$). Laugher sex had a strong influence on the proportions of call types produced, $\chi^2(5)=155.17$, $p<0.001$ (see Table Ib). Follow-up binomial tests showed that females produced significantly more voiced calls than did males ($p<0.001$), and that males produced significantly more unvoiced calls and glottal pulses than did females (p's<0.001). Laugher sex did not mediate either the acoustic complexity of laughs (as indexed by the number of call types per bout), call durations, or the number of calls produced per second [$F(1,1023)=1.83$, ns; and $F(1,3469)=0.01$, ns; $F(1,1023)=0.30$, ns, respectively].

TABLE I. Descriptive statistics associated with (a) bout- and (b) call-level analyses, separated according to laugher sex. Values in parentheses are standard deviations.

(a) Bout level	Males ($n=45$)				
Total (n)	465				
M Duration	0.95 (0.82)				
Bout type	Voiced	Unvoiced grunt-like	Unvoiced snort-like	Mixed	Glottal whistles
% Males producing	82.2	66.7	82.2	40.0	2.2
% of Total bouts	26.0	24.7	31.0	17.6	0.6
M Duration (s)	1.08 (0.83)	0.99 (0.91)	0.65 (0.49)	1.13 (0.96)	0.64 (0.42)
	Females ($n=52$)				
Total (n)	559				
M Duration	0.82 (.72)				
Bout type	Voiced	Unvoiced grunt-like	Unvoiced snort-like	Mixed	Glottal whistles
% Females producing	88.5	53.9	76.9	51.9	9.6
% of Total bouts	33.3	13.4	27.4	25.2	0.7
M duration (s)	1.04 (0.88)	0.79 (0.67)	0.53 (0.39)	0.90 (0.72)	0.51 (0.25)
(b) Call level	Males ($n=45$)				
Total (n)	1705				
M Calls per bout	3.61 (2.84)				
M Duration (s)	0.17 (0.14)				
Call type	Voiced	Unvoiced	Mixed	Glottal pulses	Glottal whistles
% Males producing	84.4	97.8	84.4	40.0	24.4
% of Total calls	27.6	52.6	13.0	5.5	1.3
M Duration (s)	0.11 (0.08)	0.20 (0.14)	0.24 (0.11)	0.03 (0.02)	0.22 (0.30)
	Females ($n=52$)				
Total (n)	1774				
M Calls per bout	3.20 (2.58)				
M Duration (s)	0.17 (0.14)				
Call-type	Voiced	Unvoiced	Mixed	Glottal pulses	Glottal whistles
% Females producing	88.5	96.2	90.4	23.1	38.5
% of Total calls	45.3	38.2	13.0	1.6	1.9
M Duration (s)	0.11 (0.08)	0.22 (0.15)	0.28 (0.17)	0.02 (0.02)	0.22 (0.20)

Further analyses examined temporal characteristics of calls within bouts. On average, 4.37 calls were produced per second, with comparable call- and intercall durations (i.e., 0.17 and 0.13 s, respectively). These two measures were also equivalent when examined only for voiced, open-mouth calls (0.11 and 0.12 s, respectively). A more fine-grained analysis examined the pattern of call- and intercall durations through the course of bouts that contained at least three but no more than eight calls. As can be seen in Fig. 3, bouts were typically initiated with comparatively long calls ($M=0.28$, s.d.$=0.15$) and followed by calls that were roughly half as long in duration ($M=0.13$, s.d.$=0.10$). This pattern was observed regardless of the number of calls per bout. The longer terminal-call durations of bouts with six or more calls contradict this general pattern, and largely reflect the prolonged inhalations and exhalations used to conclude some of these laugh episodes. The overall pattern of intercall intervals showed that regardless of the number of calls per bout, call production was denser towards the beginning of laugh bouts. Intercall durations gradually increased over the course of bouts and were longer than call durations by bout offset, especially for bouts comprised of six or more calls. Intercall intervals could become as long as twice that of call durations, but only by the seventh call in eight-call bouts.

3. Segment-level descriptive outcomes

A significant proportion of calls (30.9%) was composed of two or more discrete acoustic components. Most multisegment calls (75.8%) contained two components, an additional 20.7% contained three, and a small subset (3.5%) consisted of either four, five, or six segments. Mean segment duration was 0.11 s (s.d.$=0.11$), and there were no sex differences in the number of multisegment calls produced, $\chi^2(4)=5.50$, ns.

B. F_0-related outcomes

Descriptive statistics associated with F_0-related outcomes are shown in Table II. F_0 could be measured from 1617 voiced calls or voiced call segments. The ESPS/WAVES+ pitch-tracking algorithm performed well for about 65% of these cases, and the remaining measurements were made by extracting maximum-amplitude points from the first harmonic. Four dependent measures were of interest: mean F_0, s.d. F_0, F_0-excursion [(maximum call F_0)–(minimum call F_0)], and F_0 change [|(call-onset F_0)–(call-offset F_0)|].

Statistical tests involving F_0 measures used only those calls for which mouth position (i.e., open or closed) was readily perceptible, with a MANOVA used to test the extent

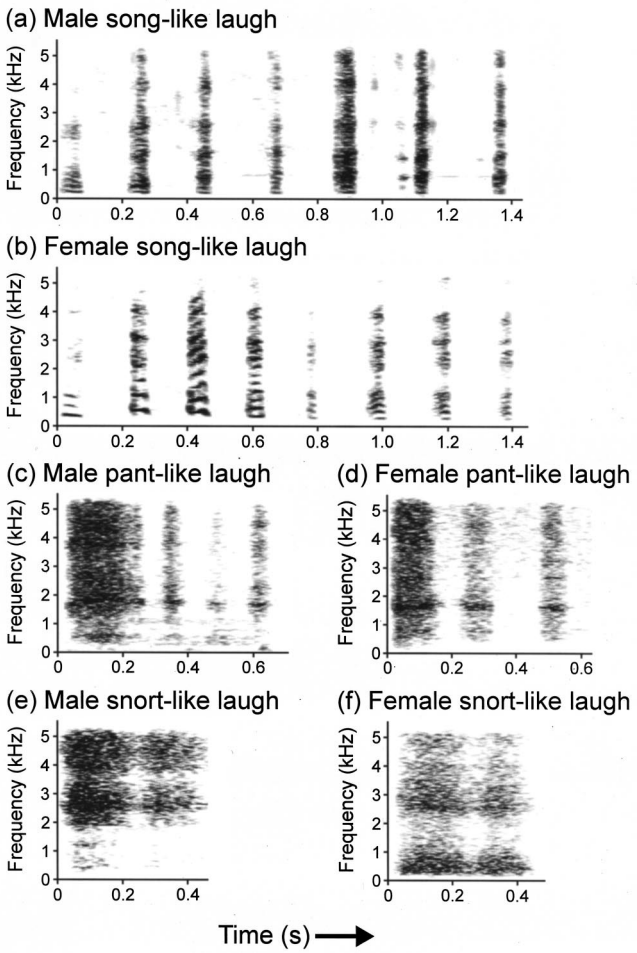


FIG. 2. Narrow-band spectrograms of (a) male and (b) female voiced laughs, wideband spectrograms of (c) male and (d) female unvoiced grunt-like laughs, and wideband spectrograms of unvoiced snort-like (e) male and (f) female laughs. Sample laughs can be heard at http://www.psy.vanderbilt.edu/faculty/bachorowski/laugh.htm

to which laugher sex and mouth position were associated with differences in the four dependent variables. Outcomes for all measures were strongly influenced by laugher sex: Results for mean F_0, s.d. F_0, F_0 excursion, and F_0 change were $F(1,1538)=165.10$, 45.58, 43.80, and 37.22, respectively (all p's<0.001). Not unexpectedly, the mean of 405 Hz (s.d.=193) measured from female laughs was considerably higher and more variable than the mean of 272 Hz (s.d.=148) found for male laughs. Also notable were the male and female absolute-maximum F_0 values of 1245 and 2083 Hz, respectively (for an example of a high F_0 call, see Fig. 4). Within-call F_0 standard deviations were quite high, on average being 21.41 and 29.98 Hz for male and female laughs, respectively. Mean F_0 excursion was also large for both sexes, but especially so for females ($M_{\text{male}}=59$ Hz, s.d.=49.74; $M_{\text{female}}=86$ Hz, s.d.=76.83). Both sexes were similarly found to have large onset to offset F_0 ranges, with females again showing the biggest change ($M_{\text{male}}=44$ Hz, s.d.=42.38; $M_{\text{female}}=64$ Hz, s.d.=63.60). There was also a significant main effect of mouth position for mean F_0, $F(1,1538)=33.43$, $p<0.001$, which was due to the higher F_0's of open- than closed-mouthed calls. Mouth position did not mediate outcomes for any of the three variability measures, and the interactions between laugher sex and mouth position were all nonsignificant.

Temporal patterning of F_0 at the call level was examined for the 297 voiced calls that were produced during the course of 96 randomly selected, predominantly voiced bouts. Using terminology common to the infant-directed speech literature (e.g., Katz, Cohn, and Moore, 1996), the F_0 contour of each call was characterized as being either ''flat,'' ''rising,'' ''falling,'' ''arched,'' or ''sinusoidal.'' Using this classification scheme, the most common contour designation was flat (38.0%). However, falling (29.0%) and sinusoidal (18.9%) types each accounted for a sizable proportion of call contours, and arched (8.1%) and rising (6.1%) contours were not uncommon.

Several remarkable aspects of laugh acoustics were highlighted by examining F_0 measures at the bout level. Using a MANOVA, bouts containing two or more voiced calls or call segments were tested, with the number of voiced segments contributing to each bout as a weighted least-squares regression coefficient (Darlington, 1990). Laugher sex and bout length were used as fixed factors, the latter being a dichotomous variable created by classifying laughs into ''short'' and ''long'' categories based on the median number of voiced segments. Short bouts therefore contained either two or three voiced segments, whereas long bouts consisted of four or more voiced segments.

As was certain to be the case given call-level outcomes, the main effects of laugher sex were significant for both mean F_0 and F_0 excursion [$F(1,388)=85.63$, $p<0.001$, and $F(1,388)=10.05$, $p<0.01$, respectively]. Both measures were also found to be strongly associated with the number of voiced segments in a laugh episode [$F(1,388)=21.20$, $p=0.01$, and $F(1,388)=56.72$, $p<0.001$, for mean F_0 and F_0 excursion, respectively]. Compared to short bouts, long bouts were found to have higher mean F_0's as well as greater F_0 excursions (see Table III). For male laughs, the difference in mean F_0 between short and long bouts was 77 Hz, whereas this difference was 48 Hz for females. Very large differences were found for F_0 excursion, with the discrepancies between short and long bouts being 161 and 189 Hz for male and female laughs, respectively. Also noteworthy were the extreme F_0 excursions that occurred during bout production, with a male maximum of 947 Hz and corresponding female value of 1701 Hz. Moreover, such extreme excursions were not altogether rare events: 7 males produced a total of 12 bouts with F_0 excursions of 500 Hz or more, and 13 females produced a total of 31 bouts with excursions of this magnitude or greater.

Patterns of mean F_0 over the course of bout production were also examined. Briefly, we found no evidence of an overall decline in F_0. For bouts with either two, three, or four voiced components, F_0 at bout offset was nearly the same as at bout onset. For bouts with greater numbers of voiced segments, F_0 routinely increased and decreased, but did not fluctuate in an obvious pattern. Here, bout-offset F_0's were often higher than bout-onset F_0's.

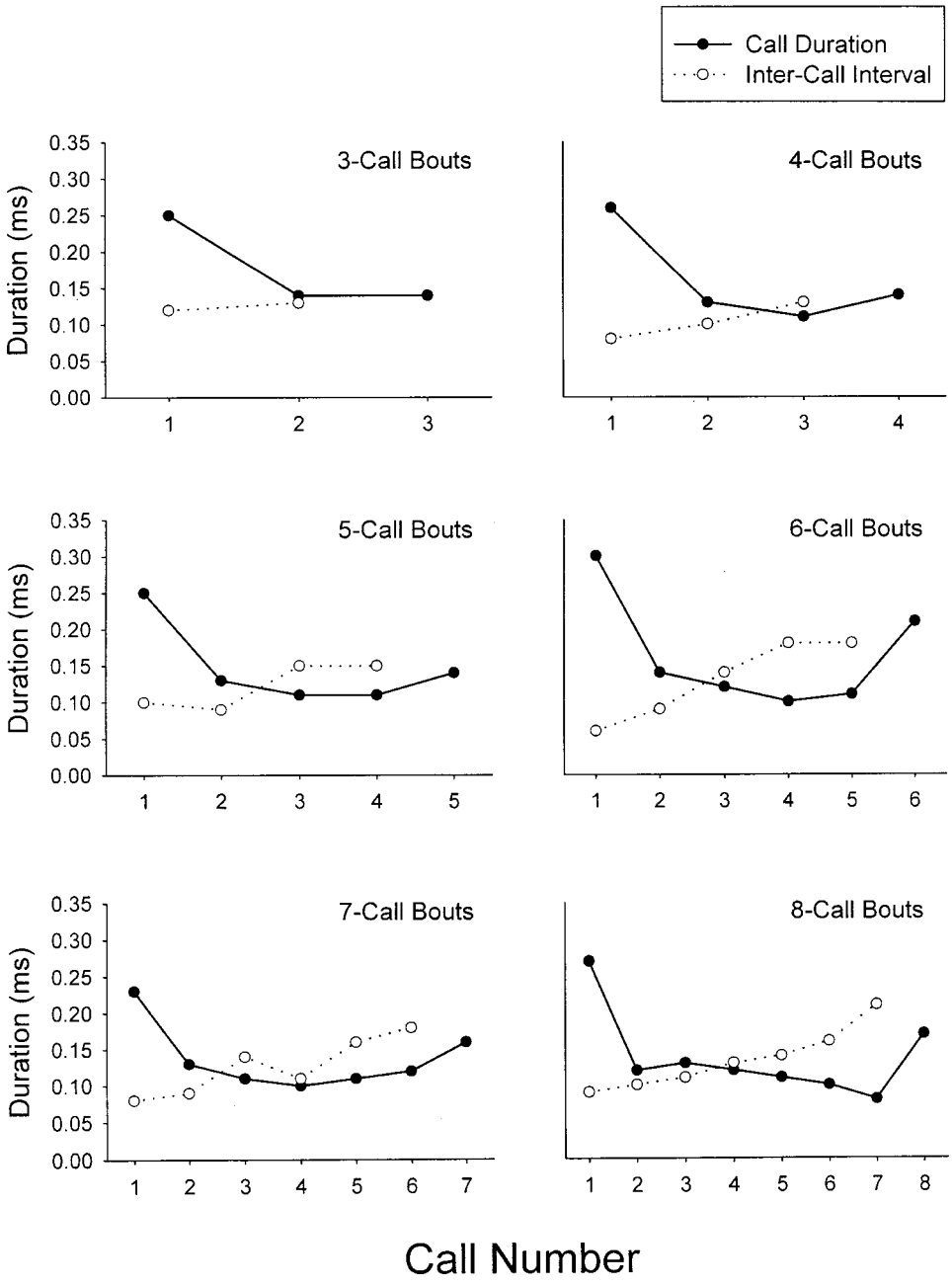


FIG. 3. Call durations and intercall intervals for laugh bouts comprised of three through eight calls.

C. Non-normative source energy

The 22 instances of open-mouth vocal fry for which F_0 could be measured showed very low F_0's, with pulses visible even on narrow-band spectrograms. A main effect of laugher sex was found for mean F_0 in vocal fry, $F(1,21) = 6.65$, $p < 0.025$ ($M_{\text{male}} = 80$ Hz, s.d. $= 19.60$; $M_{\text{female}} = 110$ Hz, s.d. $= 31.99$). However, males and females did not differ on any of the three variability indices (i.e., s.d. F_0, F_0 excursion, and F_0 change).

A total of 136 calls with nonlinear phenomena was identified (see Riede *et al.*, 2000; Wilden *et al.*, 1998). Of these, 105 were labeled as glottal whistles [see Fig. 5(a)], possibly reflecting airstream vortices induced by the medial edges of the vocal folds. These calls sounded wheeze-like, were typically low amplitude and quasiperiodic, and exhibited waveforms that were virtually indistinguishable from those of whistled /s/'s that can occur in naturally produced speech. The second sort of nonlinear phenomenon was the occurrence of subharmonics [Fig. 5(b)], typically period doubling, which was found in 26 calls. Perceptually, these sounds had a rather tinny quality. Finally, we observed five instances of biphonation, which involves the occurrence of two independent fundamental frequencies [Fig. 5(c)]. These calls sounded shrill and dissonant.

D. Formant-related outcomes

The primary goal of this series of analyses was to provide normative data concerning the spectral properties of laughter. Whenever possible, peak frequencies of five vocal-tract resonances were measured. However, accurate spectral measurements were difficult for any of several reasons. First, the noisiness of many unvoiced calls precluded adequate for-

TABLE II. F_0-related outcomes for call-level analyses, separated according to laugher sex and mouth position (i.e., open or closed). Tabled values are means, with standard deviations in parentheses.

Measures[a] (Hz)	Males		Females	
	Open mouth ($n=563$)	Closed mouth ($n=131$)	Open mouth ($n=862$)	Closed mouth ($n=276$)
MF_0	279 (146)	216 (92)	415 (193)	355 (127)
s.d. F_0	22 (17)	19 (16)	30 (24)	30 (25)
F_0-Excursion[b]	60 (51)	51 (42)	88 (78)	82 (72)
F_0-Change[c]	45 (43)	40 (39)	63 (62)	66 (69)

[a]Data from 34 males and 43 females contributed to analysis of open-mouth calls, whereas data from 25 males and 33 females were used for analysis of closed-mouth calls.
[b]F_0-Excursion=[(maximum call-F_0)−(minimum call-F_0)].
[c]F_0-Change=[|(call-onset F_0)−(call-offset F_0)|].

TABLE III. Bout-level F_0 measures, separated according to laugher sex. Values in parentheses are standard deviations.

Measures[a] (Hz)	Males		Females	
	Short bouts[b]	Long bouts[c]	Short bouts	Long bouts
MF_0	223 (168)	305 (262)	373 (266)	426 (332)
MF_0-Excursion[d]	141 (197)	299 (429)	191 (342)	405 (683)
Absolute minimum F_0-excursion	13	44	29	62
Absolute maximum F_0-excursion	741	947	991	1701

[a]Data from 37 males and 40 females contributed to short-bout analyses, whereas data from 24 males and 31 females were used in long-bout analyses.
[b]Short bouts contained either two or three voiced calls or call segments.
[c]Long bouts contained four or more voiced calls or call segments.
[d]F_0-Excursion=[(maximum call-F_0)−(minimum call-F_0)].

mant resolution. Second, LPC outcomes were occasionally driven by the high-amplitude harmonics associated with some voiced calls. Third, the harmonically sparse spectra of calls with very high F_0's left little opportunity for supralaryngeal filtering to have a visible effect. In either of these last two cases, peak frequencies were coincident with one or more harmonics and adjusting the number of coefficients had little or no impact on the LPC solution. Resonance-harmonic correspondence was observed for 428, 135, 85, 55, and 36 instances of $F1$ through $F5$ measurements, respectively. Our overall strategy was therefore to take measurements from those calls for which three or more formants were readily identifiable, and for which peak frequencies did not coincide with harmonics. As noted earlier, we did not measure formant frequencies of unvoiced, snort-like sounds because the extent to which airstream interactions with the microphone element were contributing to spectral characteristics was unclear. Finally, outcomes are not shown for either glottal pulses or whistles. The former were usually too brief for reliable measurement, and the latter were notably unstable. This overall selection procedure resulted in a sample of 1717 calls from 89 individuals. The reader is referred to Footnote 1[1] for details concerning treatment of missing data.

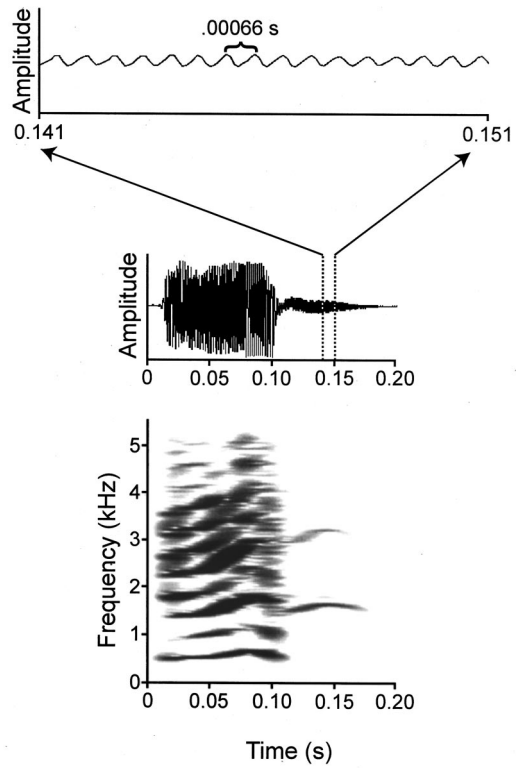


FIG. 4. Waveform (middle) and corresponding narrow-band spectrogram (bottom) of a very high F_0 call. Dotted vertical lines frame the portion of the waveform that is enlarged at the top.

A grand MANOVA confirmed that formant frequencies differed depending on call-production mode (i.e., voiced open mouth, open-mouth vocal fry, voiced close mouth, and unvoiced open mouth). Further MANOVAs were therefore conducted within each production mode, with detailed outcomes provided in Table IV. For voiced, open-mouth laughs, formant frequencies were significantly lower in males than in females, at least for $F1$, $F2$, and $F3$, $F(1,587)=115.81$, 77.06, and 316.61 (all p's<0.001). However, laugher sex did not mediate $F4$ values, $F(1,587)=0.14$, ns, and female $F5$ values were actually significantly *lower* than in males, $F(1,587)=43.34$, $p<0.001$. For voiced, closed-mouth calls, only $F3$ values distinguished between the sexes, with male sounds being lower $F(1,86)=5.20$, $p=0.025$. Vocal fry was associated with significantly lower $F2$ and $F3$ values in males than in females [$F(1,38)=5.50$, $p<0.025$, and $F(1,38)=32.67$, $p<0.001$, respectively]. As was found for voiced open-mouth calls, $F5$ values were significantly lower for female than for male fry laughter $F(1,38)=15.12$, $p<0.001$. Peak frequencies of unvoiced, open-mouth calls were significantly lower for males than for females for the lowest three formants, $F(1,358)=81.95$, 20.90, and 95.93, respectively (all p's<0.001), but laugher sex did not affect the two highest resonances.

One way to characterize these outcomes was to plot $F1$ and $F2$ values in standard vowel space representations. Plots of voiced open-mouth and unvoiced open-mouth data were made using both Peterson and Barney's (1952) classic depiction and Hillenbrand *et al.*'s (1995) more recent version. For brevity, we show outcomes only using the latter representation [Figs. 6(a)–(d)]. Regardless of laugher-sex or call-production mode, these depictions show that laughter predominantly consists of central sounds. In males, for instance, the great majority of voiced open-mouth calls fell within /ɜ˞/ and /ʌ/ ellipses. Female outcomes were more variable, but

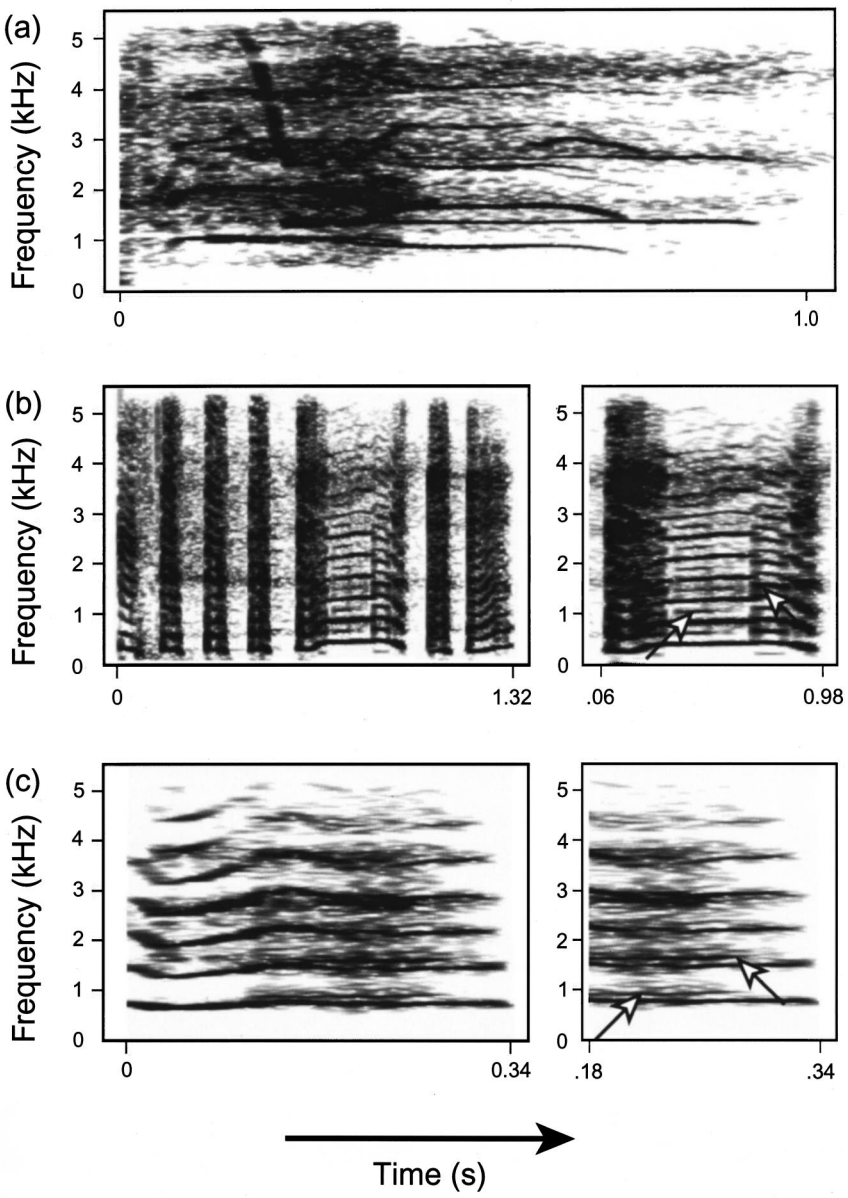


FIG. 5. Narrow-band spectrographic representations of three types of nonnormative source energy. At the top (a), the kinds of spectral nonlinearities characteristic of glottal whistles are clearly evident. In (b), subharmonics are apparent in the last three calls of this seven-call bout, with arrows on the enlarged version to the right pointing to subharmonic energy. An instance of biphonation is depicted in (c), with the narrow-band spectrogram to the left revealing independent frequencies, and arrows highlighting two of these frequencies to the right.

most cases of voiced open-mouth calls were nonetheless located within central ellipses (i.e., /ɜ˞/, /ʌ/, /ɑ/ and /ɛ/). In contrast, there were very few observations of noncentral sounds by either sex, contrary to stereotypical notions that laughter includes sounds like ''tee-hee'' or ''ho-ho.'' In fact, no observations fell into the /i/ range, and very few were found within either the /ɪ/ or /ɔ/ ellipses. Quite similar outcomes were found for male unvoiced open-mouth calls, whereas the majority of female versions of these sounds fell within /ɛ/ and /a/ ellipses and the undefined region between these two spaces.

In part to handle the large scaling differences between $F1$ and $F2$, vowel space depictions typically use nonequivalent axes. For instance, Peterson-and-Barney-type representations plot $F1$ using a linear scale but show $F2$ values on a logarithmic scale. Hillenbrand *et al.* did use linear scales for

TABLE IV. Male and female formant-frequency values according to call type. Tabled values are means, with standard deviations in parentheses.

	Sex (n)	$F1$	$F2$	$F3$	$F4$	$F5$
Voiced open mouth	M (41)	535 (112)	1592 (153)	2576 (180)	3667 (180)	4593 (160)
	F (34)	653 (155)	1713 (182)	2875 (227)	3673 (223)	4506 (157)
Voiced closed mouth	M (27)	445 (142)	1746 (187)	2527 (128)	3693 (278)	4588 (195)
	F (17)	501 (155)	1738 (291)	2636 (212)	3616 (238)	4548 (172)
Vocal fry	M (18)	582 (109)	1551 (115)	2509 (117)	3591 (126)	4574 (211)
	F (16)	638 (92)	1655 (153)	2764 (153)	3695 (220)	4290 (237)
Unvoiced open mouth	M (36)	594 (163)	1661 (155)	2589 (214)	3660 (183)	4602 (126)
	F (34)	770 (176)	1746 (171)	2826 (196)	3678 (205)	4583 (169)

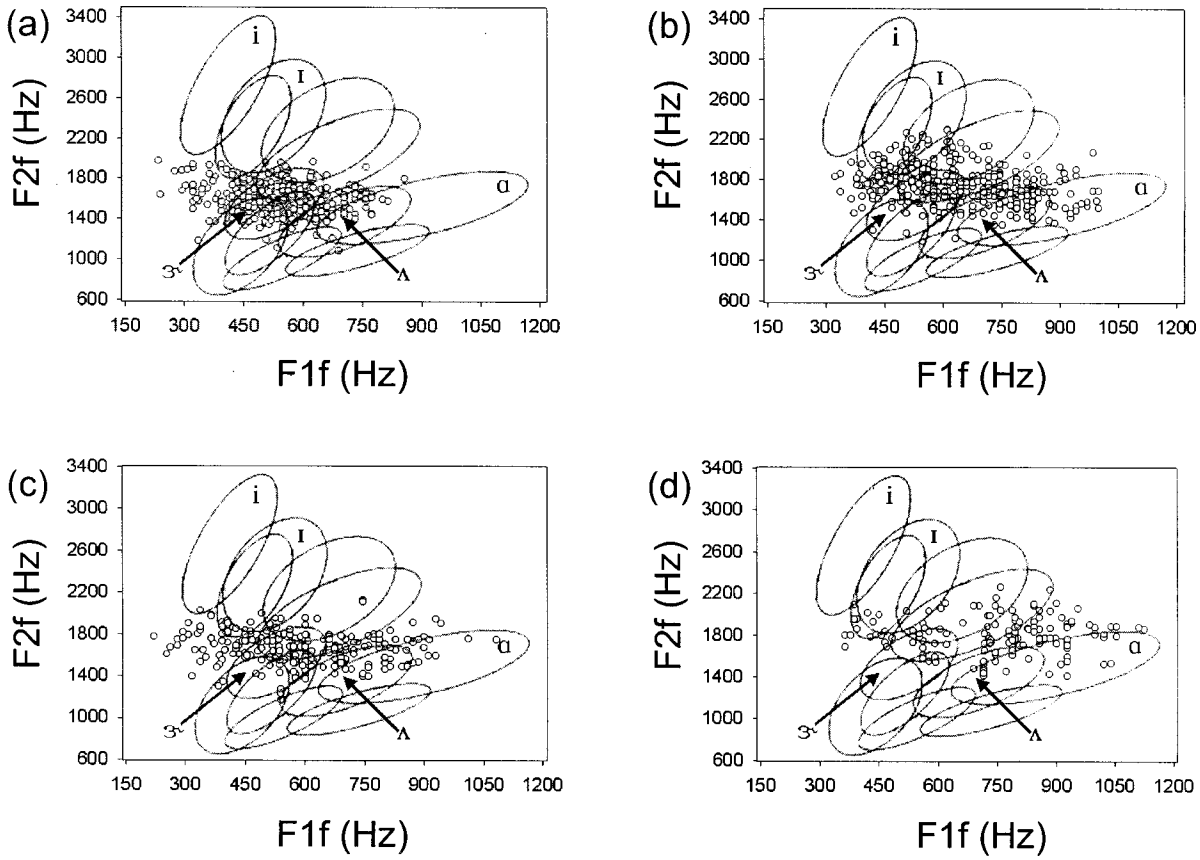


FIG. 6. Values of $F1$ and $F2$ plotted for (a) male open-mouth voiced calls; (b) female open-mouth voiced calls; (c) male open-mouth unvoiced calls, and (d) female open-mouth unvoiced calls; using Hillenbrand *et al.*'s (1995) vowel-space map.

both axes, but with different tick-mark intervals. In order to examine variability unconfounded by scaling differences, we also plotted the data using equivalent axes [Figs. 7(a)–(d)]. These representations yielded circular rather than elliptical distributions, indicating that on average the variability associated with the two resonances is essentially equivalent. Comparing the $F1$ and $F2$ distribution moments confirmed these impressions (outcomes can be obtained from author J.A.B.).

E. Acoustic correlates of laugher sex and individual identity

Earlier work involving a large set of homogeneous vowel sounds excised from running speech revealed that acoustic characteristics related to F_0 and formants play prominent but varying roles in differentiating talkers by sex and individual identity (Bachorowski and Owren, 1999). Similar analyses were conducted here, although with a smaller number of observations. This testing focused on voiced open-mouth and unvoiced open-mouth calls. For voiced calls, mean F_0, s.d. of F_0, F_0 excursion, F_0 change, $F1-F5$, VTL, and call duration were the measures used, while $F1-F5$, VTL, and call duration were examined for unvoiced calls. For each call type, only participants represented by six or more completely analyzable observations were used in classification analyses. Given these selection criteria, data from 19 males and 13 females were available for tests with voiced open-mouth sounds, whereas data from 11 males and 7 females contributed to analyses of unvoiced open-mouth calls. Eight males and five females were represented in both voiced and unvoiced call analyses.

Here, each subject was first entered as a unique independent variable in a MANOVA. Only those acoustic measures for which individual laughers differed from each other were subsequently used in discriminant-function analyses (Tabachnik and Fidell, 1996), which in practice meant that call duration, F_0 change, and $F4$ were not used in voiced-call laugher-sex analyses, $F4$ was not used in unvoiced-call laugher-sex analyses, and call duration was not used for individual laugher classification of females. The remaining variables were then entered in stepwise fashion in discriminant function analyses using the Mahalanobis-distance method, and the performance of discriminant functions was cross validated with the jackknife procedure. Functions were derived using the actual number of cases available for each subject. The overall approach was to compare outcomes for the full set of acoustic measures with particular subsets of interest.

Classification outcomes for laugher sex are given in Table V. Results are shown for classification accuracies in derivation and test phases, as well as the percent error reduction associated with the former. This last metric takes into account chance error rate, producing an unbiased measure of classification accuracy. For voiced open-mouth calls, the most successful classification (86.3%) occurred with the complete set of dependent measures, but only $F1$, $F2$, $F3$,

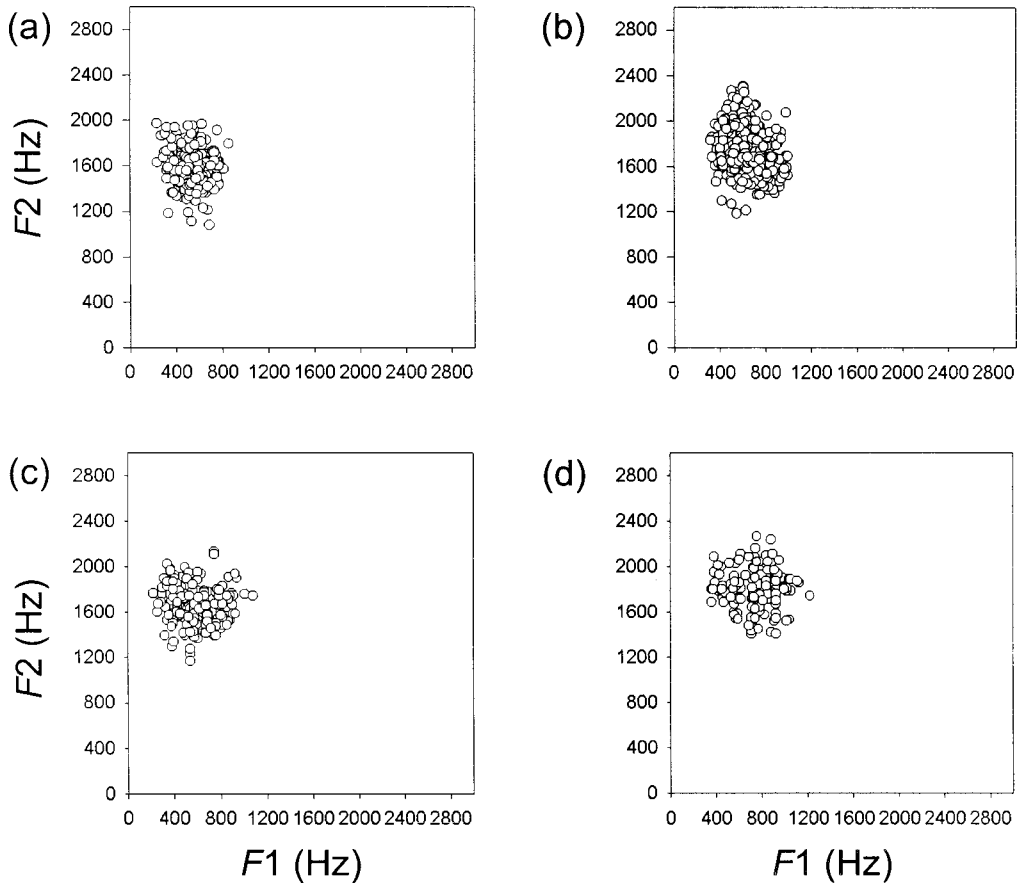


FIG. 7. Using linear axes to anchor values of both $F1$ and $F2$, data are plotted for (a) male open-mouth voiced calls; (b) female open-mouth voiced calls; (c) male open-mouth unvoiced calls; and (d) female open-mouth unvoiced calls.

and VTL met entry criteria. In other words, none of the F_0-related measures contributed significantly to classification by sex when tested in conjunction with spectrally related cues. Other comparisons also showed formant frequencies to be the most important in sorting laughers by sex. For instance, the set of four formant frequencies that entered the analysis was associated with 85.4%-correct classification (70.8% error reduction), whereas the three F_0-related measures together led to 60.6%-correct classification (only 21.2% error reduction). Similarly, VTL alone classified 79.5% of cases (59.0% error reduction), whereas mean F_0 produced only 61.2% correct (22.4% error reduction). Filter-related cues were also found to be important for sorting unvoiced calls by laugher sex. For instance, classification accuracy was 84.8% (69.6% error reduction) using only the four formant frequencies, and testing VTL alone led to virtually identical outcomes.

Classification of individual laughers within each sex was less successful. Even so, these outcomes were significantly better than expected by chance, and should be useful in developing more refined hypotheses concerning individual distinctiveness of laugh sounds. Here, we note only a few of the outcomes (also see Table VI). Overall, more effective classification occurred for female than for male calls—an outcome at least partly attributable to the smaller number of females being classified. For voiced calls produced by either sex, formant frequencies were again far more important in classifying individuals than were F_0-related measures. Whereas the former were associated with 41.2% and 49.0% correct classification for males and females, respectively, the latter produced corresponding values of only 15.4% and 22.6%. For males but not females, classification of unvoiced calls was also effective.

IV. DISCUSSION

The present study provides detailed acoustic outcomes for a large corpus of laugh sounds produced by a correspondingly large number of laughers. In addition to providing an extensive characterization of laugh acoustics, this work also suggests four broad findings concerning these sounds. First, in contrast to perspectives that emphasize stereotypy in laughter, we found this signal to be notable for its acoustic variability. Second, this variability was associated with a diversity of evident underlying vocal-production modes. Third, we found vowel-like laughs to be comprised of central, unarticulated sounds and lacking in the vowel-quality distinctions commonly thought to be present. Finally, we obtained preliminary evidence that indexical cues to laugher sex and individual identity are conveyed in laugh acoustics. The following sections elaborate on both these and other results, and include comparisons to previously reported outcomes and hypotheses concerning laugh acoustics (see Table VII for key comparisons between the current work and other studies).

TABLE V. Results of discriminant function analyses for laugher-sex classification using both the full complement of acoustic cues and theoretically derived groups of measures. Test accuracy was assessed with the jackknife procedure. Chance classification accuracy was 50%.

	Derivation accuracy	Test accuracy	Error reduction[a]
Voiced open-mouth calls[b]			
All measures[c]	86.7	86.3	72.6
$F1,F2,F3,F5$	85.4	85.4	70.8
$F1,F2,F3$	84.4	84.4	68.8
F_0-related measures[d]	60.8	60.6	21.2
VTL, mean F_0	78.9	78.7	57.4
VTL	79.5	79.5	59.0
Mean F_0	61.2	61.2	22.4
Unvoiced open-mouth calls[e]			
All measures[f]	88.2	87.4	74.8
$F1f,F2f,F3f,F5f$	84.8	84.8	69.6
$F1f,F2f,F3f$	80.7	80.3	60.6
VTL	85.4	84.6	69.2

[a]Error reduction=([(100—chance rate)−(100—observed rate)]×100)/(100—chance rate).
[b]Data came from 19 males and 13 females.
[c]Mean F_0, s.d. F_0, F_0-excursion, $F1-F5$, VTL, and call duration.
[d]Mean F_0, s.d. F_0, F_0-excursion.
[e]Data came from 11 males and 7 females.
[f]$F1-F5$, VTL, call duration.

TABLE VI. Results of discriminant function analyses for individual laughers within each sex using both the full complement of acoustic cues and theoretically derived groups of measures. Test accuracy was assessed with the jackknife procedure. Chance classification accuracies were 5.3% and 7.7% for male and female voiced open-mouth calls, and 9.1% and 14.3%, for male and female unvoiced open-mouth calls, respectively.

	Derivation accuracy	Test accuracy	Error reduction
I. Voiced open mouth			
(a) Males ($n=19$; 271 cases)			
All measures[a]	58.3	42.8	39.6
$F1,F2,F3,F4,F5$	45.5	41.2	38.1
F_0-related measures[b]	17.1	15.4	10.9
Mean F_0, VTL	24.0	21.5	17.1
Mean F_0	15.4	15.4	10.9
VTL	13.3	11.6	6.7
(b) Females ($n=13$; 211 cases)			
All measures[c]	61.3	53.2	49.3
$F1,F2,F3,F4,F5$	55.7	49.0	44.7
F_0-related measures	26.9	22.6	16.1
Mean F_0, VTL	28.5	25.8	19.6
Mean F_0	23.5	23.5	17.1
VTL	28.1	27.1	21.0
II. Unvoiced open mouth			
(a) Males ($n=11$; 207 cases)			
All measures[d]	53.5	47.6	42.4
$F1,F2,F3,F4,F5$	50.8	48.1	42.9
VTL	31.6	27.8	20.6
(b) Females ($n=7$; 63 cases)			
All measures[e]	76.3	40.7	30.8
$F1,F2,F3,F4,F5$	69.5	35.6	24.9
VTL	39.0	37.3	26.8

[a]Mean F_0, s.d. F_0, F_0-excursion, F_0-change, $F1-F5$, VTL, and call duration.
[b]Mean F_0, s.d. F_0, F_0-excursion, F_0-change.
[c]Mean F_0, s.d. F_0, F_0-excursion, F_0-change, $F1-F5$, VTL, and call duration. With the exception of call duration, measures for females were the same as those for males.
[d]$F1-F5$, VTL, and call duration.
[e]$F1-F5$ and VTL.

A. Laughter is highly variable

1. Temporal variability

On average, laugh bouts were a bit less than 1 s in duration (i.e., 870 ms) and consisted of 3.39 calls, each 170 ms long and 130 ms apart. However, considerable variability was found for every measure examined. For instance, bouts could be as short as 40 ms but as long as 5699 ms, while call durations ranged from 5 to 1053 ms. The number of calls involved was also highly variable, with many bouts consisting of only a single call but others including up to 20.

Overall, call durations and intercall durations were found to be quite comparable (cf. Ruch and Ekman, 2001). However, more detailed examinations showed that intercall intervals were markedly shorter than call durations at bout onset (see Fig. 3), with call production thus being more densely packed at the beginning than at the end of bouts. In other words, while our outcomes replicated the gradual increase in intercall interval noted by Provine (1996; Provine and Yong, 1991), we did not find evidence of a proposed monotonic decrease in call duration over the course of each bout (e.g., Provine and Yong, 1991; Ruch and Ekman, 2001). We instead found that calls produced at bout onset were much longer than later calls, with little subsequent variation among the latter.

Outcomes concerning the rate of laugh-sound production are also of interest. Using data from one male and one female laugher, Bickley and Hunnicutt (1992) found a rate of 4.7 calls/s, which is a bit greater than our obtained mean of 4.37 calls/s. Treating laugh calls as syllables, both of these rates are faster than the mean discourse rate of 3.26 syllables/s produced by comparably aged young adults (Venkatagiri, 1999). Conversely, young adults have been shown to produce laugh-like syllables at higher rates than those found here. For instance, a mean maximum-repetition rate of 5.46 was found for females producing /hʌ/ syllables (Shanks, 1970), whereas a mean maximum-repetition rate of 5.1 was reported for males producing /ʌ/ syllables (Ptacek *et al.*, 1966). Taken together, these comparisons indicate that average sound-production rates are faster in laughter than in conversational speech, without reaching the maximum possible rate.

2. Source variability

Many of the outcomes associated with F_0-related measures were remarkable. Here, we focus primarily on analyses of open-mouth calls or segments, as these accounted for the vast majority of voiced-laugh components. Consistent with several previous reports (Provine and Yong, 1991; Rothgänger *et al.*, 1998; Nwokah *et al.*, 1999; see Table VII), we found that mean F_0 of both male (282 Hz) and female (421 Hz) laughter was considerably higher than in modal speech (120 and 220 Hz for males and females, respectively). However, lower mean F_0 values have been reported by others, which we suspect may reflect either that those studies examined laughter from subjects that were tested alone (e.g.,

TABLE VII. Comparisons among the present results and other published reports. Tabled values are means, with standard deviations in parentheses.

Study	Sample size	Laugh-sampling method	Number of laugh bouts	Bout duration	Calls per bout	Minimum and Maximum calls per bout	Call duration
Bachorowski, Smoski, and Owren (2001)	45 males 52 females	Humorous video clips	1024	0.87 s (0.77)	3.39 (2.71)	1, 20	0.17 s (0.14)
Bickley and Hunnicutt (1992)	1 male 1 female	Spontaneous laughs produced during speech task	15		Laugher 1: 6.7 Laugher 2: 1.2		
Milford (1980)	15 males 15 females	Social, tension-release, humor, and tickle		1.34 s			
Mowrer, LaPointe, and Case (1987)	11 males	Humorous video clips	55	1.22 s (0.44)[a]	7.16 (2.42)	1, 25	
Nwokah, Hsu, Davies, and Fogel (1999)[b]	3 females	Mothers interacting with their infants	3	2.14	8.67	6, 14	
Provine and Yong (1991)	23 males 28 females	First ''spontaneous'' laugh after request to laugh	51		4.00	4, 16	0.08 s (0.02)[c]
Rothganger, Hauser, Cappellini, and Guidotti (1998)	20 males 20 females	Humorous video clips	187	0.75	5.90 (2.18)		0.13 s (0.06)

Study	Mean F_0 (Hz)	F_0 Range (Hz)	$F1f$ (Hz)	$F2f$ (Hz)	$F3f$ (Hz)	$F4f$ (Hz)	$F5f$ (Hz)
Bachorowski Smoski, and Owren (2001)[d]	M: 284 (155)[e] F: 421 (208)	M: 67 (76) F: 91 (85)	M: 534 (111) F: 637 (149)	M: 1589 (153) F: 1734 (193)	M: 2571 (182) F: 2887 (253)	M: 3663 (184) F: 3725 (273)	M: 4594 (161) F: 4513 (167)
Bickley and Hunnicuttt (1992)	M: 138 F: 266	M: 55 F: 315	M: 650 F: 650	M: 1700 F: 1800	M: 2200 F: 2760		
Milford (1980)	M: 175 F: 160		M: 543 F: 599	M: 1687 F: 1847			
Mowrer, LaPointe, and Case (1987)	M: 126 (42.7)	M: 69					
Nwokah, Hsu, Davies, and Fogel (1999)[b]	F: 365 (28)	F: 161					
Provine and Yong (1991)[g]	M: 276 (95) F: 502 (127)						
Rothganger, Hauser, Cappellimi, and Guidotti (1998)	M: 424 F: 475 (125)						

[a]Laugh selection required that bout duration be at least 250 ms.
[b]Some outcomes provided here were derived from results given in the original reports.
[c]Given the authors' descriptions, we assume these durations to reflect voiced portions of calls.
[d]Acoustic outcomes shown here are for voiced, open mouth calls.
[e]M=male; F=female.
[f]These formant outcomes were provided as examples rather than arithmetic means.
[g]F_0 measurements were made for the first call of each bout examined.

Bickley and Hunnicutt, 1992; see Bachorowski *et al.*, 2001), or were influenced by uncorrected errors occurring in automated pitch extraction. For example, algorithm failures likely contributed to Milford's (1980) implausibly low mean F_0 of 160 Hz for female laughter. Automated pitch-extraction errors are particularly likely to occur in laughter because F_0 variation in both calls and bouts is quite high. Although individual voiced calls were found to be quite brief, their mean F_0 excursions were nonetheless 67 and 91 Hz for males and females, respectively (also see Mowrer *et al.*, 1987; Nwokah *et al.*, 1999). Across all sounds, male F_0 was found to be as low as 43 Hz but as high as 898 Hz, whereas female F_0 was shown to be as low as 70 Hz and as high as 2083 Hz. Laughers can thus span the full range of possible F_0 variation, from the lowest vocal fry to the highest falsetto (see Hollien, Dew, and Phillips, 1971; Rothgänger *et al.*, 1998).

Consistent with Bickley and Hunnicutt's (1992) results, we found no evidence that F_0 necessarily decreases across the course of a multicall bout. In other words, laughter does not appear to exhibit the F_0-declination effect that at least some researchers report to be characteristic of human speech (see 't Hart, Collier, and Cohen, 1990). However, we did find that F_0 characteristics were markedly different depending on the length of a bout. Specifically, long bouts were associated

with both higher mean F_0 and greater F_0 variability than were shorter ones.

3. Variability in production modes

In contrast to perspectives that treat laugh sounds as being comparatively stereotyped (e.g., Grammer and Eibl-Eibesfeldt, 1990; Provine and Yong, 1991; Provine, 1996), we found laughter to be a repertoire of highly variable vocalizations that includes qualitatively distinct voiced song-like, unvoiced grunt-like, and unvoiced snort-like versions. Like the outcomes reported by Grammer and Eibl-Eibesfeldt (1990), our results showed that females produced more song-like bouts than males, whereas males produced proportionately more grunt-like laughs than females. These sex differences aside, individual laughers did not seem to rely on any particular style, with two of the three major bout types being produced by 84% of the individuals in this sample. Given that these laughs occurred within a 3.95-min window, it seems reasonable to assume that everyone produces each kind of bout at least some of the time. We also suspect that within the three broad types identified here, there may be subtypes that are acoustically and perhaps functionally important. That may be most relevant to voiced laughs, for which significant acoustic distinctions were found between shorter and longer bouts. However, a larger sample of voiced laughs would be necessary to reliably evaluate the possibility of subtypes.

While fewer than two production modes were found in most bouts, some included up to five production modes. Individual calls could also be acoustically complex, with the majority of compound calls consisting of both voiced and unvoiced components. Of these, it was more often the case that voiced segments preceded unvoiced segments than the converse (cf. Provine and Yong, 1991; Ruch and Ekman, 2001). A number of other combinations also occurred. For example, several instances of adjacent vocal fry and very high F_0 segments were noted, indicating that laughers can effect substantial and instantaneous changes in vocal-fold vibration rates (see also Rothgänger *et al.*, 1998). Variability in production modes may be driven by a number of factors, including individual style differences, linkages between laugher arousal and production processes, and social-context-based influences (Bachorowski *et al.*, 2001).

In addition to voicing distinctions, variability was also evident in the variety of non-normative source energies used in laugh production. Taken together, instances of vocal fry, glottal pulses, laryngeal whistles, subharmonics, and biphonation accounted for nearly 10% of the 3479 calls in this sample. The occurrence of subharmonics and biphonation in laughter is of particular interest (see also Riede, Wilden, and Tembrock, 1997; Švec *et al.*, 2000), as these kinds of nonlinearities are prominent features of some of the call types produced by any number of mammalian species. For instance, subharmonics have been observed in the calls of African wild dogs (Wilden *et al.*, 1998), rhesus and Japanese macaques (Riede *et al.*, 1997; Owren, 2001), and in the cries of human infants (Mende, Herzel, and Wermke, 1990; Hirschberg, 1999). Sounds with perceptually salient nonlinearities of this sort should be particularly effective in eliciting listener attention and arousal (see Owren and Rendall, 2001), and we expect the same to be true of nonlinear laugh sounds. We more specifically suspect that many instances of laugh nonlinearities are likely to be perceptually somewhat aversive if heard in isolation, but that these sounds may nonetheless enhance laughter's emotion-inducing effects when heard in conjunction with comparatively tonal calls.

B. Laughter is not articulated

Formant outcomes were generally within the bounds expected of speech acoustics for both sexes. As is typically the case due to dimorphism in supralaryngeal vocal-tract length, peak formant-frequency values of male calls were significantly lower than those of female calls for each of the lowest three resonances.

Plots of $F1$ and $F2$ outcomes in traditional vowel space showed that laugh utterances are generally clustered in /ɜ˞/ and /ʌ/ ellipses (Hillenbrand *et al.*, 1995; see also Ladefoged, 1993; Olive, Greenwood, and Coleman, 1993; Pullum and Ladusaw, 1996), thus being largely comprised of central, unarticulated sounds (see also Edmonson, 1987; Ruch and Ekman, 2001). To a lesser extent, observations also occurred within /ɑ/ and /ɛ/ ellipses. In the absence of large discrepancies between plots of voiced open-mouth and unvoiced open-mouth calls, the centrality of laugh sounds does not appear to be differentially associated with the presence or absence of harmonic energy. Alternative plots that relied on linear scalings for both *x*- and *y* axes showed that distributions of $F1$ and $F2$ were essentially normal for both call types. This impression was further supported by examining the statistical moments of these distributions, with the tight clustering of observations supporting Bickley and Hunnicutt's (1992) notion of a ''laugh vowel sound.'' Our finding that laugh sounds are consistently found in the central regions of vowel space contrasts with previous speculation that voiced laughter routinely shows vowel-quality distinctions (Provine, 1996, 2000; Provine and Yong, 1991; also see Darwin, 1872/1998; Hall and Allin, 1897; Mowrer *et al.*, 1987; Nwokah *et al.*, 1993; Nwokah *et al.*, 1999; Ruch, 1993). Hypothesizing that ''ha'' is the most prevalent, Provine has for example contended that ''ho'' and ''he'' are also common. In contrast, we observed comparatively few /ɑ/ sounds, and found no /o/ and /i/ variants.

Contrary to expectations, $F4$ frequencies for both call types were essentially the same in both sexes, rather than being higher in females. Even more surprising, female $F5$ values were actually significantly lower than in males for voiced open-mouth calls. Across the spectrum, outcomes for males were largely consistent with those expected of unarticulated sounds, which was also the case for the lowest three formants in females (e.g., Stevens, 1998). In other words, $F4$ and $F5$ outcomes in females must be considered anomalous, in spite of the conservatism of our analyses. The precautions involved included being careful not to overspecify the spectrum by using too many LPC coefficients, comparing the smoothed spectrum to corresponding narrow-band FFT representations in every case, and excluding values in which the purported formant was more likely to be ''tracking'' the energy of an individual harmonic rather than a supralaryngeal

resonance. Difficulties in formant extraction are expected when fundamental frequencies are high (e.g., Titze, Mapes, and Story, 1994), which was certainly the case here. Furthermore, as the energy of the higher harmonics of voiced sounds is typically substantially less than at lower frequencies, the most likely interpretation of the unexpected outcomes obtained for female $F4$ and $F5$ values appears to be that the measurements did not accurately reflect actual production characteristics in these individuals. Alternatively, it may be that the higher resonances are different in laughter than in speech—at least for females. This question is not readily resolvable given the current data, and will therefore be left for future research.

C. Indexical cuing in laughter

Discriminant-function analyses were used to test whether individual laugh calls could be classified according to the sex and individual identity of the person who produced the sound. These questions were of particular interest because acoustic cues to laugher identity have been proposed to play a role in listener responses (e.g., Owren and Bachorowski, 2001a; Smoski and Bachorowski, in press), and because results from voiced, open-mouth calls could be compared to findings from a previous study of an /ɛ/ vowel segment excised from naturally occurring speech (Bachorowski and Owren, 1999). One result of interest in that study was that extremely accurate classification of talker sex occurred using either mean F_0 or the lowest three formant frequencies considered as a set ($F4$ and $F5$ were not included in those analyses). However, when these frequencies were used in combination as an estimator of vocal tract length (VTL), mean F_0 and VTL together provided better overall classification than any other combination of variables.

Current results concerning classification by sex showed both similarities and differences from this earlier work. First, classification accuracy was reasonably high overall (i.e., 72.6% error reduction), but noticeably lower than with the /ɛ/ sound (i.e., 97.2%). Second, entering the $F1$ through $F3$ frequencies as a set again provided accurate classification (i.e., 68.8% error reduction), while F_0 tested alone now had very little power (i.e., 22.4%). The latter outcome is of course to be expected, given the dramatic variability observed in F_0-related measures, regardless of laugher sex. Adding the higher formants neither clarified nor improved these classification outcomes. $F4$ values were not tested because values did not differ according to sex, and classification performance of $F5$ was equivocal. When VTL was calculated using all five formants rather than just the lowest three, classification performance declined accordingly (i.e., 59.0% error reduction). As discussed above, it is simply unknown at this point whether the $F4$ and $F5$ values observed here show laughter to be different from normative speech, or instead reflect the difficulty of obtaining accurate measurements of these formants when high F_0 values are involved. However, one clear conclusion is that $F1$, $F2$, and $F3$ frequencies are primary cues to vocalizer sex, regardless of whether that individual is producing a vowel sound or a voiced, open-mouth laugh sound. The relative unimportance of F_0 in laughter left these formant characteristics as the predominant factor for successful sorting based on these sounds, as was also true for unvoiced open-mouth laughs. This outcome suggests that listeners are able to rely on the same sorts of cues in both instances (see Rendall, Owren, and Rodman, 1998, for related discussion in analogous call types produced by nonhuman primates).

Individual laughers within each sex were less successfully classified (e.g., 39.6% and 49.3% overall error reduction for males and females, respectively), both due to the larger numbers of classes (i.e., laughers) being sorted, and because within-sex acoustics were more similar than between-sex acoustics. Analogous sorting of individual talkers in the earlier study of /ɛ/ vowels also showed reduced accuracy, but there the decline was less precipitous (e.g., 78.6% and 64.3%). Results in both cases nonetheless showed that filter-related cues were again much more important than F_0-related cues in successful classification of individuals, whether male or female. Thus, individually distinctive cues appear to be less prominent in laugh sounds than in vowel sounds, but are nonetheless present in the form of supralaryngeal filtering effects. Classification performance based on formant frequencies in unvoiced, open-mouth calls was similar for males, while somewhat less successful for females. Outcomes for the latter were nonetheless significantly above chance.

D. Theoretical comments

Several aspects of the present findings provide support for our broader theoretical perspective concerning the use and functions of laughter (reviewed in Owren and Bachorowski, 2001a; see also Bachorowski and Owren, 2001; Owren and Bachorowski, 2001b). Drawing on Owren and Rendall's (1997, 2001) model of nonhuman primate vocal signaling, we have proposed that laughter largely functions to elicit emotional responses in listeners and thereby shape their subsequent behavior. Laughter is hypothesized to influence listeners through two mechanisms. For the first, signal acoustics are thought to directly affect listener attention, arousal, and emotional response processes. Laughs with features such as abrupt rise times, high F_0's, perceptually salient F_0 modulation, and perhaps acoustic nonlinearities should be particularly effective in engaging listener response systems. Some empirical support for direct-effect notions comes from the results of perceptual studies, which showed that listeners had significantly more positive emotional responses to voiced than to unvoiced laughs (Bachorowski and Owren, 2001). Further work will more specifically delineate the features and combinations of features that most effectively elicit listener responses. In the meantime, the present results show that many laughs have acoustic features likely to directly ''tweak'' listeners. For the second, more indirect mechanism, learned, positive emotional responses are thought to occur as a result of repeated pairings of the laugher's distinctive acoustics with positive affect occurring in a listener. It was therefore important to find here that both voiced and unvoiced sounds could be statistically classified by individual laugher. Additional work along these lines is thus warranted, for instance testing listener responses to familiar and unfamiliar laugh acoustics.

Another important piece of our theoretical perspective involves sex differences in the use of laughter. We have reported elsewhere that both the rate and selected acoustic features of the laughs analyzed here varied according to social context (Bachorowski *et al.*, 2001; see also Grammer and Eibl-Eibesfeldt, 1990). Overall, those results indicated that variability in individual male laughter is associated with his relationship to his social partner (i.e., friend or stranger), whereas individual female laughter is more closely associated with the sex of her social partner. We interpreted these outcomes to indicate that both males and females use laughter nonconsciously but strategically, in accordance with evolved, sex-based psychological mechanisms (Owren and Bachorowski, 2001a). In this perspective, the remarkable acoustic variability documented here is interpreted as being functionally significant. Individuals of either sex are expected to produce laughs with direct effects on listener response systems when arousal induced in the listener elicits or heightens positive affect, but expected not to when the effect of such arousal is to exacerbate a negative state in that individual (see also Patterson, 1976). Acoustic variability is probably also related to a number of other factors, such as the potency of laugh-eliciting stimuli, individual differences in emotion-based response processes, and sociocultural influences on both the rate and form of signal production. We nonetheless suggest that interactions among laugher sex and social context are likely primary determinants of acoustic variability in laughter.

V. CONCLUSIONS

The data considered here show that laughter is a highly complex vocal signal. A variety of types involving distinct production modes was evident, with song-, grunt-, and snort-like versions being most readily discernible. The observed variability highlights the need for large sample sizes in studying laughter, and suggests that previous work has tended to underestimate the range of acoustic features involved. Although some aspects of laughter were found to resemble speech, most outcomes showed notable differences between the two signals. For instance, voiced laughter showed much more striking source-related variability than is associated with normative speech. Furthermore, while supralaryngeal-filtering effects were as much in evidence in laughter as in speech vowels and sonorants, there was no evidence of analogous articulation. Instead, voiced laughter in American-English speakers overwhelmingly consists of sounds located close to the center of their vowel space. Finally, classification results involving vocalizer sex and individual identity based on the characteristics of individual sounds resembled typical findings from speech in showing filter-related cues to be disproportionately important. However, there was much less of a role for F_0-related features.

Overall, these results stand in contrast to claims that laughter is a stereotyped vocal signal and highlight the difficulty of trying to characterize laughter as being a single acoustic form (e.g., Provine, 1996; Provine and Yong, 1991; cf. Grammer and Eibl-Eibesfeldt, 1990; Rothgänger *et al.*, 1998). Instead, laughter appears to be better conceptualized as a repertoire of sounds, with the prevalence of various subtypes perhaps best gauged by recording laughter that occurs in response to controlled laugh-eliciting stimuli. That approach is also likely to be crucial in eventually understanding the functional importance of the various production modes and the acoustic features associated with them, as we have found in testing variously composed subject dyads (Bachorowski *et al.*, 2001). Important extensions will necessarily involve examining the use of laughter in explicitly interactive circumstances (e.g., Smoski and Bachorowski, in press). Finally, the impact of laugh subtypes on listener responsiveness should be examined through perceptual testing. Other research could include testing the extent to which bout-level temporal patterning is individually distinctive and thereby contributes to indexical cueing (see Owren and Rendall, 1997), examining whether acoustically coherent subtypes occur within the broad categories identified here, and studying the functional importance of unvoiced laughs.

ACKNOWLEDGMENTS

Jo-Anne Bachorowski was supported in part by an NSF POWRE award during acoustic analyses and manuscript preparation. Equipment and funding for data collection and analyses were also provided by funds from Vanderbilt University and Cornell University. Work on portions of this manuscript was completed while the first author was hosted as a Visiting Scholar by the Department of Psychology, Cornell University. We thank Ralph Ohde and Johan Sundberg for their thoughts concerning ''glottal whistles,'' Elizabeth Milligan Spence for her assistance with data collection and digitization of laugh sounds, and Bill Hudenko for his work on inter-rater reliability. Some aspects of this work were presented at meetings of the Acoustical Society of America, Annual Interdisciplinary Conference, and International Society for Research in Emotions.

[1]Formant-frequency values were considered ''missing'' both for instances of harmonic-resonance overlap and for cases in which there was no spectrographic evidence of a resonance in an expected region (see also Hillenbrand *et al.*, 1995). So that multivariate statistics could be used, missing values were replaced with the mean of the relevant formant frequency on a subject-by-subject basis. One should note, however, that this procedure can constrain true variability. Additional precautions were thus taken to ensure that the data set was not unduly influenced by mean replacements. First, preliminary analyses indicated that formant frequencies measured from voiced open-mouth, voiced closed-mouth, vocal fry, and unvoiced open-mouth calls were significantly different from each other. Therefore, replacement values were calculated separately for the four call types. While thus increasing the likelihood of finding differences associated with call-production mode, this approach was preferred because treating all call types as one would create the converse problem of obscuring differences that did exist. Second, statistical outliers were identified on a formant-by-formant basis as those values that were either less than or greater than 3 s.d.s' from the mean for that subject. The 19 cases identified in this fashion were then treated as missing. Third, and again on a formant-by-formant basis, mean replacements were only conducted for instances in which four or more measurements were available and replacements did not account for more than half of a given subject's formant-frequency values. Replacements were not conducted for 17 subjects because too few formant measurements were taken. For the remaining data, a total of 963 replacements was made (i.e., 13% of the observations used in statistical analyses), which was found to change the resulting mean only by approximately 4 Hz. More replacements were made for $F1$ ($n=300$) and $F5$ ($n=281$) than for $F2$ through $F4$ ($n=100$, 132, and 150, respectively).

Bachorowski, J.-A., and Owren, M. J. (**2001**). ''Not all laughs are alike: Voiced but not unvoiced laughter elicits positive affect in listeners,'' Psychological Science **12**, 252–257.
Bachorowski, J.-A., and Owren, M. J. (**1999**). ''Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech,'' J. Acoust. Soc. Am. **106**, 1054–1063.
Bachorowski, J.-A., Smoski, M. J., and Owren, M. J. (**2001**). (unpublished).
Bickley, C., and Hunnicutt, S. (**1992**). ''Acoustic analysis of laughter,'' in Proceedings of the International Conference on Spoken Language Processing 2, 927–930.
Darlington, R. B. (**1990**). *Regression and Linear Models* (McGraw-Hill, New York).
Darwin, C. (**1872/1998**). *The Expression of the Emotions in Man and Animals* (Oxford University Press, New York).
Deacon, T. W. (**1997**). *The symbolic species: The Co-evolution of Language and the Brain* (Norton, New York).
Edmonson, M. S. (**1987**). ''Notes on laughter,'' Anthro. Ling. **29**, 23–33.
Grammer, K. (**1990**). ''Strangers meet: Laughter and nonverbal signs of interest in opposite-sex encounters,'' J. Nonverbal Beh. **14**, 209–236.
Grammer, K., and Eibl-Eibesfeldt, I. (**1990**). ''The ritualization of laughter,'' in *Naturlichkeit der Sprache und der Kultur: Acta colloquii*, edited by W. Koch (Brockmeyer, Bochum, Germany), pp. 192–214.
Hall, G. S., and Allin, A. (**1897**). ''The psychology of tickling, laughing, and the comic,'' Am. J. Psychol. **9**, 1–44.
Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.
Hirschberg, J. (**1999**). ''Dysphonia in infants,'' Int. J. Pediatr. Otorhinolaryngol. **49** (Suppl. 1), S293–S296.
Hollien, H., Dew, D., and Phillips, P. (**1971**). ''Phonation frequency ranges of adults,'' J. Speech Hear. Res. **14**, 755–760.
Katz, G. S., Cohn, J. F., and Moore, C. A. (**1996**). ''A combination of vocal f_0 dynamic and summary features discriminates between three pragmatic categories of infant-directed speech,'' Child Dev. **67**, 205–217.
Ladefoged, P. (**1993**). *A Course in Phonetics*, 3rd ed. (Harcourt Brace, New York).
Lieberman, P., and Blumstein, S. E. (**1993**). *Speech Physiology, Speech Perception, and Acoustic Phonetics* (University Press, Cambridge, MA).
Mende, W., Herzel, H., and Wermke, K. (**1990**). ''Bifurcations and chaos in newborn infant Cries,'' Phys. Lett. A **145**, 418–424.
Milford, P. A. (**1980**). ''Perception of laughter and its acoustical properties,'' unpublished doctoral dissertation, Pennsylvania State University.
Mowrer, D. E., LaPointe, L. L., and Case, J. (**1987**). ''Analysis of five acoustic correlates of laughter,'' J. Nonverbal Beh. **11**, 191–199.
Nwokah, E. E., Davies, P., Islam, A., Hsu, H. C., and Fogel, A. (**1993**). ''Vocal effect in 3-year-olds: A quantitative acoustic analysis of child laughter,'' J. Acoust. Soc. Am. **94**, 3076–3090.
Nwokah, E. E., Hsu, H. C., Davies, P., and Fogel, A. (**1999**). ''The integration of laughter and speech in vocal communication: A dynamic systems perspective,'' J. Speech Lang Hear Res. **42**, 880–894.
Olive, J. P., Greenwood, A., and Coleman, J. (**1993**). *Acoustics of American English Speech* (Springer, New York).
Owren, M. J. (**2001**). ''Acoustic evidence of stable and unstable vocal-fold action in rhesus monkey vocalizations,'' XVIII Congress of the International Primatological Society, Adelaide, Australia.
Owren, M. J., and Bachorowski, J.-A. (**2001a**). (unpublished).
Owren, M. J., and Bachorowski, J.-A. (**2001b**). ''The evolution of emotional expression: A 'selfish-gene' account of smiling and laughter in early hominids and humans,'' in *Emotions: Current Issues and Future Directions*, edited by T. J. Mayne and G. A. Bonanno (Guilford, New York), pp. 152–191.
Owren, M. J., and Casale, T. M. (**1994**). ''Variations in fundamental frequency peak position in Japanese macaque (*Macaca fuscata*) coo calls,'' J. Comp. Psych. **108**, 291–297.
Owren, M. J., and Rendall, D. (**1997**). ''An affect-conditioning model of nonhuman primate vocal signaling,'' in *Perspectives in Ethology: Volume 12. Communication*, edited by D. H. Owings, M. D. Beecher, and N. S. Thompson (Plenum, New York), pp. 299–346.
Owren, M. J., and Rendall, D. (**2001**). ''Sound on the rebound: Bringing form and function back to the forefront in understanding nonhuman primate vocal signaling,'' Evo. Anthro. **10**, 58–71.
Owren, M. J., Seyfarth, R. M., and Cheney, D. L. (**1997**). ''The acoustic features of vowel-like *grunt* calls in chacma baboons (*Papio cyncephalus ursinus*): Implications for production processes and functions,'' J. Acoust. Soc. Am. **101**, 2951–2962.
Patterson, M. L. (**1976**). ''An arousal model of interpersonal intimacy,'' Psychol. Rev. **83**, 235–245.
Peterson, G. E., and Barney, H. L. (**1952**). ''Control methods used in a study of the vowels,'' J. Acoust. Soc. Am. **24**, 175–184.
Provine, R. R. (**1996**). ''Laughter,'' Am. Sci. **84**, 38–45.
Provine, R. R. (**2000**). *Laughter: A Scientific Investigation* (Penguin, New York).
Provine, R. R., and Yong, Y. L. (**1991**). ''Laughter: A stereotyped human vocalization,'' Ethology **89**, 115–124.
Ptacek, P. H., Sander, E. K., Maloney, W. H., and Jackson, C. C. (**1966**). ''Phonatory and related changes with advanced age,'' J. Speech Hear. Res. **9**, 353–360.
Pullum, G. K., and Ladusaw, W. A. (**1996**). *Phonetic Symbol Guide*, 2nd ed. (University of Chicago Press, Chicago).
Rendall, D., Owren, M. J., and Rodman, P. S. (**1998**). ''The role of vocal tract filtering in identity cueing in rhesus monkey (*Mocaca mulatta*) vocalizations,'' J. Acoust. Soc. Am. **103**, 602–614.
Riede, T., Herzel, H., Mehwald, D., Seidner, W., Trumler, E., Böhme, G., and Tembrock, G. (**2000**). ''Nonlinear phenomena in the natural howling of a dog–wolf mix,'' J. Acoust. Soc. Am. **108**, 1435–1442.
Riede, T., Wilden, I., and Tembrock, G. (**1997**). ''Subharmonics, biphonation, and frequency jumps—common components of mammalian vocalization or indicators for disorders?,'' Z. f. Saeugetierkunde **62** (Suppl. 2), 198–203.
Rothgänger, H., Hauser, G., Cappellini, A. C., and Guidotti, A. (**1998**). ''Analysis of laughter and speech sounds in Italian and German students,'' Naturwissenschaften **85**, 394–402.
Ruch, W. (**1993**). ''Exhilaration and humor,'' in *Handbook of Emotions*, edited by M. Lewis and J. M. Haviland (Guilford, New York), pp. 605–616.
Ruch, W., and Ekman, P. (**2001**). ''The expressive pattern of laughter,'' in *Emotion, Qualia, and Consciousness*, edited by A. W. Kaszniak (World Scientific, Tokyo), pp. 426–443.
Shanks, S. J. (**1970**). ''Effect of aging upon rapid syllable production,'' Percept. Motor Skills **30**, 687–690.
Smoski, M. J., and Bachorowski, J.-A. (**in press**). ''Antiphonal laughter between friends and strangers,'' Cognition & Emotion.
Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge).
Struhsaker, T. T. (**1967**). ''Auditory communication among vervet monkeys (*Cercopithecus aethiops*),'' in *Social Communication among Primates*, edited by S. A. Altmann (University of Chicago, Chicago), pp. 281–324.
Švec, J. G., Horáček, J., Sram, F., and Veselý, J. (**2000**). ''Resonance properties of the vocal folds: *In vivo* laryngoscopic investigation of the externally excited laryngeal vibrations,'' J. Acoust. Soc. Am. **108**, 1397–1407.
Tabachnik, B. G., and Fidell, L. S. (**1996**). *Using Multivariate Statistics* (HarperCollins, New York).
t' Hart, J., Collier, R., and Cohen, A. (**1990**). *A Perceptual Study of Intonation* (Cambridge University Press, Cambridge).
Titze, I. R., Mapes, S., and Story, B. (**1994**). ''Acoustics of the high tenor voice,'' J. Acoust. Soc. Am. **95**, 1133–1142.
Venkatagiri, H. S. (**1999**). ''Clinical measurement of rate of reading and discourse in young adults,'' J. Fluency Disord. **24**, 209–226.
Wilden, I., Herzel, H., Peters, G., and Tembrock, G. (**1998**). ''Subharmonics, biphonation, and deterministic chaos in mammal vocalization,'' Bioacoustics **9**, 171–196.

Challenging the notion of innate phonetic boundaries

Susan Nittrouer
Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

(Received 27 September 2000; accepted for publication 17 April 2001)

Numerous studies of infants' speech perception abilities have demonstrated that these young listeners have access to acoustic detail in the speech signal. Because these studies have used stimuli that could be described in terms of adult-defined phonetic categories, authors have concluded that infants innately recognize stimuli as members of these categories, as adults do. In fact, the predominant, current view of speech perception holds that infants are born with sensitivities for the universal set of phonetic boundaries, and that those boundaries supported by the ambient language are maintained, while those not supported by the ambient language dissolve. In this study, discrimination abilities of 46 infants and 75 3-year-olds were measured for several phonetic contrasts occurring in their native language, using natural and synthetic speech. The proportion of children who were able to discriminate any given contrast varied across contrasts, and no one contrast was discriminated by anything close to all of the children. While these results did not differ from those reported by others, the interpretation here is that we should reconsider the notion of innate phonetic categories and/or boundaries. Moreover, success rates did not differ for natural and synthetic speech, and so a minor conclusion was that children are not adversely affected by the use of synthetic stimuli in speech experiments. [DOI: 10.1121/1.1379078]

PACS numbers: 43.71.Ft, 43.71.An, 43.71.Pc [CWT]

I. INTRODUCTION

In 1971, Eimas, Siqueland, Jusczyk, and Vigorito reported that one- and four-month-old infants could discriminate between two synthetic stop-vowel syllables that differed along an acoustic dimension associated with the voicing of initial stop consonants. The voicing of initial stops is usually described by voice onset time (VOT), which is the start time of vocal-fold vibration relative to the release of closure (Lisker and Abramson, 1964). Positive values of VOT indicate that vocal-fold vibration started after the release of closure, while negative VOT values indicate that vocal-fold vibration preceded release. The acoustic correlate of VOT manipulated by Eimas *et al.* was $F1$-cutback, which is the time of first formant ($F1$) onset relative to the release of closure. As with English-speaking adults, these infants were found to discriminate between stimuli with VOTs of +20 ms and +40 ms. These VOTs placed the syllables on opposite sides of the phoneme boundary for English /p/ and /b/. When a 20-ms difference between syllables was used that placed both stimuli on the same side of the phoneme boundary, infants failed to discriminate between them, as adults fail to do. From these results, Eimas *et al.* concluded that infants are sensitive to the acoustic dimension that defines adult voicing categories, even before they have experience with language.

That report sparked a great deal of research over the next two decades investigating infants' capacities for speech perception. The collective conclusion of these many studies was that infants approximately nine months of age or younger were able to discriminate virtually all phonetic contrasts presented to them, regardless of whether or not the contrasts were in the infant's native language (e.g., Werker, 1991). This result was demonstrated with natural and synthetic stimuli, across a range of contrasts (e.g., Eilers *et al.*, 1982, 1977; Kuhl, 1979b; Moffitt, 1971; Morse, 1972; Streeter, 1976). Although not investigated as frequently, evidence was also found to support the second of Eimas *et al.*'s results, that infants fail to discriminate within-category acoustic differences. Unlike the between-category experiments, this kind of test can be conducted only with synthetic stimuli. When stimuli differ by the same acoustic distance as a between-category pair, but both fall within the same category, infants fail to discriminate them (e.g., Aslin *et al.*, 1981; Eimas, 1974, 1975).

Those early studies of infant speech perception led to the widely accepted view that infants are born with sensitivities to phonetic boundaries for all languages (i.e., the universal set). Experience listening to a native language during the first year of life, the theory holds, maintains those boundaries supported by the ambient language, and causes those boundaries not supported by the ambient language to dissolve. This view of perceptual development is what Aslin and Pisoni (1980) call a ''universal'' theory. Reviews of the work supporting this theoretical position are numerous (e.g., Eimas *et al.*, 1987; Jusczyk, 1995; Kuhl, 1979a, 1987; Morse, 1985; Werker, 1989). Even if only by default it has become the predominant theory of infant speech perception.

As early as the 1970s, however, there were a few discrepant findings that presented some challenge to the intransigent nature of the speech processing mechanism suggested by this model. First, it was found that some phonetic contrasts were discriminated more readily than others by infants. For example, Holmberg *et al.* (1977) counted the number of trials required for 6-month-olds to learn to discriminate pairs of stimuli. Using a criterion of eight correct responses out of ten consecutive trials (half change and half no-change),

Holmberg *et al.* reported that it required 64 trials, on average, for 6-month-olds to meet the criterion for /f/ versus /θ/, but only 33 trials, on average, for /s/ versus /ʃ/. Similarly, Eilers *et al.* (1977) found that infants were able to discriminate some contrasts (out of the ten presented to them), but not others.

Even when infants were found to discriminate stimuli, responses were not always strictly categorical. For example, Eimas and Miller (1980) found that 2- to 4-month-olds were able to discriminate synthetic tokens located on the same side of a /b/-to-/m/ boundary better than would be expected if perception was strictly categorical. Furthermore, even when responses were of a categorical nature, category boundaries did not always appear where they would be expected. For example, Lasky *et al.* (1975) investigated the abilities of infants in a Spanish language environment to discriminate three voicing contrasts for syllable-initial stops. One contrast placed stimuli on opposite sides of the Spanish VOT boundary: −20 ms versus +20 ms. Two contrasts placed stimuli within a Spanish voicing category: −60 ms versus −20 ms VOT and +20 ms versus +60 ms VOT. A particularly important manipulation in this study was that the +20 ms versus +60 ms contrast placed stimuli on opposite sides of the English VOT boundary. The infants in Lasky *et al.*'s study (ages 4 to $6\frac{1}{2}$ months) discriminated this contrast, even though it was not in their native language. They also discriminated the −60 ms versus −20 ms contrast, even though both of these stimuli fall within the Spanish voiced category. In fact, the one contrast they failed to discriminate was the −20 ms versus +20 ms, which crosses the Spanish voicing boundary. In spite of these seemingly contradictory findings, however, the notion of innate phonetic boundaries has persisted.

Of course, descriptions offered by various authors differ somewhat, particularly with respect to whether the focus is on the boundary or on the contrast. For example, Jusczyk (1995) writes ''Findings of this sort [as those described at the outset] have led to the view that infants are born with the capacity to discriminate contrasts that could potentially appear in any of the world's languages. Experience with language appears to have its impact by getting the infant to focus on those contrasts that play a critical role in distinguishing words in the native language.'' (p. 269) Similarly Best (1994) states ''Current findings suggest that infants begin life with language-universal abilities for discriminating segmental phonetic contrasts but that, by the second half-year of life, listening experience with the native language has begun to influence the perception of contrasts that are nondistinctive in the native phonological system.'' (p. 168) Kuhl (e.g., 1979b, 1980; Kuhl and Miller, 1982) reminds us repeatedly that an important component of any theory of innate capacities for speech-sound categorization must be a demonstration of similarity judgments for acoustically disparate members of a category. Also, there have been various modifications of the basic view. For example, Werker (1994) suggests that the loss of non-native boundaries may not be a permanent loss, as originally thought, and that boundaries may differ in how long it takes them to dissolve. Jusczyk (1998) specifically invokes the Lasky *et al.* (1975) finding to suggest that rather than there being innate phonetic categories that line up with the categories of specific languages, perhaps there is a ''language-general categorization of speech information'' (p. 56). According to this view, the ability of the infants in the Lasky *et al.* study to discriminate the English voicing contrast, but not the Spanish, can be explained as evidence that the English voicing contrast comes closer to infants' innate perceptual boundaries than does the Spanish contrast. Kuhl proposes a model in which ''acoustic space'' is linear at birth, but the regions around phonetic boundaries become warped as a result of language experience during the first six months of life (Grieser and Kuhl, 1989; Kuhl, 1991, 1993). These variations, however, fail to contradict the basic tenets of the universal theory. The predominant view continues to rely on notions of innate, universal boundaries as the starting point for human speech perception, with some form of loss as the main mechanism for developmental change.

The purpose of this brief report is to encourage reconsideration of the ''universal'' theory as it applies to infant speech perception. The experiment reported here evolved from efforts in this laboratory to extend findings and hypotheses concerning the speech perception of children roughly $3\frac{1}{2}$ to 7 years to even younger listeners. This work has shown that, at least for some phonetic distinctions, children in this age range weight the various acoustic properties upon which phonetic decisions are made differently than adults do (Nittrouer, 1992, 1996; Nittrouer *et al.*, 1998, 2000; Nittrouer and Miller, 1997a, b; Nittrouer and Studdert-Kennedy, 1987). This finding has been corroborated by others (e.g., Greenlee, 1980; Krause, 1982; Morrongiello *et al.*, 1984; Parnell and Amerman, 1978; Wardrip-Fruin and Peach, 1984), who also report that when making the same phonetic decision, children, compared to adults, pay more attention to some acoustic properties and less attention to others. Combining these two general findings (that infants are born with innate phonetic boundaries and that children weight acoustic properties differently from adults in making phonetic decisions) led to an apparent contradiction: If indeed infants are born with capacities to recognize all the phonetic contrasts in their native language (i.e., the mechanism of maintenance alone accounts for their presence into childhood), how is it that differences in phonetic decision-making are observed for children and adults? The hypothesis that emerged was that perhaps infants' discrimination abilities are based on different weighting strategies than those of adults. That is, even though infants make the same discriminations as adults, the way that they come to make these discriminations could be different. Thus efforts were undertaken to examine the relative weighting of acoustic properties in infants' phonetic decisions.

The initial assumption was that infants would surely be able to discriminate the contrast of interest. Much of the work examining developmental shifts in perceptual weighting strategies has been done using /s/-vowel versus /ʃ/-vowel contrasts, and Holmberg *et al.* (1977) showed that this contrast is well within the capabilities of 6-month-olds to discriminate. So that contrast was selected for use with infants. Quickly, however, it became clear that infants could not

readily make this discrimination, even when all properties covaried appropriately (e.g., when natural tokens were used). As a result, other contrasts were introduced that were presumed to be even more discriminable: specifically, two vowels from the corners of the vowel triangle and a VOT contrast. The focus of the study then shifted to examining the ''universal theory.''

A secondary goal of this work was to determine if infants perform differently in their perception of speech when synthetic signals are used instead of natural tokens. This question arose largely from concern expressed informally (e.g., in manuscript reviews) that children may perform differently from adults in tests using synthetic speech because children have less experience hearing these signals. The implication of such statements has clearly been that children perform worse with synthetic stimuli than they would with natural stimuli. Eilers *et al.* (1977), on the other hand, suggested that perhaps infants perform better with synthetic stimuli than they would with natural stimuli because there is no variation in acoustic properties across stimuli, save the one on which the discrimination must be made.

Finally, in addition to infants, children just 3 years of age participated in this study. In earlier work we have tested children no younger than $3\frac{1}{2}$ years largely because younger children do not perform well on the labeling tasks used. However, in addition to difficulty with the task, it has appeared that some of the poor performance of children in this age range may actually be due to phonetic categories that are more poorly specified for children than for adults. We wanted to explore that possibility with a discrimination task, using just slightly younger children than in our labeling experiments.

II. METHOD

A. Participants

Two groups of children participated: 46 infants and 75 preschoolers. Infants were between the ages of 6 and 14 months. Several investigators have successfully used headturning procedures with infants up to 14 months of age (e.g., Eilers *et al.*, 1977; Moore *et al.*, 1975), and Kuhl (1985) states that it is appropriate for infants between 5.5 and 18 months. Preschoolers were between 2 years, 6 months and 3 years, 4 months. All children (infants and preschoolers) were full-term births, with no prenatal or perinatal histories that would put them at risk for language problems. All children lived in homes with English as the only language. Children were excluded if they had a sibling or parent with a speech or language problem, or if they were not developing speech as expected. Specifically parents were asked about two well-recognized milestones, if the infant or preschooler was old enough to have reached the milestone. Children must have shown evidence of canonical babbling by nine months of age and must have started using two-word utterances by two years. All children were free from significant histories of otitis media, defined as having no more than one episode during the first year of life and no more than three episodes total. All children passed a hearing screening of the frequencies 0.5, 1.0, 2.0, 4.0, and 6.0 kHz presented free-field at 25 dB HL using either a visually reinforced headturning procedure (infants) or play audiometry (preschoolers).

B. Stimuli

All stimuli were digitized at a 20-kHz sampling rate, and low-pass filtered at 10 kHz.

1. Natural stimuli

Five sets of natural stimuli were made. Two sets consisted of stimuli that differed only in the vowel, either /sɑ/ versus /su/ or /ʃɑ/ versus /ʃu/. Because relatively stable regions of spectral information signal these contrasts they should be readily discriminated. One set of stimuli consisted of syllables with initial alveolar stops that differed in VOT, /tɑ/ versus /dɑ/. This voicing contrast was included because it has a long and thorough history of investigation, dating back to the seminal work of Lisker and Abramson (1964), and it is widely accepted that infants (e.g., Eimas *et al.*, 1971) as well as nonhuman animals (e.g., Kuhl and Miller, 1978) can make VOT discriminations. Two sets of stimuli differed only in fricative place of constriction, /sɑ/ versus /ʃɑ/ and /su/ versus /ʃu/. Much of the earlier work in this laboratory with $3\frac{1}{2}$- to 7-year-olds has focused on the /s/-/ʃ/ contrast, and so we were particularly anxious to investigate this contrast with younger listeners. For each set, three tokens of each syllable were obtained from the same speaker. Using multiple tokens made it possible to present stimuli with a roving standard and a roving comparison to ensure that discrimination was based on phonetic change, rather than on changes in other, irrelevant acoustic properties. At the same time, restricting samples to those from one speaker minimized factors that could interfere with the encoding of speech by listeners (Jusczyk *et al.*, 1992).

2. Synthetic stimuli

Synthetic versions of /su/ versus /ʃu/ and /sɑ/ versus /ʃɑ/ were prepared using a Klatt (1980) software synthesizer. The fricative noises have been used in other labeling experiments (Nittrouer, 1992, 1996; Nittrouer and Miller, 1997a, b), had a single pole, and were 230 ms long. The center frequency of the /s/ noise was 3.8 kHz, and the center frequency of the /ʃ/ noise was 2.2 kHz. Vocalic portions were 270 ms long, and two portions were synthesized for each vowel: one with a second-formant ($F2$) transition appropriate for /ʃ/ and one with an $F2$ transition appropriate for /s/. For the two /u/ portions, fundamental frequency ($f0$) started at 120 Hz and fell throughout to an ending frequency of 100 Hz. The $F1$ was constant at 250 Hz, and the third formant ($F3$) was constant at 2100 Hz. Because $F3$ was similar in frequency to the pole of the /ʃ/ noise, energy was present in this frequency region across the entire syllable when the /ʃ/ noise was used, but not when the /s/ noise was used. Stevens (1985) has suggested that one cue to fricative identity for /s/ versus /ʃ/ is the amount of amplitude change in the $F3$ region across the noise/voicing boundary, and this cue was appropriately manipulated here. For both /u/ portions, $F2$ fell through the entire portion to an ending frequency of 850 Hz. For /(s)u/, $F2$ started at 1600 Hz; for /(ʃ)u/, it started at 1800 Hz.[1]

For the two /ɑ/ portions, $f0$ started at 100 Hz, and fell through the portion to 80 Hz. For both /ɑ/ portions, $F1$ started at 450 Hz and rose over the first 50 ms to a steady-state frequency of 650 Hz. $F3$ remained constant at 2500 Hz. Again this setting maintained the relative amplitude cue in the $F3$ region described by Stevens (1985). For both /ɑ/ portions, $F2$ fell over the first 100 ms to a steady-state frequency of 1130 Hz. For the /(s)ɑ/ portion, $F2$ started at 1300 Hz; for /(ʃ)ɑ/, $F2$ started at 1500 Hz.

3. Hybrid stimuli

The vocalic portion of each of the three tokens of each natural fricative-vowel syllable was separated from the fricative noise, and combined with the synthetic /s/ and /ʃ/ noise such that the place of constriction specified by formant transitions matched that specified by the noise. Although one purpose of this study was to compare discrimination of natural and synthetic stimuli, these hybrid stimuli were also included because such tokens are frequently used in our testing with children.

C. Equipment

All testing took place in a sound-attenuated chamber. A one-way window connected the chamber with an adjacent control room. A Madsen audiometer was used to screen hearing. Speech stimuli were presented free field using a computer, a Data Translation 2801A digital-to-analog converter, a Frequency Devices 901F filter, a Tascam PA30-B amplifier, and a JBL Control-1 speaker. A special purpose board with two boxes attached to it controlled the presentation of stimuli, recorded responses, and turned on reinforcers. One box had three foot pedals attached to it which allowed the experimenter in the chamber with the infant or preschooler to start the presentation of standard stimuli, initiate trials of comparison stimuli, and temporarily interrupt the presentation of all stimuli without the child or the second experimenter observing pedal presses.[2] The other box had four buttons, and was in the control room. By pressing any one of the buttons, the experimenter in the control room recorded that a response had occurred and presented reinforcement. Reinforcement was provided by one of three Plexiglas boxes, each containing a mechanical animal, or by a graphics monitor that displayed brightly colored shapes. A total of ten mechanical animals was kept in stock, so they could be replaced between visits for any one child. For infants, a supply of quiet toys helped maintain forward eye gaze between trials. For preschoolers, reinforcement was contingent on the press of a large button, mounted on a board. The button was not connected to anything, but when the child pressed it the experimenter in the control room recorded the response.

D. Procedures

Procedures were very similar to those of most studies using a visually reinforced headturning procedure (e.g., Kuhl, 1985; Werker and Tees, 1984). One experimenter (E1) was in the chamber with the child and the child's parent. The arrangement in the chamber (shown in Fig. 1) was modified slightly from that of other investigators, who often have the child sit on the parent's lap. We found that children were less

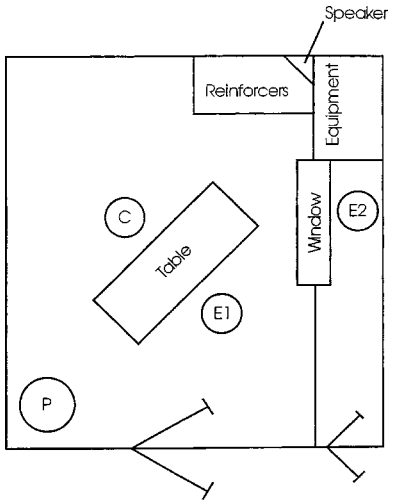


FIG. 1. Arrangement of booth during testing. C=child; P=parent; E1 and E2=experimenters.

restless if they sat in a seat by themselves (a table-mounted chair for infants; a high chair for preschoolers).[3] The child sat across a table from E1, with the parent well off to the side. The speaker and reinforcers were on the opposite side of the table from the parent. A second experimenter (E2) was in the control room.

The parent listened the entire time she was in the chamber to monologues by a male radio personality (Garrison Keillor), presented over headphones. We found that at comfortable listening levels these monologues more effectively masked the stimuli being presented than did music, probably because the $f0$'s of Garrison Keillor and of the stimuli (whether natural or synthetic) were similar. E1 listened to the stimuli during training phases, but listened to the monologues during testing. E2 listened to the stimuli during training, but then simply switched off the speaker in the control room so that stimuli were not heard during testing.

E1 used the foot pedals to initiate the presentation of the standard stimulus, and to introduce trials. Stimuli were presented at a peak intensity of 68 dB SPL, at a rate of one every 2 s. The presentation level was selected based both on Dobie and Berlin's (1979) report that a normal conversational level is between 65 and 70 dB SPL, and on Nozza's (1987) demonstration that infants' discriminations were more successful at a level close to 70 dB, rather than at the lower levels of 50 to 60 dB commonly used (e.g., Eilers *et al.*, 1977; Werker *et al.*, 1981). Three stimuli were presented during each comparison trial (i.e., 6-s trials), and stimulus presentation returned to the presentation of the standard if no response occurred. E2 pressed a button in the control room if she judged that a response had occurred (i.e., headturns on the part of infants or button-presses on the part of preschoolers). If the response was to a change trial, reinforcement lasting 3 s was presented.

Children were scheduled to participate in up to four sessions, over as many consecutive days.[4] One contrast was presented per session, although no one child received more than two phonetically different contrasts (see later in this work). During training, all trials were change trials. For preschoolers, training was straightforward: the task was ex-

plained, and they were given the training trials. For infants, training procedures required traditional conditioning. Initially, reinforcement was presented after the first presentation of the stimulus during the change trial (causing the infant to turn to look), but gradually the presentation of reinforcement became contingent on a headturn. To pass training, a child had to respond to three consecutive change trials with no prompting. Twenty trials were provided in which to meet this criterion. Again, the fricative contrast was anticipated to be the most difficult of the three contrasts used, and Holmberg *et al.* (1977) reported that it required an average of 11.2 trials for infants to train on this contrast. Consequently, it seemed reasonable to expect infants and preschoolers to train on these contrasts within the 20-trial limit, if they were going to train at all. For both infants and preschoolers, the training phase was also used to decrease the probability of false positives. For those children who initially demonstrated frequent false-positive responses, the interval between change trials was deliberately lengthened, thus diminishing those responses (Werker and Tees, 1984).

During testing, 15 trials were presented: ten change and five no-change. The criterion for passing a test phase was to get eight of ten correct responses to change trials, with no more than one response during no-change trials. This criterion is similar to that of Werker and Tees (1984).

Only E2 had a vote in deciding if a response had occurred. This experimenter was unaware of when a trial was occurring because that was controlled by E1. In many studies with infants, two experimenters must judge that a headturn occurred for it to be considered a response, although the use of just one judge for these decisions is not novel (e.g., Hirsch-Pasek *et al.*, 1987; Werker and Tees, 1984). The decision to base reinforcement on the judgment of a headturn by just one experimenter was related to the choice of ratios for change/no-change trials. We used a 2/1 ratio of change/no-change trials, instead of the more common 1/1 ratio (i.e., five change and five no-change trials) (Kuhl, 1985).[5] We chose to implement the higher ratio of change/no-change trials because the use of a 1/1 ratio gives the same weight to a lack of a headturn for a no-change trial as to a headturn for a change trial. Thus, even if 90% correct responses are required to satisfy the test criterion, a child need only respond to four changes (out of ten trials) to satisfy that criterion; that is, simply failing to turn one's head in the presence of five no-change trials would count as five correct responses. We wanted to see stronger evidence of the child responding in the presence of a change in stimulus. However, there was one drawback to using this stricter criterion. With this higher ratio of change/no-change trials, the probability of a trial being a change trial increased, so the experimenter who knew if a trial was occurring (E1) might be biased to vote that a headturn had occurred. Thus, that experimenter did not get a vote.

1. Infants

Half the infants heard a vowel contrast first and half heard the VOT contrast first. Of the infants hearing the vowel contrast first, half of them heard the contrast with syllable-initial /ʃ/ and half heard it with /s/. Also, the vowel that was the standard and the vowel that was the comparison (/ɑ/ or /u/) varied across infants. Those infants who met the test criterion for this first contrast were next presented with a fricative contrast using natural tokens, on the second day. For those infants who first heard a vowel contrast, the syllable that had served as the comparison for that contrast remained the comparison for the fricative contrast. For example, if an infant heard the vowel contrast /sɑ/ versus /su/ (with /su/ as the comparison), then the infant heard the fricative contrast /ʃu/ versus /su/ (again with /su/ as the comparison). This was done because it cannot be known whether the child is responding to a change in stimuli, or to the presence of the stimulus associated with reinforcement. If the latter, contingencies would not change for the infant from the vowel contrast to the fricative contrast. The fricative contrasts presented to infants hearing the VOT contrast first were randomly assigned. Every infant who met the test criterion for the natural fricative contrast was subsequently presented with a synthetic fricative contrast, on the third day, and the synthetic fricative-vowel syllables used with any one child remained the same as those of the natural contrast. Infants who met the test criterion for their first contrast, but failed to meet it for the natural fricative contrast, came back on the third day to repeat the first contrast. Infants who met the test criterion for these synthetic fricative stimuli were dismissed. Infants who did not meet the criterion would return for a fourth day, to be retested with the natural fricative stimuli.

2. Preschoolers

The focus of investigation with preschoolers was on fricative perception, and so the VOT contrast was not used. All preschoolers heard a vowel contrast on the first day of testing. Those who met the test criterion with vowels were presented with a fricative contrast on the second day. As with infants hearing the vowel contrast first, testing for any one child was planned so that the comparison stimulus remained the same across all contrasts. The kind of fricative stimuli first presented (natural, synthetic, or hybrid) was randomly varied across preschoolers. Children who heard natural or hybrid stimuli for the first fricative contrast, and met the test criterion, returned for a third day of testing with the synthetic fricative contrast. In this way we could ask if synthetic speech *per se* presents problems for children. As with infants, preschoolers were retested with the last contrast on which they were successful, if they failed to meet the criterion for a contrast.

For all children then, except those infants hearing the VOT contrast first, the phonetic structure of the stimulus associated with reinforcement remained constant across all conditions.

III. RESULTS

A. Infants

Of the 23 infants tested with a vowel contrast, 15 (65%) met the test criterion.[6] Of the 23 infants tested with the VOT contrast, eight (35%) met the test criterion. Of the 15 children who were able to do the vowel contrast, six (40%) were

subsequently successful with the natural fricative contrast. The nine who were not successful with the fricative contrast were all able to meet the test criterion for the vowel contrast when retested. All six infants who could discriminate the natural fricative contrast were also able to discriminate the synthetic fricatives. None of the eight infants who were successful with the VOT contrast were able to discriminate the natural fricatives, but all were able to discriminate the VOT contrast when retested.

B. Preschoolers

Forty-two preschoolers (56%) were successful with the vowel contrast. One of those children did not participate in further testing due to illness. Of the remaining 41 children tested with one of the three fricative contrasts, 24 (59%) were successful. There were no differences among the proportions of children who succeeded with the natural, synthetic, and hybrid stimuli. The 17 preschoolers who did not discriminate the fricative contrast were able to perform the vowel contrast when retested. All children tested with the natural or hybrid stimuli first were subsequently able to discriminate the fricative contrast with synthetic stimuli.

IV. DISCUSSION

The data reported here were originally collected as part of what was to be a pilot experiment, developing methods for investigating the weighting strategies of infants for the various acoustic properties that define linguistic segments. In line with the work that has been done with $3\frac{1}{2}$- to 7-year-olds, the plan was to manipulate the acoustic structure of fricative-vowel syllables to examine whether infants base discrimination judgments more on differences in the fricative noise or on differences in formant transitions. However, the principal experiment was never conducted because infants and 3-year-olds demonstrated unreliable results discriminating even clear tokens of fricative-vowel syllables, regardless of whether the fricative or vowel differed within the pair, as well as unreliable results for stimuli differing in voicing.[7] To ask the question of how much weight is given to each acoustic property in discrimination decisions would require manipulations of the stimuli that would make them somewhat perceptually ambiguous. There is every reason to believe that such manipulations would render stimuli undiscriminable, even for the infants and 3-year-olds who did discriminate these clear tokens.

What is left then are these sparse data demonstrating how difficult it is for infants, and even children as old as 3 years, to discriminate speech stimuli based on phonetic category. It is, of course, tempting to dismiss these results by suggesting that the success rates were low because of poor procedures. However, the success rates reported here are not different from those reported by others who report success rates for infants. Jusczyk and colleagues always report attrition, and they generally dismiss 40% to 45% of the infants in their studies due to ''fussiness'' (e.g., Bertoncini *et al.*, 1988; Jusczyk *et al.*, 1992; Levitt *et al.*, 1988). (Other infants may be dismissed for other reasons, as well.) The implicit assumption of that work is that infants dismissed due to fussiness would have had similar success rates as those of the nonfussy infants, if only they had not been fussy. In our work with children $3\frac{1}{2}$ to 7 years of age, however, we have not found that to be the case. We find that children may become uncooperative precisely because they cannot discriminate the stimuli presented: If these children return to the laboratory on a different day they usually become uncooperative with the same or similar stimuli, but if presented with stimuli that are not minimal pairs, these same children cooperate and perform the task appropriately. While we cannot conclusively draw a parallel from those findings with older children to work with infants, it would be inappropriate to assume that the dismissed infants would have performed as the infants who were not dismissed.

Another reason to discount the notion that procedures might have been nonoptimal, accounting for the low success rate, is that success rates differed across contrasts. If procedures accounted for a large proportion of variance in success rates, we would have expected those rates to be similar across contrasts. Of particular interest was the low proportion of infants who reached criterion on the VOT contrast. The English /dɑ/ versus /tɑ/ contrast has been used extensively to support the argument that the auditory system provides regions of enhanced sensitivity along some psychophysical continua, and those regions form natural boundaries between phonetic classes (e.g., Kuhl, 1981; Kuhl and Miller, 1978; Sinex and McDonald, 1989; Sinex *et al.*, 1991). This study was unable to address the notion of enhanced sensitivity, but clearly infants were not as successful at discriminating stimuli differing in VOT as the notion suggests they should have been.

Finally, the fact is that it is simply not that difficult to institute a headturning procedure with infants or a button-pressing procedure with 3-year-olds. These procedures are used routinely in audiology clinics to measure auditory thresholds in infants and 3-year-olds. In those settings, the procedures have proven to be fairly robust to variations in procedures, and so it was that screening children's hearing in this study presented no problems.

A minor conclusion drawn from this work was that infants and preschoolers are perfectly capable of perceiving synthetic speech. There was not one instance in which a child was able to discriminate a contrast with natural or hybrid stimuli, but unable to do so with synthetic stimuli.

Overall these results fail to provide support for claims that universal phonetic boundaries are in place at birth. It is emphasized that the findings of this study do not really differ from those of others: success rates are similar across studies. What differs is the willingness of authors to use the results to support claims of innate phonetic boundaries. In fact, some earlier studies provide evidence that could be taken to refute overtly such claims. Again, Eimas and Miller (1980) found that infants could discriminate between stimuli that both fell within an adult phonetic category, and Lasky *et al.* (1975) found that infants failed to discriminate between tokens that fell into different categories. In sum, the data across experiments do not support the proposition that infants have clearly established phonetic categories, separated by well-defined boundaries. As a field, it is important for us to bear this point

in mind because the nature of the processing that we attribute (or fail to attribute) to infants affects the questions we ask about the speech perception of older listeners. Perhaps we should not even be asking if infants have well-formed phonetic categories, separated by boundaries, but rather if any language users do. In other words, the very concept of categories, and even more so of boundaries, needs to be reconsidered.

The concept of phonetic boundaries arose from the early categorical studies using synthetic signals in which articulatorily and acoustically impossible speech sounds were constructed by manipulating a single dimension of the signal, while holding all other dimensions constant. Boundaries were, and remain, a statistical term: they are defined as the points on the distributions where half of the responses are for one phonetic category and half are for another phonetic category. In the early studies, this statistical term helped investigators to describe the acoustic correlates of phonetic segments. In more recent studies, the term helps us to understand how multiple properties influence phonetic decisions. For example, we can examine how the boundary along an acoustic continuum of one property shifts when another property is manipulated, and so understand better the nature of effect of these two properties. However, we have no evidence that boundaries exist in the natural world, or any account of how or why they might have evolved by natural selection. To extend to them any degree of psychological reality is unsupportable, and deleterious to efforts to understand how phonetic structure is indeed instantiated and retrieved from the speech signal.

ACKNOWLEDGMENTS

This work was supported by Grant No. R01 DC00633 from the National Institute on Deafness and Other Communication Disorders. Carol Manning and Gina Meyer helped with data collection. Michael Studdert-Kennedy, Carol A. Fowler, Doug Keefe, Maureen Higgins, and Arlene Carney provided helpful comments on earlier drafts of this manuscript.

[1]Throughout this manuscript, the fricative shown in parentheses indicates the one for which the *F*2 transition was appropriate.

[2]Although the capability to interrupt the presentation of stimuli during testing existed, it was rarely used. It was there merely as an option, in case it would be needed if the infant were to spit up, have a serious episode of coughing, or such.

[3]Investigators who have difficulty using a headturning task with infants older than 12 months typically do so for one of two reasons, according to Kuhl (1985): either the infant becomes restless sitting on the parent's lap or the infant wants to look for the mechanical animal in the Plexiglass box, producing frequent false-positive headturns. The first of these concerns was eliminated by our use of an infant chair, and the second concern was eliminated by the training procedure, to be discussed.

[4]As it turned out, no child had to attend more than three sessions.

[5]In fact, ratios of change/no-change trials as high as 3/1 have been used successfully (Moore *et al.*, 1975).

[6]It will not be specified here whether children failed to pass the training or the testing phase. In the end, it does not matter because in either case the child was judged not to discriminate the stimuli in the contrast.

[7]The lessons learned from this experiment did help to develop adaptive procedures that were used with 3 year olds (mean age 3 years, 7 months) to explore their perceptual weighting of some acoustic cues in speech perception (Nittrouer, 1996).

Aslin, R. N., and Pisoni, D. B. (**1980**). ''Some developmental processes in speech perception,'' in *Child Phonology*, edited by G. H. Yeni-Komshian, J. F. Kavanagh, and C. A. Ferguson (Academic, New York), Vol. 2, pp. 67–96.

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., and Perey, A. J. (**1981**). ''Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience,'' Child Dev. **52**, 1135–1145.

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P. W., Kennedy, L. J., and Mehler, J. (**1988**). ''An investigation of young infants' perceptual representations of speech sounds,'' J. Exp. Psychol. Gen. **117**, 21–33.

Best, C. T. (**1994**). ''The emergence of native-language phonological influences in infants: A perceptual assimilation model,'' in *The Development of Speech Perception*, edited by J. C. Goodman and H. C. Nusbaum (MIT, Cambridge, MA), pp. 167–224.

Dobie, R. A., and Berlin, C. I. (**1979**). ''Influence of otitis media on hearing and development,'' in *Otitis Media and Child Development: Speech, Language and Education*,'' Ann. Otol. Rhinol. Laryngol. **88** (Suppl. 60), 48–53.

Eilers, R. E., Gavin, W., and Oller, D. K. (**1982**). ''Cross-linguistic perception in infancy: Early effects of linguistic experience,'' J. Child Lang. **9**, 289–302.

Eilers, R. E., Wilson, W. R., and Moore, J. M. (**1977**). ''Developmental changes in speech discrimination in infants,'' J. Speech Hear. Res. **20**, 766–780.

Eimas, P. D. (**1974**). ''Auditory and linguistic processing of cues for place of articulation by infants,'' Percept. Psychophys. **16**, 513–521.

Eimas, P. D. (**1975**). ''Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants,'' Percept. Psychophys. **18**, 341–347.

Eimas, P. D., and Miller, J. L. (**1980**). ''Discrimination of information for manner of articulation,'' Infant Behav. Dev. **3**, 367–375.

Eimas, P. D., Miller, J. L., and Jusczyk, P. W. (**1987**). ''On infant speech perception and the acquisition of language,'' in *Categorical Perception: The Groundwork of Cognition*, edited by S. Harnard (Cambridge U. P., New York), pp. 161–195.

Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., and Vigorito, J. (**1971**). ''Speech perception in early infancy,'' Science **171**, 304–306.

Greenlee, M. (**1980**). ''Learning the phonetic cues to the voiced-voiceless distinction: A comparison of child and adult speech perception,'' J. Child Lang. **7**, 459–468.

Grieser, D., and Kuhl, P. K. (**1989**). ''Categorization of speech by infants: Support for speech-sound prototypes,'' Dev. Psychol. **25**, 577–588.

Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Wright Cassidy, K., Druss, B., and Kennedy, L. (**1987**). ''Clauses are perceptual units for young infants,'' Cognition **26**, 269–286.

Holmberg, T. L., Morgan, K. A., and Kuhl, P. K. (**1977**). ''Speech perception in early infancy: Discrimination of fricative consonants,'' J. Acoust. Soc. Am. **62**, S99.

Jusczyk, P. W. (**1995**). ''Language acquisition: Speech sounds and the beginning of phonology,'' in *Speech, Language, and Communication: Handbook of Perception and Cognition, Vol. 11*, edited by J. L. Miller and P. D. Eimas (Academic, San Diego), pp. 263–301.

Jusczyk, P. W. (**1998**). *The Discovery of Spoken Language* (MIT, Cambridge, MA).

Jusczyk, P. W., Pisoni, D. B., and Mullennix, J. (**1992**). ''Some consequences of stimulus variability on speech processing by two-month-old infants,'' Cognition **43**, 253–291.

Klatt, D. (**1980**). ''Software for a cascade/parallel formant synthesizer,'' J. Acoust. Soc. Am. **67**, 971–995.

Krause, S. E. (**1982**). ''Vowel duration as a perceptual cue to postvocalic consonant voicing in young children and adults,'' J. Acoust. Soc. Am. **71**, 990–995.

Kuhl, P. K. (**1979a**). ''The perception of speech in early infancy,'' in *Speech and Language: Advances in Basic Research and Practice*, edited by N. J. Lass (Academic, New York), pp. 1–47.

Kuhl, P. K. (**1979b**). ''Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories,'' J. Acoust. Soc. Am. **66**, 1668–1679.

Kuhl, P. K. (**1980**). ''Perceptual constancy for speech-sound categories in early infancy,'' in *Child Phonology, Vol. 2: Perception*, edited by G. H. Yeni-Komshian, J. F. Kavanaugh, and C. A. Ferguson (Academic, New York), pp. 41–66.

Kuhl, P. K. (**1981**). ''Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech sound catego-

ries,'' J. Acoust. Soc. Am. **70**, 340–349.
Kuhl, P. K. (**1985**). ''Methods in the study of infant speech perception,'' in *Measurement of Audition and Vision in the First Year of Postnatal Life*, edited by G. Gottlieb and N. A. Krasnegor (Ablex, Norwood, NJ), pp. 223–251.
Kuhl, P. K. (**1987**). ''Perception of speech and sound in early infancy,'' in *Handbook of Infant Perception, Vol. 2*, edited by P. Salapatek and L. Cohen (Academic, New York), pp. 275–382.
Kuhl, P. K. (**1991**). ''Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not,'' Percept. Psychophys. **50**, 93–107.
Kuhl, P. K. (**1993**). ''Developmental speech perception: Implications for models of language impairment,'' Ann. N.Y. Acad. Sci. **682**, 248–263.
Kuhl, P. K., and Miller, J. D. (**1978**). ''Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli,'' J. Acoust. Soc. Am. **63**, 905–917.
Kuhl, P. K., and Miller, J. D. (**1982**). ''Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants,'' Percept. Psychophys. **31**, 279–292.
Lasky, R. E., Syrdal-Lasky, A., and Klein, R. E. (**1975**). ''VOT discrimination by four- to six-and-a-half-month-old infants from Spanish environments,'' J. Exp. Child. Psychol. **20**, 215–225.
Levitt, A. G., Jusczyk, P. W., Murray, J., and Carden, G. (**1988**). ''Context effects in two-month-old infants' perception of labiodental/interdental fricative contrasts,'' J. Exp. Psychol. [Hum Learn] **14**, 361–368.
Lisker, L., and Abramson, A. S. (**1964**). ''A cross-language study of voicing in initial stops: Acoustical measurements,'' Word **20**, 384–422.
Moffitt, A. R. (**1971**). ''Consonant cue perception by twenty- to twenty-four-week-old infants,'' Child Dev. **42**, 717–731.
Moore, J. M., Thompson, G., and Thompson, M. (**1975**). ''Auditory localization in infants as a function of reinforcement conditions,'' J. Speech Hear. Disord. **40**, 29–34.
Morrongiello, B. A., Robson, R. C., Best, C. T., and Clifton, R. K. (**1984**). ''Trading relations in the perception of speech by five-year-old children,'' J. Exp. Child Psychol. **37**, 231–250.
Morse, P. A. (**1972**). ''The discrimination of speech and nonspeech stimuli in early infancy,'' J. Exp. Child Psychol. **14**, 477–492.
Morse, P. A. (**1985**). ''Infant speech perception,'' in *Advances in the Study of Communication and Affect: Auditory Development in Children, Vol. 10*, edited by S. E. Trehub and B. Schneider (Plenum, New York), pp. 215–230.
Nittrouer, S. (**1992**). ''Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries,'' J. Phonetics **20**, 1–32.
Nittrouer, S. (**1996**). ''Discriminability and perceptual weighting of some acoustic cues to speech perception by 3-year-olds,'' J. Speech Hear. Res. **39**, 278–297.
Nittrouer, S., and Miller, M. E. (**1997a**). ''Developmental weighting shifts for noise components of fricative-vowel syllables,'' J. Acoust. Soc. Am. **102**, 572–580.
Nittrouer, S., and Miller, M. E. (**1997b**). ''Predicting developmental shifts in perceptual weighting schemes,'' J. Acoust. Soc. Am. **101**, 2253–2266.
Nittrouer, S., and Studdert-Kennedy, M. (**1987**). ''The role of coarticulatory effects in the perception of fricatives by children and adults,'' J. Speech Hear. Res. **30**, 319–329.
Nittrouer, S., Crowther, C., and Miller, M. E. (**1998**). ''The relative weighting of acoustic properties in the perception of [s]+stop clusters by children and adults,'' Percept. Psychophys. **60**, 51–64.
Nittrouer, S., Miller, M. E., Crowther, C. S., and Manhart, M. J. (**2000**). ''The effect of segmental order on fricative labeling by children and adults,'' Percept. Psychophys. **62**, 266–284.
Nozza, R. J. (**1987**). ''Infant speech-sound discrimination testing: Effects of stimulus intensity and procedural model on measures of performance,'' J. Acoust. Soc. Am. **81**, 1928–1937.
Parnell, M. M., and Amerman, J. D. (**1978**). ''Maturational influences on perception of coarticulatory effects,'' J. Speech Hear. Res. **21**, 682–701.
Sinex, D. G., and McDonald, L. P. (**1989**). ''Average discharge rate representation of voice onset time in the chinchilla auditory nerve,'' J. Acoust. Soc. Am. **83**, 1817–1827.
Sinex, D. G., McDonald, L. P., and Mott, J. B. (**1991**). ''Neural correlates of nonmonotonic temporal acuity for voice onset time,'' J. Acoust. Soc. Am. **90**, 2441–2449.
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, Orlando), pp. 243–255.
Streeter, L. (**1976**). ''Language perception shows effects of both innate mechanisms and experience,'' Nature (London) **259**, 39–41.
Wardrip-Fruin, C., and Peach, S. (**1984**). ''Developmental aspects of the perception of acoustic cues in determining the voicing feature of final stop consonants,'' Lang. Speech. **27**, 367–379.
Werker, J. F. (**1989**). ''Becoming a native listener,'' Am. Sci. **77**, 54–59.
Werker, J. (**1991**). ''The ontogeny of speech perception,'' in *Modularity and the Motor Theory of Speech Perception*, edited by I. G. Mattingly and M. Studdert-Kennedy (Lawrence Erlbaum, Hillsdale, NJ), pp. 91–109.
Werker, J. F. (**1994**). ''Cross-language speech perception: Developmental change does not involve loss,'' in *The Development of Speech Perception*, edited by J. C. Goodman and H. C. Nusbaum (MIT, Cambridge, MA), pp. 93–120.
Werker, J. F., and Tees, R. C. (**1984**). ''Cross-language speech perception: Evidence for perceptual reorganization during the first year of life,'' Infant Behav. Dev. **7**, 49–63.
Werker, J. F., Gilbert, J. H. V., Humphrey, K., and Tees, R. C. (**1981**). ''Developmental aspects of cross-language speech perception,'' Child Dev. **52**, 349–355.

A robust method to study stress "deafness"[a)]

Emmanuel Dupoux[b)]
Laboratoire de Sciences Cognitives et Psycholinguistique (EHESS/CNRS), 54 Boulevard Raspail, 75006 Paris, France

Sharon Peperkamp[c)]
Laboratoire de Sciences Cognitives et Psycholinguistique (EHESS/CNRS), 54 Boulevard Raspail, 75006 Paris, France and Département de Sciences du Langage, Université de Paris VIII, 2 Rue de la Liberté, 93535 Saint Denis, France

Núria Sebastián-Gallés[d)]
Departamente de Psicologia Bàsica, Universitat de Barcelona, P. de la Vall d'Hebron, 171, 08035 Barcelona, Spain

(Received 11 September 2000; revised 16 April 2001; accepted 30 April 2001)

Previous research by Dupoux *et al.* [J. Memory Lang. **36**, 406–421 (1997)] has shown that French participants, as opposed to Spanish participants, have difficulties in distinguishing nonwords that differ only in the location of stress. Contrary to Spanish, French does not have contrastive stress, and French participants are ''deaf'' to stress contrasts. The experimental paradigm used by Dupoux *et al.* (speeded ABX) yielded significant group differences, but did not allow for a sorting of individuals according to their stress ''deafness.'' Individual assessment is crucial to study special populations, such as bilinguals or trained monolinguals. In this paper, a more robust paradigm based on a short-term memory sequence repetition task is proposed. In five French–Spanish cross-linguistic experiments, stress ''deafness'' is shown to crucially depend upon a combination of memory load and phonetic variability in $F0$. In experiments 3 and 4, nonoverlapping distribution of individual results for French and Spanish participants is observed. The paradigm is thus appropriate for assessing stress deafness in individual participants. [DOI: 10.1121/1.1380437]

PACS numbers: 43.71.Hw, 43.71.Es, 43.71.An [KRK]

I. INTRODUCTION

The way in which we perceive speech sounds depends on the properties of our native language. This phenomenon has been noticed by linguists (Sapir, 1921; Polivanov, 1931), and has been investigated by psycholinguists mostly for segmental categories. For instance, Japanese participants map the English /r/ and /l/ onto a single /ɾ/ category and have trouble discriminating between these English segments (Goto, 1971; Miyawaki *et al.*, 1981). The impact of the mother tongue on the perception of speech segments has attracted considerable attention among psycholinguists, and several models have been proposed to account for it. Some of these models propose that infants learn to focus their attention onto the acoustic cues that are most relevant for their language (Jusczyk, 1993; Nusbaum and Goodman, 1994). Other models postulate the existence of an abstract, language-specific, phoneme detector that segments the continuously varying acoustic signal into discrete categories. According to these models, non-native segments that are close enough to a segment in the native language are assimilated to it; consequently, two non-native segments that are assimilated to the same native category will be very difficult to distinguish (Best, 1994; Flege, 1995; Kuhl, 2000). Several researchers have focused on the age at which this phonological processing level is fixed (Werker and Tees, 1984; Best, McRoberts, and Sithole, 1988; Kuhl *et al.*, 1992; Jusczyk *et al.*, 1993), the extent of individual variability in late learners (Flege, MacKay, and Meador, 1999), and the possible effect of extensive training (Lively, Logan, and Pisoni, 1993; Francis, Baldwin, and Nusbaum, 2000). Hence, this line of research relates to theoretical questions regarding brain plasticity and the existence of a critical period. Moreover, it has practical implications concerning the development of training procedures for second language learning.

Languages differ not only in their repertoire of phonemes, but also in their suprasegmental properties: some use tones (Mandarin), pitch accent (Japanese), length (Finnish), or stress (Spanish) to make lexical distinctions; others do not use any of these suprasegmental properties to distinguish lexical items (French). The impact of this type of variation has been studied less extensively, and its incorporation into models of speech processing and acquisition is still awaited. As to the perception of stress, Dupoux *et al.* (1997) found that native speakers of French, a language with fixed word-final stress, have difficulties with the discrimination of nonwords that differ only in the position of stress (e.g., [vásuma] vs [vasúma] vs [vasumá]). Spanish listeners, by contrast, do not have any difficulties, stress being contrastive in their language. More research has focused on the perception of tone

[a)]Portions of this work were presented in ''Perception of stress by French, Spanish, and bilingual subjects,'' Proceedings of EuroSpeech '99, Budapest, September 1999, Vol. 6, pp. 2683–2686.
[b)]Electronic mail: dupoux@lscp.ehess.fr.
[c)]Electronic mail: sharon@lscp.ehess.fr
[d)]Electronic mail: nsebastian@psi.ub.es

by speakers of nontonal languages. For instance, it has been shown that English listeners have difficulties with the perception of Mandarin Chinese tones (Kiriloff, 1969; Bluhme and Burr, 1971; Wang *et al.*, 1999). Wang *et al.* (1999) showed that American students with one or two semesters in Mandarin correctly identified the four tones in only 69% of the cases. After extensive training, this performance jumped to 90% correct, but still only one out of the eight participants attained native-like performance. Gandour (1983) compared the perception of tone by speakers of English and four tone languages. Compared to the speakers of the latter, the English participants paid more attention to $F0$ and less to contour information in order to identify tones. Finally, Lee and Nusbaum (1993) found that Mandarin but not English listeners are slowed down by an irrelevant change in pitch level when they make a segmental classification. In other words, native speakers of Mandarin perceive pitch and segmental information in an integral fashion, whereas the two dimensions are perceived as orthogonal by native speakers of English.

In this paper, we focus on the perception of stress. Unlike some of the tonal and segmental contrasts, stress contrasts have massive acoustic correlates (duration, $F0$, and energy); it is, therefore, surprising that French participants have any problem at all with the perception of stress. Indeed, informal testing suggests that the acoustic correlates of stress are salient enough for listeners to have little difficulty in an identification paradigm similar to that used in Wang *et al.* (1999) for the perception of tones. Dupoux *et al.* (1997) reported that when tested with a standard AX discrimination task, French participants had *no* detectable problem with the stress contrast. It is only when a more demanding task was used (such as ABX with talker changes) that French participants began to have problems with stress. This suggests two things. First, unlike what happens with the perception of consonants for which within-category discrimination is very difficult, French listeners can use the acoustic stress cues in order to perform standard discrimination tasks flawlessly. Second, French listeners are nevertheless ''deaf'' to stress contrasts at a more abstract processing level, which is revealed only with tasks that are more demanding as far as memory and perceptual resources are concerned. Our aim in this paper, then, is to explore more systematically the effect of these variables in order to build a more robust paradigm to study stress ''deafness.'' Importantly, we require our paradigm to give individual results, such that the study of stress perception in special populations (for instance bilinguals, second language learners, or trained monolinguals) becomes possible.

In the ABX paradigm used in Dupoux *et al.* (1997), participants heard three successive items in three different voices, and had to judge whether the third item was identical to the first or to the second one. This task required a short-term working memory buffer because the decision had to be delayed until the final stimulus was heard. Furthermore, the stimuli A, B, and X were pronounced by three different talkers. This phonetic variability made an acoustically based response strategy more difficult to use than in a standard single-token AX paradigm. In the present study, we set up a short-term memory task and we manipulate both memory load and phonetic variability in order to find the most effective combination for a robust stress deafness effect to arise.

It should be noted that the ABX task used in Dupoux *et al.* (1997) was probably not optimal. First, the observed deafness was far from being total. For instance, in experiment 1, French participants made 20% errors in the stress contrast, whereas Spanish participants made only 4% errors. This difference was highly significant ($p<0.001$ both by items and participants), but still, the French participants performed much better than chance (50%). This might mean either that participants relied on some residual phonological representation of stress, or that the ABX paradigm was not demanding enough and allowed for alternative strategies involving an acoustic level of representation. In this paper, we evaluate whether French participants perform significantly better than chance across the several versions of our new paradigm.

Second, the results showed considerable individual variability, and an inspection of the distribution of individual errors in the stress discrimination task for French and Spanish participants revealed a substantial overlap in the distribution of errors across the two populations. That is, some French participants were as good as typical Spanish participants (three French participants made less than 5% errors), while, conversely, some Spanish participants were as bad as many of the French participants (one Spanish participant made 15% errors). This overlap might be due to the fact that some French participants succeeded in representing stress phonologically, while some Spanish failed; alternatively, it might be due to noise in the experimental method. In the present experiments, we compute the overlap in individual results across populations as well as the reliability of the effects, in order to tease apart the variation due to the participants from the variation due to the method.

To sum up, we propose to study the perception of stress contrasts in French and Spanish participants using a short-term memory task. The same task is used in all experiments. In the experiments, we compare the recall performance of a stress contrast with that of a control phonemic contrast, across different levels of memory load. Experiments differ in the amount of phonetic variability that is introduced in the stimuli tokens. In experiment 1, we use different tokens spoken by the same talker for each item. Experiments 2 and 3 add variability on the mean pitch by using speech resynthesis. Experiment 4 uses tokens from two talkers, and experiment 5 uses, as a control, a single token for each item. In all these experiments, we perform a group analysis in which we compare performance in the stress contrast to performance in the control phonemic contrast as well as to chance performance. We also analyze individual data by evaluating the overlap between the French and Spanish populations. Finally, we assess the reliability of the paradigm by computing a split-half reliability index for each experiment.

II. EXPERIMENT 1

The experiment was divided into two parts. In each part, participants were required to learn two CVCV nonwords that are a minimal pair differing only in one phonological dimen-

sion, i.e., place of articulation of the second consonant or location of stress. In each part, participants were taught to associate the two nonwords to the keys [1] and [2], respectively, of a computer keyboard. After some training with an identification task, participants listened to longer and longer random sequences of the two items, which they were required to recall and transcribe as sequences of [1] and [2]. The phonemic contrast in the first part was meant to be equally easy for speakers of French or Spanish, and was used to establish baseline performance. In order to diminish the likelihood that participants use recoding strategies, the stimuli were short and the tokens in the sequences were separated from one another by a very short interval. Moreover, in order to prevent participants from using echoic memory, every sequence was followed by the word ''OK'' (Morton, Crowder, and Prussin, 1971; Morton, Marcus, and Ottley, 1981).

In this experiment, some phonetic variability was present in that each word was instantiated by one of six acoustically different tokens. Moreover, memory load was manipulated by an increase of the sequence length from two to four and eventually six. We predicted that French participants should make many more errors in the stress than in the phoneme condition, whereas Spanish participants should have a similar performance in both conditions.

A. Method

1. Materials

Two minimal pairs were constructed, one involving a segmental contrast, i.e., [túku, túpu], and the other one involving a stress contrast, i.e., [píki, pikí]. All items are nonwords in both French and Spanish. They were recorded ten times each by a female trained phonetician who is a native speaker of Dutch. Six recordings of each word were selected. Their mean duration was 491 ms. In addition, the word ''OK'' was recorded once by a male talker. All recorded items were digitized at 16 kHz at 16 bits, digitally edited, and stored on a computer disk.

The mean durations of the tokens with the phonemic and the stress contrast were 345 and 351 ms, respectively. As to the tokens with the stress contrast, stressed vowels were on average 20 ms longer than unstressed vowels, a significant difference [$F(1,10)=14.5$, $p<0.003$]. Stressed vowels also had a higher pitch than unstressed vowels; in particular, the maximum $F0$ value of the stressed vowels was on average 45.3 Hz higher than that of the unstressed vowels, corresponding to a significant difference of 3.9 semitones [$F(1,10)=441$, $p<0.001$]. Finally, stressed vowels were on average 1.6 dB louder than unstressed vowels, again a significant difference [$F(1,10)=20.6$, $p<0.001$].

For each minimal pair three experimental blocks were constructed, each containing eight sequences of the two nonwords. The first block contained two-word sequences, the second block contained four-word sequences, and the third block contained six-word sequences. There are four logically possible two-word sequences, which each appeared twice in the first block. In the other two blocks, the eight sequences were all different. Out of the 16 possible sequences of four repetitions, eight among the most varied ones were selected, making up the second block. For instance, 1211, containing one transition from 1 to 2 and another one from 2 to 1, has more variation than 2111, containing only a transition from 2 to 1. There are six possible sequences with two transitions, which were all selected for the second block. Two sequences with one transition were selected to complete this block. Similarly, out of the 64 possible sequences of six repetitions, eight were selected for the third block; four of these contained three transitions and four contained four transitions. (The maximum number of transitions, five, gives rise to the completely regular, hence easy, patterns 121212 and 212121.) All selected sequences are listed in the Appendix. The overall design was $2\times2\times3$: language$\times$contrast $\times$sequence length.

2. Procedure

Participants were first tested on the minimal pair containing the phonemic contrast. Participants were told that they were going to learn two words in a foreign language. They could listen to the various tokens of these two words by pressing the number keys [1] and [2] as many times as they wanted. The nonword [túku] was associated to key [1], while its counterpart [túpu] was associated to key [2]. Pressing each one of these keys resulted in the playing of one token of the corresponding word. Subsequently, it was verified that participants had learned the distinction between the two words as well as the correct association between the words and the number keys. That is, they heard a token of one of the items and had to press the associated key, [1] or [2]. A message on the screen informed participants whether their responses were correct. The message was ''OK!'' or ''ERROR!,'' and was displayed for 800 ms. We defined a success criterion of five correct responses in a row. After having reached this criterion, participants turned to the main experiment.

During the test, participants listened to 24 sequences constituted by repetitions of the two words, divided into three blocks as described above. Their task was to reproduce each sequence by typing the associated keys in the correct order. For each participant, the order of the eight sequences in each block was randomized, and each item was instantiated randomly by one of the six recorded tokens. In order to diminish the likelihood that participants mentally translate the words into the associated numbers while listening to the sequence, the silent period between the items in a sequence was kept very short, i.e., 80 ms. Each trial consisted of a sequence followed by the word OK, and participants could not begin typing their response until they had heard this word. Participants did not receive feedback as to whether their responses were right or wrong. A 1500-ms pause separated each response from the next trial.

The whole procedure was repeated with the minimal pair containing a stress contrast. The nonword with stress on the first syllable, [píki], was associated to key [1], while its counterpart with stress on the second syllable, [pikí], was associated to key [2].

On average, the entire experiment lasted about 15 min. Responses were recorded on a computer disk and classified as follows. Responses that were a 100%-correct transcription

TABLE I. Percent error with phoneme and stress contrast as a function of sequence length for 12 French and 12 Spanish participants in experiment 1.

Sequence length	2	4	6	Mean
		French		
Phoneme	6.3%	11.3%	43.8%	20.5%
Stress	18.8%	38.6%	72.3%	43.2%
		Spanish		
Phoneme	8.3%	14.6%	61.5%	28.1%
Stress	1.0%	16.7%	58.3%	25.3%

of the input sequence were coded as correct; all other responses were coded as incorrect. Among the incorrect responses, those that were a 100%-incorrect transcription—i.e., with each token of the sequence labeled incorrectly—were coded as *reversals*. Participants with more reversals than correct responses in either the phoneme or the stress condition were rejected, the high percentage of reversals suggesting that they might have confused the number key associated with the first item with the one associated with the second item.

3. Participants

Twelve native French speakers, aged between 20 and 38, and 12 native Spanish speakers, aged between 18 and 21, were tested individually. None of the participants had a known hearing deficit. Two additional French participants and one additional Spanish participant were tested and excluded from the results on the basis of the rejection criterion defined above. For both French participants, the reversals outnumbered the correct responses in the case of the stress contrast, while the Spanish participant had too many reversals with the phoneme contrast.

B. Results

Error rates for French and Spanish participants for the phonemic and the stress contrast as a function of sequence length are shown in Table 1.

These data were subjected to an analysis of variance (ANOVA) with the between-participant factor language (French vs Spanish), and within-participant factors contrast (phoneme vs stress) and sequence length (2 vs 4 vs 6). As predicted on the basis of the results in Dupoux *et al.* (1997), there was a significant interaction between language and contrast [$F(1,22)=11.3$, $p<0.003$]. This interaction was due to the fact that there was an effect of contrast for French participants, with stress yielding more errors than phoneme [$F(1,11)=13.7$, $p<0.003$], but not for Spanish participants [$F(1,11)<1$, $p>0.1$]. *Post hoc* comparisons indicated a significant effect of contrast in French participants for sequence lengths 4 and 6 (Bonferroni-corrected $p=0.024$ and $p=0.03$, respectively), but not for sequence length 2 (Bonferroni-corrected $p=0.27$). However, there was no significant interaction between length and contrast [$F(2,22)=1$, $p>0.1$]. The Spanish participants showed no significant difference between phoneme and stress at any length ($p>0.1$).

For the French participants, we compared the percentage of correct responses with the stress contrast to chance performance at each sequence length. The chance level is defined as the probability of making a *correct* response by responding randomly. According to the binomial law, it is equal to $1/2^n$ with n being the sequence length. The chance level is thus $1/4=25\%$ at length 2, $1/16=6\%$ at length 4, and $1/64=2\%$ at length 6. For the comparisons, we ran one-tailed t-tests. The comparisons were significant at each length (Bonferroni-corrected $p<0.001$). The overall performance across lengths was significantly better than chance [$F(1,11)=62.30$, $p<0.0001$].

For each participant, the difference score is defined as the percentage of errors with the stress contrast minus the percentage of errors with the phoneme contrast. The *overlap* between populations is defined as the percentage of participants whose difference scores lie in the area that is common to both distributions. In this experiment, the overlap was 83%. Next, we compute the *optimal classification score* as follows. First, we establish an arbitrary separation criterion in the range of the observed difference scores. Second, we classify the individual participants based on their difference scores: all participants with a difference score higher than the separation criterion are classified as French, whereas all participants below the criterion are classified as Spanish. Third, we compute the percentage of participants that are correctly classified according to such a criterion. Fourth, the optimal classification score is now defined as the separation criterion that yields the best classification score. In this experiment, the optimal classification score was 79.2%. That is, if we tried to guess the mother tongue of individual participants based on their results in the experiment, we would correctly classify 79.2% of the participants as either French or Spanish.

Finally, to run a reliability test, we split the data of each participant in two halves in the following way: for each contrast and each sequence length, the first four responses of the participant were put in bin 1 and the final four responses were put in bin 2. We derived a difference score (stress minus phoneme) for these two bins, and ran a correlation analysis across participants. The correlation coefficient r between the two halves was 0.696 ($p<0.001$).

C. Discussion

As predicted, French participants had significant difficulties with the stress contrast compared to the phoneme contrast, whereas Spanish participants had equal performance with the two contrasts. The difficulty with stress in the French was significant only with length 4 and 6, although there was a numerical trend in the same direction at length 2. Note also that the interaction between sequence length and contrast was not significant, which suggests that the three sequence lengths are not qualitatively different. We also found, as in Dupoux *et al.* (1997), that the mean performance with the stress contrast was significantly better than chance; this was true at all sequence lengths.

In order to compare the robustness of the present paradigm with the ones used previously, we reanalyzed experi-

TABLE II. Separation of responses to stress discrimination in French versus Spanish participants in experiment 1 and 3 in Dupoux *et al.* (1997) and in experiment 1 of the present series.

	Stress errors Dupoux *et al.* (1997) Experiment 1	Stress-phoneme score Dupoux *et al.* (1997) Experiment 3	Stress-phoneme score Experiment 1
Number of trials	96	96	24
Number of participants	15	20	24
ANOVA	$F(1,30)=17.1$	$F(1,38)=5.7$	$F(1,22)=11.3$
Overlap of the distributions	46.8%	77.5%	83%
Optimal classification score	81.3%	65%	79.2
Split-half correlation coefficient	$r=0.895$	$r=0.435$	$r=0.696$

ments 1 and 3 of Dupoux *et al.* (1997) that used the ABX discrimination paradigm (see Table II).

As seen in Table II, the robustness of the results in the present experiment was actually weaker than that obtained in experiment 1 in Dupoux *et al.* (1997). Indeed, the overlap of the two distributions of scores for the Spanish vs the French participants was larger in the present experiment than in experiment 1 of Dupoux *et al.* Moreover, the reliability of the present experiment was also smaller. Nevertheless, the results of the present experiment were encouraging for the following reasons:

First, our experiment used only one talker with limited phonetic variability; this may have allowed the French participants to rely on acoustic information to discriminate and classify the two stress patterns. In contrast, experiment 1 and 3 in Dupoux *et al.* used three talkers. In our next experiment, we introduce more phonetic variability by way of pitch manipulation. Second, the present experiment has only 40 data points per participant, whereas experiments 1 and 3 of Dupoux *et al.* (1997) used 96 data points. This might explain the weaker reliability of our experiment. In experiments 3 and 4 of the present series, we increase the number of sequences in order to have a more comparable data set. Finally, the score that we derived in our experiment is relative to a baseline. By contrast, in experiment 1 in Dupoux *et al.*, the scores were absolute error rates for a stress contrast. Absolute scores have intrinsically lower errors of measurement than difference scores, since in the latter the error of measurement appears twice. However, the absence of a baseline induces a potential confound with population sampling biases. Indeed, irrelevant variables such as age, IQ, or motivation can obscure or erroneously increase differences in the mean performance in a task across two populations of participants. Therefore, a within-participant design with a control baseline is preferable. In fact, a comparison between experiment 3 of Dupoux *et al.* (1997), which also uses a baseline condition, and the present experiment reveals much more similar robustness and reliability of the two paradigms. Therefore, we had good reasons to hope that the present paradigm could be modified such as to become more robust and reliable than the ABX discrimination task.

III. EXPERIMENT 2

This experiment was a replication of experiment 1 with only one change: we introduced a variation in the pitch of the tokens through speech resynthesis. These changes were only plus or minus 5% of the original pitch. We expected that this modification would make it difficult for the French participants to rely on an acoustic representation. By contrast, pitch variations alone should not affect the Spanish participants too much, since Spanish speakers represent stress abstractly in the phonological representation. We thus predicted a larger difference between the two populations than in experiment 1.

A. Method

1. Materials

The same nonwords as those in experiment 1 were used. However, pitch variation was obtained in the various tokens by means of a resynthesis algorithm in the waveform editor COOL96,[1] with the percentages 105, 103, 101, 99, 97, and 95, respectively.

2. Procedure

The procedure was the same as in experiment 1, with one modification: within each sequence, a single token could not appear more than once.

3. Participants

Twelve native French speakers, aged between 20 and 40, and 12 native Spanish speakers, aged between 18 and 21, were tested individually. None of the participants had participated in experiment 1, and none had a known hearing deficit. Three additional French speakers were tested and excluded from the results, due to too many reversals among their responses with the stress contrast.

B. Results

Error rates for French and Spanish participants for the phonemic and the stress contrast as a function of sequence length are shown in Table III.

These data were subjected to an ANOVA with the between-participant factor language (French vs Spanish), and within-participant factors contrast (phoneme vs stress) and sequence length (2 vs 4 vs 6). As in the previous experiment, there was a significant interaction between language and contrast [$F(1,22)=17.3$, $p<0.001$]. The interaction was due to the fact that there was an effect of contrast for the French participants [$F(1,11)=77.8$, $p<0.001$], but not for the Spanish participants [$F(1,11)<1$, $p>0.1$]. *Post hoc* comparisons indicated a significant effect of contrast in the French partici-

TABLE III. Percent error with phoneme and stress contrast as a function of sequence length for 12 French and 12 Spanish participants in experiment 2.

Sequence length	2	4	6	Mean
		French		
Phoneme	11.3%	32.5%	71.3%	38.4%
Stress	58.8%	73.8%	97.5%	76.7%
		Spanish		
Phoneme	8.7%	25.0%	63.5%	32.4%
Stress	13.5%	26.9%	74.0%	38.1%

pant at all sequence lengths (Bonferroni-corrected p values: 0.009, 0.001, and 0.009 for sequence lengths 2, 4, and 6, respectively). There was no significant interaction between sequence length and contrast for the French participants [$F(2,22)<1$, $p>0.1$]. The Spanish participants showed no significant difference at any length ($p>0.1$). For the French participants, we ran a t-test at each sequence length to compare the performance in the stress condition to chance performance. The comparison was significant for lengths 2 and 4, but not for length 6 (Bonferroni-corrected one-tailed $p<0.016$, $p<0.0015$, and $p>0.2$, respectively). The overall performance across lengths was significantly better than chance [$F(1,11)=16.51$, $p<0.002$].

The overlap between the two populations was 62%, with an optimal classification score of 71.7%. As for the reliability test, the correlation coefficient r between the two halves of the experiment was 0.566 ($p<0.004$).

C. Discussion

In this experiment, we introduced a change in the pitch of the experimental tokens varying between -5% and $+5\%$. The group results are very similar to those in experiment 1, except that the stress deafness effect for the French participants, i.e., the difference in performance between the stress and the phoneme condition, is numerically larger and now significant even at sequence length 2. Yet, French participants are still better than chance with the stress contrast at sequence lengths 2 and 4. Finally, compared to experiment 1, the distribution of scores between the French and the Spanish participants is more separated. This is shown by the fact that the classification error is divided by 2 and the overlap between the two distributions is reduced. Hence, the introduction of a pitch change was successful in making more difficult an acoustically based strategy for the French participants, thus increasing the size of the language-specific effect.

Note, however, that there was one Spanish participant who made a great number of errors in the stress contrast. This outlier was responsible for the high degree of overlap in the two distributions of scores that remains in this experiment. After an interview with this participant, it turned out that one of the [pikí] tokens was perceived as a little ambiguous in terms of stress. Therefore, we decided to run a replication of this experiment using a novel set of stimuli. Another point worth noting is that the reliability of this experiment was not good ($r=0.566$). In the next experiment, we also added more sequences, in order to get more stable individual data.

IV. EXPERIMENT 3

In this experiment, we introduced a novel set of stimuli that was better controlled for stress. In experiment 2, the minimal pairs consisted of nonwords having identical vowels in the two syllables ([túku, túpu] and [píki, pikí]). This might have made the stimuli confusable in short-term memory. In order to test whether the obtained effects generalize to a situation with a different vowel in the two syllables, we constructed two new minimal pairs, consisting of nonwords having different vowels in the two syllables. We also increased the power of this experiment by adding sequences of length 3 and 5. As in experiment 2, we introduced a $\pm 5\%$ change in pitch in the different tokens.

A. Method

1. Materials

Two minimal pairs were constructed, one involving a segmental contrast, i.e., [kúpi, kúti], the other one involving a stress contrast, i.e., [mípa, mipá]. All items are nonwords in both French and Spanish. They were recorded about ten times by a female talker, the same one who recorded the items used in experiments 1 and 2. The stimuli were judged by a Spanish phonetician and only tokens with unambiguous stress patterns were used. Six recordings of each item were selected. All recorded items were digitized at 16 kHz at 16 bits, digitally edited, and stored on a computer disk.

The mean durations of the tokens with the phonemic and the stress contrast were 439 and 513 ms, respectively. As in experiment 2, more variation was obtained in the six tokens of the four items, in that they had their pitch changed by means of the COOL96 waveform editor, with the percentages 105, 103, 101, 99, 97, and 95, respectively. As to the tokens with the stress contrast, stressed vowels were on average 21 ms longer than unstressed vowels, a significant difference [$F(1,5)=43.20$, $p<0.001$]. Stressed vowels also had a higher pitch than unstressed vowels; in particular, the maximum $F0$ value of the stressed vowels was on average 52.6 Hz higher than that of the unstressed vowels, corresponding to a significant difference of 4.9 semitones [$F(1,10)=521$, $p<0.001$]. Finally, stressed vowels were on average 3.7 dB louder than unstressed vowels, again a significant difference [$F(1,5)=78.20$, $p<0.001$].

Two blocks, containing eight three-word sequences and eight five-word sequences, respectively, were added. Thus, there were five test blocks for each contrast, containing sequences of length 2, 3, 4, 5, and 6. There are exactly eight logically possible three-word sequences, all of which appeared once in the second block. Out of the 32 possible sequences of five repetitions, eight among the most varied ones were selected for the fourth block; half of them contained two transitions; the other half contained three transitions (see the Appendix). As to the remaining blocks with sequences of length 2, 4, and 6, the same sequences as in experiments 1 and 2 were selected. The overall design was $2\times2\times5$: language$\times$contrast$\times$sequence length.

2. Procedure

The procedure was as in experiment 2, with the following modifications. First, the stimuli were presented in a different manner. That is, participants were first asked to press the number key [1], upon which they heard all tokens of the first item. They were then asked to press the number key [2], upon which they heard all tokens of the second item. Subsequently, participants could continue listening to the various tokens of the two items by pressing the associated keys; as in the previous experiments, pressing each one of these keys resulted in the playing of one token of the corresponding item. They could thus hear as many tokens of the two items as they desired. Second, in the training phase, we increased the number of correct responses in the success criterion from five to seven. Third, in order to diminish the amount of noise in the results, participants were warned whenever they entered a sequence with a length that did not correspond to the length of the input string and asked to enter their reply again.

For the phoneme contrast, [kúpi] was associated with key [1] and [kúti] with key [2]. For the stress contrast, [mípa] was associated with key [1] and [mipá] with key [2].

On average, the experiment lasted between 15 and 20 min. Responses were recorded on a computer disk.

3. Participants

Twelve native French speakers, aged between 20 and 42, and 12 native Spanish speakers, aged between 23 and 29, were tested individually. None of the participants had participated in experiments 1 or 2, and none had a known hearing deficit. Two additional Spanish speakers were tested and excluded from the results, due to too many reversals among their responses. For one of these participants this was the case in the phoneme condition, and for the other one it was in the stress condition.

B. Results

Error rates for French and Spanish participants for the phonemic and the stress contrast as a function of sequence length are shown in Table IV.

These data were subjected to an ANOVA with the between-participant factor language (French vs Spanish), and within-participant factors contrast (phoneme vs stress) and sequence length (2 vs 3 vs 4 vs 5 vs 6). As in experiments 1 and 2, there was a significant interaction between language and contrast [$F(1,22)=70.3$, $p<0.0001$]. This interaction was due to the fact that stress yielded significantly more errors than phoneme for French participants [$F(1,11)=71.0$, $p<0.0001$], whereas there was a nonsignificant trend in the other direction for Spanish participants [$F(1,11)=3.7$, $0.1>p>0.05$]. *Post hoc* comparisons indicated a significant effect of contrast in French participants for all sequence lengths (Bonferroni-corrected p values: 0.035, 0.015, 0.001, 0.001, and 0.001 for sequence lengths 2, 3, 4, 5, and 6, respectively). The interaction between sequence length and contrast was significant for French participants [$F(4,44)=7.09$; $p<0.001$]. Spanish participants showed no significant difference at any length ($p>0.1$).

TABLE IV. Percent error with phoneme and stress contrast as a function of sequence length for 12 French and 12 Spanish participants in experiment 3.

Sequence length	2	3	4	5	6	Mean
			French			
Phoneme	2.8%	7.3%	15.6%	18.7%	33.3%	15.4%
Stress	29.2%	28.1%	59.4%	64.6%	86.5%	53.5%
			Spanish			
Phoneme	7.3%	5.2%	12.5%	35.4%	61.5%	24.4%
Stress	0.0%	4.2%	10.4%	32.3%	53.1%	20.0%

For French participants, we ran a t-test at each sequence length to compare the performance with the stress contrast to chance performance. This comparison was significant at each length (Bonferroni-corrected one-tailed $p<0.001$), except at length 6 where the comparison was only marginally significant ($p=0.066$). The overall performance across lengths was significantly better than chance [$F(1,11)=37.04$, $p<0.001$].

The overlap between the two populations was 0%, with an optimal classification score of 100%. The two distributions of points are actually separated by a gap whose size is 9.4% of the total range. As for the reliability test, the correlation coefficient r between the two halves of the experiment was 0.882 ($p<0.001$).

In this experiment, the number of training trials was saved. Note that due to a programming error, the training criterion was seven correct answers in a row for the French but only six for the Spanish participants. For the French participants, the number of training trials was 7.3 for the stress and 7.1 for the phoneme condition, a nonsignificant difference [$F(1,11)<1$, $p>0.1$]. For the Spanish participants, the number of training trials was 6.1 for the stress and 6.7 for the phoneme condition, again, a nonsignificant difference [$F(1,11)=1.31$, $p>0.1$]. In other words, most French and Spanish participants passed the training without any error. In a global ANOVA, there was no significant interaction between contrast and language [$F(1,22)=2.02$, $p>0.1$].

C. Discussion

This experiment replicated experiment 2 using novel and more controlled stimuli, as well as more sequence lengths. The group analysis was very similar to the one in experiment 2: we found a stress deafness effect in the French participants and not in the Spanish participants, and the effect was significant at all sequence lengths, while performance for the stress contrast was better than chance at all sequence lengths (except at length 6, where it was only marginal). In the individual analyses, however, the results were stronger than in experiment 2: there was no overlap in the distributions of the French and Spanish populations and the split-half reliability of the test was quite good.

The fact that the French participants were still able to perform the memory task with the stress contrast better than chance can be interpreted in two ways: either they managed to apply an acoustic strategy with the stress contrast, or they have a residual phonological representation of stress, which would mean that the deafness is not total. In order to distinguish between these two hypotheses, we introduced more

phonetic variability by using two talkers in our next experiment. If the residual capacity to distinguish stress is based on an acoustic strategy, it should be harder to apply such a strategy on tokens with increased phonetic variability, and hence the size of the deafness effect should increase as well.

V. EXPERIMENT 4

This experiment was a replication of experiment 3 with even more phonetic variability, in that we introduced a second, male, talker. The big difference in $F0$ is predicted to make it more difficult for the participants to use acoustic information. For the Spanish participants, this should not be a problem in either the phoneme or the stress condition, since they can use their phonological representations of the target items to perform the task. The same holds for the French participants in the phoneme condition. By contrast, the French participants should have increased problems in the stress condition if the residual performance that we observed in the previous experiments were due at least in part to acoustic strategies. In brief, we predict that if stress discrimination in the French participants is acoustically based, the introduction of a new talker should increase the size of the language-specific effect measured previously, and the difference scores of the two populations should be even further apart than in experiment 3.

A. Method

1. Materials

The same minimal pairs as in experiment 3 were used, but a new recording with a male voice was made. This second talker was a native speaker of French, who imitated the tokens produced by the female talker. A trained phonetician corrected his productions till they were deemed satisfactory. For each of the four test items ([kúpi], [kúti], [mípa], and [mipá]), three tokens from the male talker were selected. These tokens replaced three of the six tokens per item of the female talker used in experiment 3. A total of six tokens per item was thus obtained: three produced by the female talker and three produced by the male talker. Given that the tokens produced by the male talker were shorter than those produced by the female talker, they were stretched such that they matched exactly the length of the tokens from the female talker they replaced.[2]

As to the items for the stress contrast, stressed vowels had a higher pitch than unstressed vowels; in particular, the maximum $F0$ value of the stressed vowels was on average 48.4 Hz higher than that of the unstressed vowels, corresponding to a significant difference of 5.2 semitones [$F(1,10)=304$, $p<0.001$]. Moreover, stressed vowels were on average 6.0 dB louder than unstressed vowels, again a significant difference [$F(1,10)=45.7$, $p<0.001$].

The mean $F0$ of the tokens of the female talker was 181 Hz, and the mean $F0$ of the tokens of the male talker was 144 Hz; hence, there was an $F0$ variation between the two talkers of 20%. As in experiment 2 and 3, $\pm 5\%$ variation in pitch was obtained, in that the three tokens of both sets had their pitch changed by means of the PSOLA algorithm, with the percentages 105, 101, and 95, respectively. The design was the same as in experiment 3.

2. Procedure

The same procedure as in experiment 3 was used.

3. Participants

Twelve native French speakers, aged between 18 and 29, and 12 native Spanish speakers, aged between 23 and 28, were tested individually. None of the participants had participated in the previous experiments, and none had a known hearing deficit. One Spanish participant had to be replaced due to too many complete reversals among his responses in the phoneme condition.

B. Results

Error rates for French and Spanish participants for the phonemic and the stress contrast as a function of sequence length are shown in Table V.

These data were subjected to an ANOVA with the between-participant factor language (French vs Spanish), and within-participant factors contrast (phoneme vs stress) and sequence length (2 vs 3 vs 4 vs 5 vs 6). The interaction between language and contrast was highly significant [$F(1,22)=61.6$, $p<0.0001$]. The interaction was due to the fact that stress yielded significantly more errors than phoneme with the French participants [$F(1,11)=81.1$, $p<0.0001$], whereas there was a nonsignificant trend in the other direction with the Spanish participants [$F(1,11)=3.5$, $0.1>p>0.05$]. *Post hoc* comparisons indicate a significant effect of contrast with the French participants for all sequence lengths (Bonferroni-corrected p values: 0.025, 0.001, 0.001, 0.001, and 0.01 for sequence length 2, 3, 4, 5, and 6, respectively). The sequence length by contrast interaction for the French participants was not significant [$F(4,44)=2.28$, $p=0.075$]. The Spanish participants showed no significant difference at any length ($p>0.1$). For the French participants, we ran a t-test at each sequence length to compare the performance in the stress condition to chance performance. The comparison was significant at lengths 2, 3, and 4 (Bonferroni-corrected one-tailed $p<0.003$, $p<0.003$, $p<0.01$, respectively) but not at lengths 5 and 6 (Bonferroni-corrected one-tailed $p=0.2$ and $p>0.2$, respectively). The overall performance across lengths with the stress contrast was significantly better than chance [$F(1,11)=22.94$, $p<0.001$].

The overlap between the two populations was 0%, with an optimal classification score of 100%. The distributions were actually separated by a gap whose size was 2.7% of the total range. As to the reliability test, the correlation coefficient r between the two halves of the experiment was 0.82 ($p<0.001$).

In this experiment, the number of training trials before criterion was saved. For the French participants, the number of training trials was 23.6 for the stress condition and 11.2 for the phoneme condition, a nonsignificant difference [$F(1,11)=3.10$, $p>0.1$]. The bulk of the difference was due to two participants who had more than 50 trials in the stress condition. For the Spanish participants, the number of training trials was 9.6 for the stress condition and 8.8 for the phoneme condition, again, a nonsignificant difference

TABLE V. Percent error with phoneme and stress contrast as a function of sequence length for 12 French and 12 Spanish participants in experiment 4.

Sequence length	2	3	4	5	6	Mean
			French			
Phoneme	7.3%	5.2%	21.9%	50.0%	63.5%	29.5%
Stress	34.4%	48.9%	75.0%	88.5%	94.8%	68.3%
			Spanish			
Phoneme	6.2%	9.4%	31.2%	56.2%	71.9%	35.0%
Stress	8.3%	13.5%	19.8%	38.5%	57.3%	27.5%

$[F(1,11)<1,\ p>0.1]$. In a global ANOVA, the interaction between contrast and language was only marginal $[F(1,22)=3.4,\ 0.1>p>0.05]$.

C. Discussion

This experiment was identical to experiment 3 except that two talkers were used instead of only one. This introduced a threefold increase in $F0$ variability across tokens, as well as other differences due to timbre, and fine phonetic variations of the segments and of suprasegmental information. Yet, this experiment replicated almost exactly experiment 3. In particular, the stress ''deafness'' effect was not stronger with two talkers than with one talker. The overall interaction between language and contrast, the overlap between the French and Spanish participants, as well as the reliability score, were all very similar in the two experiments. The only difference was that the overall error rate was slightly higher in this experiment than in experiment 3 (40% instead of 28%); this difference was probably due to the fact that the change in talker made the memory task more difficult (see Nygaard, Sommers and Pisoni, 1995). Consequently, the most difficult conditions, i.e., sequence lengths 5 and 6 with the stress contrast for the French participants, were at chance level.

In the last experiment, we tested whether memory load alone, without any phonetic variability, is sufficient to induce a stress deafness effect.

VI. EXPERIMENT 5

In this experiment, we tested whether French participants still display a stress deafness when there is no phonetic variability at all. That is, we used only a single token for each of the test items. With a single token, it is in principle possible to encode a sequence of stimuli in terms of ''same'' and ''different,'' these two categories being definable acoustically. Assuming that participants have access to same/different judgments at the acoustic level, and that they can encode this information in short-term memory in keeping with the presentation rate, we expected that they should be able to perform the task even with contrasts that are nonnative. Hence, we predicted that French and Spanish participants would have equal performance on the task with both the phonemic and the stress contrast.

A. Method

1. Materials

The same minimal pairs as in experiment 4 were used, but only a single token from the female voice for each item was used. The design was the same as in experiment 4.

2. Procedure

The same procedure as in experiment 4 was used.

3. Participants

Twelve native French speakers, aged between 17 and 50, and 12 native Spanish speakers, aged between 23 and 29, were tested individually. None of the participants had participated in one of the previous experiments, and none had a known hearing deficit.

B. Results

Error rates for French and Spanish participants for the phonemic and the stress contrast as a function of sequence length are shown in Table VI.

These data were subjected to an ANOVA with the between-participant factor language (French vs Spanish), and within-participant factors contrast (phoneme vs stress) and sequence length (2 vs 3 vs 4 vs 5 vs 6). The interaction between language and contrast was not significant $[F(1,22)<1,\ p>0.1]$. A separate analysis for the French and the Spanish participants revealed only an effect of sequence length for both groups $[F(4,44)=30.3,\ p<0.001,$ and $F(4,44)=36.0,\ p<0.001,$ respectively]. There was no significant effect of contrast either globally $[F(1,22)<1,\ p>0.1]$ or for individual sequence lengths $[p>0.1]$.

For the French participants, we ran a t-test at each sequence length to compare the performance in the stress condition to chance performance. The comparison was significant at each length (Bonferroni-corrected one-tailed $p<0.001$ at every length). The overall performance across lengths was significantly better than chance $[F(1,11)=131.89,\ p<0.0001]$.

The overlap between the two populations was 95.8%. The optimal classification score was 54.2%. As to the reliability test, the correlation coefficient r between the two halves of the experiment was 0.55 $(p<0.007)$.

In this experiment, the number of training trials before criterion was saved. For the French participants, the number of training trials was 10.1 for the stress and 9.7 for the phoneme condition, a nonsignificant difference $[F(1,11)<1,\ p>0.1]$. For the Spanish participants, the number of training trials was 8.9 for the stress and 8.3 for the phoneme condition, again, a nonsignificant difference $[F(1,11)<1,\ p>0.1]$. In a global ANOVA, there was no significant interaction between contrast and language $[F(1,22)>1,\ p>0.1]$.

C. Discussion

This experiment demonstrates that with no phonetic variability, no stress ''deafness'' emerges in the French participants even when the memory load is high. Note that although the performance in this experiment was overall better

TABLE VI. Percent error with phoneme and stress contrast as a function of sequence length for 12 French and 12 Spanish participants in experiment 5.

Sequence length	2	3	4	5	6	Mean
			French			
Phoneme	7.3%	6.2%	12.5%	39.6%	51.0%	23.3%
Stress	8.3%	12.5%	16.7%	38.5%	57.3%	26.7%
			Spanish			
Phoneme	3.1%	1.0%	7.3%	27.1%	47.9%	17.3%
Stress	1.0%	4.2%	12.5%	31.2%	47.9%	19.4%

than in the experiments with greater variability, we cannot explain the lack of a cross-linguistic difference by a ceiling effect. Indeed, even with sequence lengths matched in difficulty with those of the previous experiments (for instance, sequences 4 to 6 in the present experiment versus sequences 3 to 5 in experiment 4), no difficulty with stress emerged in French participants. This replicates and extends the findings in Dupoux *et al.* (1997), where it was reported that the problem of the French participants to discriminate stress disappears in an AX paradigm with no phonetic variability. What the current experiment shows is that the lack of phonetic variability is sufficient to make the stress deafness effect disappear; in particular, the presence of a high memory load alone does not induce the stress deafness effect.

VII. GENERAL DISCUSSION

In a series of five experiments, we demonstrated that the stress deafness effect in French listeners can be replicated with a new paradigm. Moreover, we have substantially improved the methodology, in that the results can be interpreted on an individual basis. In the following discussion we address three issues. First of all, we discuss the effect of memory load and phonetic variability on the group results. We then discuss the residual capacity of French participants to perceive the stress contrast. Finally, we consider the broader implications of our findings regarding the effects of the native language on speech perception.

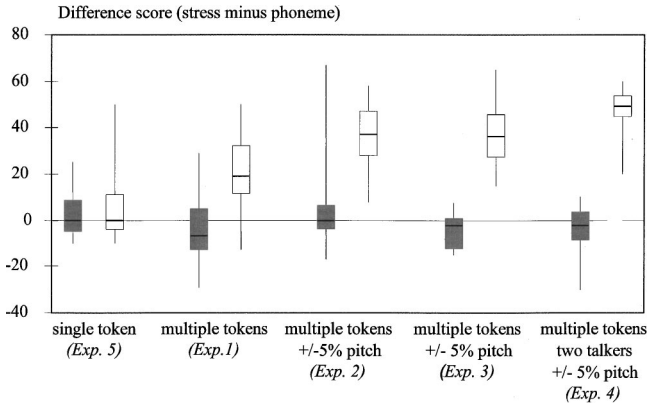


FIG. 1. Distribution of difference scores (phoneme minus stress) as a function of phonetic variability in five experiments. The Spanish difference scores are in gray, and the French in white. The minimum and maximum scores are indicated by the bottom and top of the vertical lines, respectively, the median scores by the thick horizontal lines, and the boxes contain scores that fall within the 25%–75% percentile.

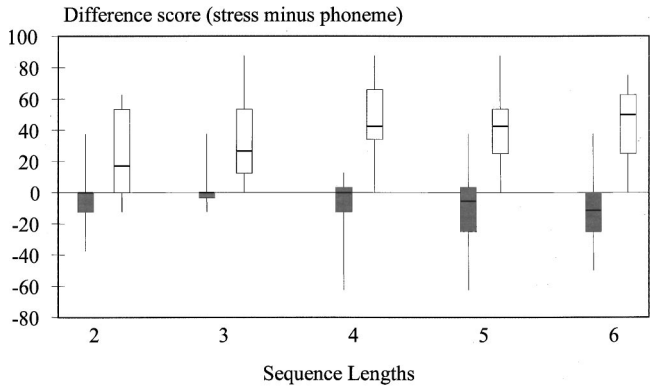


FIG. 2. Distribution of difference scores (phoneme minus stress) in experiments 3 and 4 as a function of sequence length. The Spanish difference scores are in gray, and the French in white. The minimum and maximum scores are indicated by the bottom and top of the vertical lines, respectively, the median scores by the thick horizontal lines, and the boxes contain scores that fall within the 25%–75% percentile.

A. Memory load and phonetic variability

Our findings can be summarized in three points. First, the size of the ''deafness'' effect increases with the amount of phonetic variability. The effect of phonetic variability, however, seems to reach a plateau (see Fig. 1). Specifically, we found that adding $\pm 5\%$ pitch variations on stimuli produced by a single talker (experiments 2 and 3) yielded a similar effect as adding this variation on stimuli produced by two talkers, one male and one female (experiment 4).

Second, although memory load had a strong effect on mean performance in that shorter sequences yielded less errors than longer ones, the stress-phoneme difference score in the French participants across experiments 1 to 4 was found to have roughly the same size at sequence length 2 and at sequence length 6 (means of 28.3% and 34.8%, respectively). In fact, if anything, the effect seemed numerically larger at sequence length 4 (mean of 41.3%). This might be due to the fact that error scores were squeezed by a floor effect at length 2 and a ceiling effect at length 6 (see Fig. 2). Note, however, that with sequences of size 1, as used in the training session, there was no longer a significant difference between stress and phonemes in French participants. In other words, it is the presence of memory load that matters, not the amount of it.

Third, phonetic variability and memory load displayed an interaction. On the one hand, in experiment 5 with zero phonetic variability, no deficit with the stress contrast was found in the French participants, not even with sequences of size 6. On the other hand, in the training sessions (sequence length 1), we found no measurable problem to master the stress contrast, not even if there was a high phonetic variability as in experiment 4 (two talkers, extra pitch variation). Hence, it is only in the presence of both factors that the ''deafness'' effect emerges.

These findings can be interpreted as follows. The sequence repetition task requires participants to encode the information in their short-term memory buffer in order to recall the sequence. Several coding strategies are available according to the task demands and stimulus characteristics.

TABLE VII. Coding strategies available to Spanish and French participants as a function of memory load and phonetic variability.

	Spanish participants	French participants
Sequence 1, no variability (training of Expt. 5)	Explicit categorization Acoustic mismatch Phonological coding	Explicit categorization Acoustic mismatch
Sequence 1, pitch variability (training of Expt. 2–4)	Explicit categorization Phonological coding	Explicit categorization
Sequence 2–6, no variability (Expt. 5)	Acoustic mismatch Phonological coding	Acoustic mismatch
Sequence 2–6, pitch variability (Expt. 2–4)	Phonological coding	No strategy available

We distinguish two acoustic and one phonological strategy (see Table VII). A first coding strategy is based on the fact that stress has massive acoustic correlates. Given enough time, participants can use these cues to explicitly categorize the stimuli, for instance, by comparing them to stored exemplars, by focusing on the melody, or by repeating the token and monitoring for one's own pitch. This strategy is not automated and, therefore, cannot be used in case of long and rapid sequences of stimuli. A second coding strategy is based on the acoustic mismatch signal. This signal has been shown to be automatically generated in the auditory cortex whenever an acoustic signal differs from its predecessor(s) (Näätanen *et al.*, 1997). This acoustic mismatch might be used to recode sequences of stimuli in terms of same/different. The last token can then be explicitly categorized in order to reconstruct the underlying sequence. Of course, such a strategy only works in the absence of phonetic variability. When the tokens are phonetically varied, even sequences of ''same'' tokens give rise to a mismatch signal. A final coding strategy uses the automatic encoding which is provided by the language-specific phonological representation. In the case of stress, such phonological encoding is of course available to the Spanish but not to the French participants participants, since stress is not encoded phonologically in the latter language.

Our analysis of the available strategies has a practical impact regarding how to construct a paradigm that will show maximal cross-linguistic sensitivity. In order to eliminate nonphonological compensatory strategies it is essential to limit the amount of time that subjects have available. In all of our experiments, the duration of the tokens was rather short, and the interstimulus interval, ISI, between them was only 80 ms. Although we have not explicitly varied this variable, informal testing reveals that if participants were given more time—by having either longer items or a longer ISI—they would be able to explicitly recode the tokens (as they do in the training phase, where they have all the time necessary to produce a response). This, then, would have the effect of reducing the size of the cross-linguistic difference.

B. Residual capacity to discriminate stress

In all our experiments, we found that French participants, although they made massively more errors with the stress than with the phoneme contrast, were still significantly better than chance. This was true even for all the individual sequence lengths except 6, which was not significantly different from chance in experiments 2, 3, and 4. Note, however, that it is a matter of experimental power; that is, when these three experiments are pooled together, the performance at sequence length 6 is still better than chance $[t(35) = 2.90, p < 0.006]$.

We would like to discuss three possible interpretations of this residual effect. First, an acoustic strategy might survive both long sequence lengths and phonetic variability. This is somewhat unlikely, since, for instance, sequences of length 6 last for about 4 s, which is longer than the span for the echoic store (Guttman and Julesz, 1963; Crowder and Morton, 1969). Second, there might be a residual encoding of stress within the phonological representation itself. This encoding should be much weaker or less accessible in French than in Spanish participants. One possible reason why French participants would not entirely ignore stress is that they use stress in word segmentation. In French, stress falls on the word's final nonschwa syllable. Items with word-initial stress like [píki] might therefore be perceived as containing a word boundary after the first syllable. Consequently, [píki] and [pikí], although identical in their segmental contents, would differ in their perceived segmentation pattern.[3] Third, there might be another type of representation, i.e., a phonetic representation, which can be used to encode the stimuli in a language-universal fashion, but yet in a more abstract format than the acoustic representation. Goldinger (1998) has shown, for instance, that in an immediate repetition task, participants can imitate certain phonetic details of stimuli that are not used phonologically in their native language. More research is needed to tease apart these various interpretations.

C. Implications for future research

Our results raise a number of issues regarding the influence of the maternal language on speech perception. We found that French speakers have difficulties with distinguishing a stress contrast, but only with phonetically variable stimuli, high memory load, and limited time.

This suggests that, first, the extent of influences of the maternal language might be underestimated when using standard discrimination procedures for which participants have much time to perform the task, and which usually do not introduce phonetic variability. It is, therefore, important to study speech processing under constrained resources, in order to limit the possibility for participants to use nonlinguistic response strategies.

Second, our findings raise the issue of the specificity of suprasegmental features. Indeed, we showed that for French participants, acoustically based response strategies are readily available. This is clearly not the case for the perception of non-native consonantal contrasts (see Goto, 1971). It remains to be investigated whether the availability of acoustically-based response strategies depends upon the size of the acoustic correlates, and whether it extends to other suprasegmental contrasts.

Third, our findings raise the issue of the linguistic and developmental conditions under which stress ''deafness'' arises during language acquisition. Do all languages with noncontrastive stress yield stress ''deafness'' (Dupoux and Peperkamp, in press; Peperkamp and Dupoux, in press)? Conversely, can speakers of languages with contrastive stress exhibit a stress ''deafness'', provided the number of minimal stress pairs in their language is small (Cutler, 1986)?

A final question is whether native speakers of French who learn Spanish can acquire the perception of stress, and whether age of acquisition matters in this respect (Flege, MacKay, and Meador, 1999; Flege, Schmidt, and Wharton, 1996; Pallier, Bosch, and Sebastián-Gallés, 1997; Peperkamp, Dupoux, and Sebastián-Gallés, 1999).

In brief, the existence of stress ''deafness'' raises a number of central issues in speech perception and language acquisition. The availability of an easy to use and reliable methodology allows us to address these issues in the near future.

ACKNOWLEDGMENTS

This research was supported in part by a grant from the Fyssen Foundation to the second author, by the Groupement d'Interet Scientifique: Sciences de la Cognition, and by the Spanish Ministerio de Educación y Cultura, Grant No. PB97-0977. This project benefited from many discussions with Jacques Mehler, who also invented the term stress ''deafness.'' We thank Katherine White and Mar Rodríguez for help in running participants, and Laura Bosch, Anne Christophe, Christophe Pallier, Franck Ramus, and two anonymous reviewers for comments and discussion.

APPENDIX: SEQUENCES USED IN THE EXPERIMENTS

Two-word sequences: (experiments 1–5) 11, 12, 21, 22, 11, 12, 21, 22.

Three-word sequences: (experiments 3, 4, 5) 111, 112, 121, 122, 211, 212, 221, 222.

Four-word sequences: (experiments 1–5) 1121, 1122, 1211, 1221, 2111, 2112, 2122, 2212.

Five-word sequences: (experiments 3, 4, 5) 11121, 12112, 12122, 12211, 21112, 21211, 21221, 22122.

Six-word sequences: (experiments 1–5) 112121, 112212, 121112, 122121, 211221, 212112, 221212, 222121.

[1]Syntrillium Software Corporation (www.synrtillium.com).

[2]We used the compression algorithm in the PRAAT software (www.fon.hum.uva.nl/praat/).

[3]Another possibility stems from the fact that word stress is final in French. Stress-final items could be perceived as more prototypical than stress-initial ones. Prototypicality could then be used to distinguish the two sets of items. However, Dupoux *et al.* (1997) compared discrimination of [vasúma] vs [vasumá] (where the latter is the more prototypical one) with that of [vásuma] vs [vasúma] (where the two are equally nonprototypical), and failed to find a difference.

Best, C., McRoberts, G., and Sithole, N. (**1988**). ''Examination of perceptual reorganization for non-native speech contrasts; Zulu click discrimination by English-speaking adults and infants,'' J. Exp. Psychol. Hum. Percept. Perform. **14**, 345–360.

Best, C. (**1994**). ''The emergence of native-language phonological influence in infants; A perceptual assimilation model,'' in *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, edited by J. Goodman and H. Nusbaum (MIT Press, Cambridge), pp. 167–224.

Bluhme, S., and Burr, R. (**1971**). ''An audio-visual display of pitch for teaching Chinese tones,'' Studies Ling. **22**, 51–57.

Crowder, R.G., and Morton, J. (**1969**). ''Precategorical acoustic store (PAS),'' Percept. Psychophys. **5**, 365–373.

Cutler, A. (**1986**). ''*Forbear* is a homophone: Lexical prosody does not constrain lexical access,'' Lang. Speech **29**, 201–220.

Dupoux, E., Pallier, C., Sebastián-Gallés, N., and Mehler, J. (**1997**). ''A destressing 'deafness' in French?'' J. Memory Lang. **36**, 406–421.

Dupoux, E., and Peperkamp, S. (**in press**). ''Fossil markers of language development: Phonological 'deafnesses' in adult speech processing,'' in *Cognitive Phonology*, edited by J. Durand and B. Laks (Oxford University Press, Oxford).

Flege, J.E. (**1995**). ''Second language speech learning: Theory, findings, and problems,'' in *Speech Perception and Linguistic Experience*, edited by W. Strange (York, Baltimore), pp. 233–272.

Flege, J., MacKay, I., and Meador, D. (**1999**). ''Native Italian speakers' perception and production of English vowels,'' J. Acoust. Soc. Am. **106**, 2973–2987.

Flege, J., Schmidt, A., and Wharton, G. (**1996**). ''Age of learning affects rate-dependent processing of stops in a second language,'' Phonetica **53**, 143–161.

Francis, A., Baldwin, K., and Nusbaum, H. (**2000**). ''Effects of training on attention to acoustic cues,'' Percept. Psychophys. **62**, 1668–1680.

Gandour, J.T. (**1983**). ''Tone perception in Far Eastern languages,'' J. Phonetics **11**, 149–175.

Goldinger, S.D. (**1998**). ''Echoes of echoes? An episodic theory of lexical access,'' Psychol. Rev. **105**, 251–279.

Goto, H. (**1971**). ''Auditory perception by normal Japanese adults of the sounds 'l' and 'r','' Neuropsychologia **9**, 317–323.

Guttman, N., and Julesz, B. (**1963**). ''Lower limits of auditory periodicity analysis,'' J. Acoust. Soc. Am. **35**, 610.

Jusczyk, P.W. (**1993**). ''From general to language-specific capacities: the WRAPSA Model of how speech perception develops,'' J. Phonetics **21**, 3–28.

Jusczyk, P.W., Friederici, A.D., Wessels, J.M.I., Svenkerud, V.Y., and Jusczyk, A.M. (**1993**). ''Infants' sensitivity to the sound patterns of native language words,'' J. Memory Lang. **32**, 402–420.

Kiriloff, C. (**1969**). ''On the auditory perception of tones in Mandarin,'' Phonetica **20**, 63–67.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., and Lindblom, B. (**1992**). ''Linguistic experience alters phonetic perception in infants by 6 months of age,'' Science **255**, 606–608.

Kuhl, P.K. (**2000**). ''Language, mind, and brain: Experience alters perception,'' in *The New Cognitive Neuroscience*, edited by M. Gazzaniga (MIT Press, Cambridge, MA), pp. 99–114.

Lee, L., and Nusbaum, H. (**1993**). ''Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese,'' Percept. Psychophys. **53**, 157–165.

Lively, S.E., Logan, J.S., and Pisoni, D.B. (**1993**). ''Training Japanese listeners to identify English /r/ and /l/. II. The role of phonetic environment and talker variability in learning new perceptual categories,'' J. Acoust. Soc. Am. **94**, 1242–1255.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins, J., and Fujimura, O. (**1981**). ''An effect of linguistic experience; The discrimination of /r/ and /l/ by native speakers of Japanese and English,'' Percept. Psychophys. **18**, 331–340.

Morton, J., Marcus, S., and Ottley, P. (**1981**). ''The acoustic correlates of speechlike: A use of the suffix effect,'' J. Exp. Psychol. **110**, 568–593.

Morton, J., Crowder, R., and Prussin, H. (**1971**). ''Experiments with the stimulus suffix effect,'' J. Exp. Psychol. **91**, 169–190.
Näätanen, R., Lehtokovski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R., Luuk, A., Allik, J., Sinkkonen, J., and Alho, K. (**1997**). ''Language-specific phoneme representations revealed by electric and magnetic brain responses,'' Nature (London) **385**, 432–434.
Nusbaum, H., and Goodman, J. (**1994**). ''Learning to hear speech as spoken language,'' in *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, edited by H. Nusbaum and J. Goodman (MIT Press, Cambridge, MA), pp 299–338.
Nygaard, L.C., Sommers, M.S., and Pisoni, D. (**1995**). ''Effects of stimulus variability on perception and representation of spoken words in memory,'' Percept. Psychophys. **57**, 989–1001.
Pallier, C., Bosch, L., and Sebastián-Gallés, N. (**1997**). ''A limit on behavioral plasticity in speech perception,'' Cognition **64**, B9–B17.
Peperkamp, S., Dupoux, E., and Sebastián-Gallés, N. (**1999**). ''Perception of stress by French, Spanish, and bilingual subjects,'' Proceedings of Euro-Speech '99, Vol. 6, 2683–2686.
Peperkamp, S., and Dupoux, E. (**in press**). ''A typological study of stress 'deafness','' in *Laboratory Phonology VII*, edited by C. Gussenhoven and N. Warner (Mouton de Gruyter, Berlin).
Polivanov, E. (**1931**). ''La perception des sons d'une langue étrangère,'' Travaux du Cercle Linguistique de Prague **4**, 79–96.
Sapir, E. (**1921**). *Language* (Harcourt Brace Jovanovich, New York).
Wang, Y., Spence, M., Jongman, A., and Sereno, J. (**1999**). ''Training American listeners to perceive Mandarin tones,'' J. Acoust. Soc. Am. **106**, 3649–3658.
Werker, J., and Tees, R. (**1984**). ''Cross language speech perception: Evidence for perceptual reorganization during the first year of life,'' Infant Behav. Dev. **7**, 49–63.

Minimum spectral contrast needed for vowel identification by normal hearing and cochlear implant listeners

Philipos C. Loizou[a] and Oguz Poroy
Department of Electrical Engineering, University of Texas at Dallas, P.O. Box 830688, EC 33, Richardson, Texas 75083-0688

(Received 23 December 2000; accepted for publication 31 May 2001)

The minimum spectral contrast needed for vowel identification by normal-hearing and cochlear implant listeners was determined in this study. In experiment 1, a spectral modification algorithm was used that manipulated the channel amplitudes extracted from a 6-channel continuous interleaved sampling (CIS) processor to have a 1–10 dB spectral contrast. The spectrally modified amplitudes of eight natural vowels were presented to six Med-El/CIS-link users for identification. Results showed that subjects required a 4–6 dB contrast to identify vowels with relatively high accuracy. A 4–6 dB contrast was needed independent of the individual subject's dynamic range (range 9–28 dB). Some cochlear implant (CI) users obtained significantly higher scores with vowels enhanced to 6 dB contrast compared to the original, unenhanced vowels, suggesting that spectral contrast enhancement can improve the vowel identification scores for some CI users. To determine whether the minimum spectral contrast needed for vowel identification was dependent on spectral resolution (number of channels available), vowels were processed in experiment 2 through n ($n=4,6,8,12$) channels, and synthesized as a linear combination of n sine waves with amplitudes manipulated to have a 1–20 dB spectral contrast. For vowels processed through 4 channels, normal-hearing listeners needed a 6 dB contrast, for 6 and 8 channels a 4 dB contrast was needed, consistent with our findings with CI listeners, and for 12 channels a 1 dB contrast was sufficient to achieve high accuracy (>80%). The above-mentioned findings with normal-hearing listeners suggest that when the spectral resolution is poor, a larger spectral contrast is needed for vowel identification. Conversely, when the spectral resolution is fine, a small spectral contrast (1 dB) is sufficient. The high identification score (82%) achieved with 1 dB contrast was significantly higher than any of the scores reported in the literature using synthetic vowels, and this can be attributed to the fact that we used natural vowels which contained duration and spectral cues (e.g., formant movements) present in fluent speech. The outcomes of experiments 1 and 2, taken together, suggest that CI listeners need a larger spectral contrast (4–6 dB) than normal-hearing listeners to achieve high recognition accuracy, not because of the limited dynamic range, but because of the limited spectral resolution. [DOI: 10.1121/1.1388004]

PACS numbers: 43.71.Es, 43.71.Ky, 43.66.Ts [CWT]

I. INTRODUCTION

The vowel spectra are typically characterized by high-amplitude peaks and relatively low-amplitude valleys. Although the frequencies of the spectral peaks are considered to be the primary cues to vowel identity, the spectral contrast, i.e., the difference between the spectral peak and the spectral valley, needs to be maintained to some extent for accurate vowel identification. The importance of spectral contrast in vowel identification was investigated by Leek *et al.* (1987) using four vowel-like complexes constructed as a sum of 30 100 Hz harmonics. The amplitudes of two consecutive harmonics that defined the (formant) peaks appropriate for the vowels /i ae a U/ varied over a range of 1–8 dB above background harmonics. Results showed that normal-hearing listeners required a 1–2 dB peak-to-valley difference to identify four vowel-like harmonic complexes with relatively high (75% correct) accuracy. Alcantara and Moore (1995) showed that the minimum spectral contrast needed for vowel identification depended on, among other factors, the fundamental frequency, presentation level, and the component phase (cosine versus random) used for the synthesis of the harmonic complexes. Vowel identification was higher with cosine phase, and improved with higher presentation levels and lower fundamental frequency (50 Hz). Subjects needed a 3 dB contrast to identify six vowel-like harmonic complexes with 75% accuracy. In another study, Turner and Van Tasell (1984) showed that normal-hearing listeners could detect a 2 dB notch in a vowel-like spectrum. In summary, the above-mentioned studies indicate that only a small spectral contrast is needed by normal-hearing listeners for vowel identification.

This remarkable ability of the normal auditory system to detect small amplitude changes in the spectrum was not observed in hearing-impaired listeners (Turner and Holte, 1987; Leek *et al.*, 1987). Leek *et al.* (1987) have shown that listeners with a flat, moderate hearing loss required a 6–7 dB peak-to-valley difference for vowel identification. This was attributed to the lack of suppression and the abnormally

[a]Author to whom correspondence should be addressed; electronic mail: loizou@utdallas.edu

broad auditory filters associated with hearing loss (e.g., Pick *et al.*, 1977; Wightman *et al.*, 1977). Spectral contrast is reduced when vowels are processed through broad filters due to the shallow filter roll-off. As a result, the internal vowel representation is ''blurred'' leading to poorer vowel identification.

Unlike normal-hearing and hearing-impaired listeners, cochlear implant (CI) listeners have a limited spectral resolution and a limited dynamic range. Spectral contrast is reduced in cochlear implant listeners, not because of the abnormally broad auditory filters—which are bypassed with electrical stimulation—but primarily because of the reduced dynamic range and amplitude compression. The large acoustic dynamic range is typically compressed in implant speech processors using a logarithmic function to a small electrical dynamic range, 5–15 dB. This compression results in a reduction of spectral contrast. We believe that the reduction in spectra contrast was one of the reasons that vowel recognition performance decreased in studies looking for the effect of reduced dynamic range on speech recognition (Zeng and Galvin, 1999; Loizou *et al.*, 2000). Another factor that could potentially reduce spectral contrast is the steepness of the compression function used for mapping acoustic amplitudes to electric amplitudes. A highly compressive mapping function, for instance, would yield a small spectral contrast, even if the dynamic range were large. It is therefore conceivable that a patient may have a large dynamic range, but a small effective spectral contrast because of a steep mapping function. The sensitivity setting, which affects the input gain, can also affect the spectral contrast. If a patient sets the sensitivity too high, then the acoustic amplitudes would be mapped to the high end of the compression function (above the knee point) producing a relatively flat (i.e., small spectral contrast) electrical channel amplitude pattern. Last, additive background noise could also reduce spectral contrast (e.g., Leek and Summers, 1996), probably to a larger degree in cochlear implant listeners compared to normal-hearing listeners due to the limited electrical dynamic range.

Given the above-mentioned factors that could reduce spectral contrast in CI users and consequently affect vowel identification in quiet or in noise, then what is the minimum spectral contrast needed for vowel identification by cochlear implant listeners? The answer to this question is important for the design of CIs for two main reasons. First, it will tell us whether current speech processing strategies preserve enough spectral contrast information needed for vowel identification. Second, it will help us devise new speech processing strategies that will enhance the incoming signal to have a certain spectral contrast. Such strategies could potentially be used to enhance vowel recognition in quiet or in noise. Hawks *et al.* (1997), for instance, increased the vowel spectral contrast by narrowing the formant bandwidths, and noted improvement in vowel identification. Given that the number of channels currently supported by commercial implant processors (Loizou, 1998) varies from a low of 6 channels to a high of 22 channels, it is very important to also ask the question whether the minimum spectral contrast needed for vowel identification is dependent on spectral resolution, i.e., the number of channels available. The above-mentioned questions are addressed in experiments 1 and 2 using cochlear implant listeners and normal-hearing listeners, respectively.

In experiment 1, six CI users fitted with a Continuous Interleaved Sampling (CIS) processor are used to determine the minimum spectral contrast needed for vowel identification. Vowel stimuli are processed off-line and the channel amplitudes are manipulated to have a peak-to-trough ratio ranging from 1 to 10 dB. In experiment 2, normal-hearing listeners are used to investigate possible interaction between spectral resolution (number of channels) and spectral contrast, based on the hypothesis that there might be a trade-off relationship between spectral resolution (number of channels) and spectral contrast. This hypothesis is based on the view that when speech is processed through a small number of channels the relative differences in across-channel amplitudes must be used to code frequency information. In this view, if spectral contrast was reduced, then vowel recognition ought to decline. On the other hand, when speech is processed through a large number of channels, a large spectral contrast might not be needed, since the frequency information can be coded by the channels that have energy. These questions are investigated in experiment 2 with normal-hearing listeners, where we assess speech intelligibility as a function of number of channels and as a function of spectral contrast. Normal-hearing listeners are used because the channels and contrast manipulations cannot be independently controlled with implant listeners due to the many confounding factors associated with electrical stimulation. To produce speech with varying degrees of spectral resolution and varying degrees of spectral contrast, we synthesized speech as a linear combination of sine waves and manipulated the amplitudes of the sine waves to have a 1–20 dB peak-to-trough ratio.

II. EXPERIMENT 1: MINIMUM VOWEL SPECTRAL CONTRAST NEEDED BY COCHLEAR-IMPLANT LISTENERS

A. Method

1. Subjects

The subjects were six postlingually deafened adults who had used a 6-channel CIS processor for periods ranging from three to four years. All the patients had used a 4-channel, compressed-analog signal processor (Ineraid) for at least four years before being switched to a CIS processor. The patients ranged in age from 40 to 68 years and they were all native speakers of American English. Biographical data for each patient are presented in Table I. All subjects were fitted with a 6-channel CIS processor, except for subject S1, who was fitted with a 5-channel processor.

2. Vowel stimuli

Eight monopthong vowels produced by a male speaker were used for testing. The vowels were contained in the words ''heed, hid, head, had, hod, hud, hood, who'd,'' and were produced by a male speaker ($F0=115$ Hz) randomly selected from the vowel database used by Hillenbrand *et al.*

TABLE I. Biographical data of the six cochlear-implant users who participated in this study.

Subject	Gender	Age (years) at detection of hearing loss	Age at which hearing aid gave no benefit	Age fit with Ineraid	Age at testing	Etiology of hearing loss	Score on H.I.N.T sentences in quiet	Score on NU-6 words in quiet
S1	F	10	46	47	55	Unknown	44	20
S2	F	7	31	33	40	Unknown/ hereditary	100	80
S3	F	23	48	51	57	Unknown	100	71
S4	M	5	43	48	58	Unknown	92	43
S5	M	20	46	63	68	Unknown	88	46
S6	M	19	19	29	41	Cogan's syndrome	100	93

(1995). The vowel formant frequencies, estimated at the steady-state portion of the vowel, are shown in Table II.

3. CIS implementation and experimental setup

The vowel stimuli were first processed off-line using the CIS strategy, saved in a file, and then presented to the CI listeners. Off-line processing was used to ensure that the channel amplitudes had the desired peak-to-trough ratio.

The CIS strategy, which involves bandpass filtering, amplitude envelope estimation and compression, was implemented in MATLAB. Signals were first processed through a pre-emphasis filter (2000 Hz cutoff), with a 3 dB/octave rolloff, and then bandpassed into 6 frequency bands using sixth-order Butterworth filters. The center frequencies of the six bandpass filters were 461, 756, 1237, 2025, 3316, and 5428 Hz. The envelopes of the filtered signals were extracted by full-wave rectification and low-pass filtering (second-order Butterworth) with a 400 Hz cutoff frequency. The six envelope amplitudes $A_i\,(i=1,2,...,6)$ were mapped to electrical amplitudes E_i using a logarithmic transformation:

$$E_i = c\,\log(A_i) + d, \tag{1}$$

where c and d are constants chosen so that the electrical amplitudes fall within the range of threshold and most-comfortable levels. The electrical amplitudes E_i were processed through a spectral contrast algorithm (see Sec. II A 4) which manipulated the six channel amplitudes, estimated in each cycle, to have a prescribed peak-to-trough ratio. The spectrally enhanced channel amplitudes were saved in a file, and the experimental setup shown in Fig. 1 was used to load the saved channel amplitudes. The envelope amplitudes were finally used to modulate biphasic pulses of duration 40 μs/phase at a stimulation rate of 2100 pulses/s. The electrodes were stimulated in the same order as in the subjects' daily processors. For most subjects, the electrodes were stimulated in ''staggered'' order. The sensitivity setting on our laboratory speech processor was fixed and was identical for all subjects.

TABLE II. The formant frequencies of the vowels used in this study.

Vowel	$F1$ (Hz)	$F2$ (Hz)	$F3$ (Hz)
(h)a(d)	647	1864	2561
(h)o(d)	871	1204	2595
(h)ea(d)	555	1851	2624
(h)i(d)	441	2080	2721
(h)ee(d)	367	2390	2777
(h)oo(d)	468	1115	2492
(h)u(d)	629	1146	2643
(wh)o('d)	366	919	2378

The experiments were performed on our laboratory cochlear implant processor (Poroy and Loizou, 2000) using the experimental setup shown in Fig. 1. To accommodate for off-line data file processing, an I/O card (installed in the PC) was used. The six output lines of the I/O card in the PC were connected to six general-purpose I/O pins of the DSP in the laboratory speech processor, forming a 6 bit, parallel, unidirectional data bus. Since the cochlear implant was connected to the DSP during the experiments, it was necessary to isolate the PC from the rest of the circuitry, which was battery powered. This was achieved using three Burr–Brown ISO150 dual, isolated, digital coupling chips. The speech materials were pre-processed as described in the following and the amplitudes of the current pulses to be presented to the electrodes were stored in binary data files in the hard drive of the PC. During the experiments, these files were downloaded to the DSP over the isolated data bus, and were read in and stored in RAM by an assembly program running on the DSP. Finally, the amplitude data were retrieved word by word from RAM and sent to the current sources using a serial port in the DSP. A MATLAB interface program was used for loading and ''playing back'' the binary data files.

4. Spectral contrast enhancement algorithm

Unlike previous studies on spectral contrast (e.g., Leek *et al.*, 1987; Turner and Van Tassell, 1984; Alcantara and

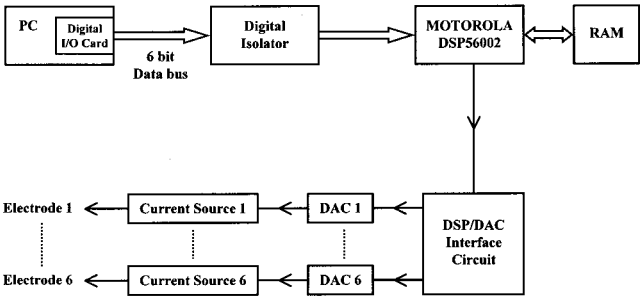


FIG. 1. Block diagram of the experimental setup.

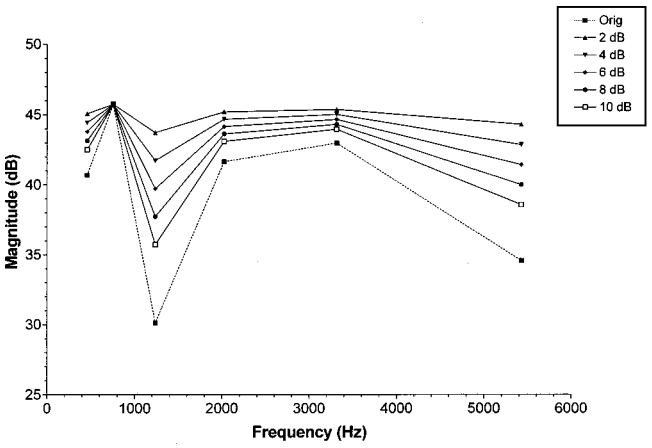


FIG. 2. Example of spectral modification of the vowel /ε/ to 2–10 dB contrast. The original, unenhanced, channel amplitudes are shown in the dotted line.

Moore, 1995) which manipulated synthetic vowels, this study manipulated naturally produced vowels. The main advantage in using natural vowels over synthetic vowels is that the natural stimuli contain both dynamic and static spectral cues commonly present in fluent speech. Manipulating the spectrum of natural vowels to have a certain peak-to-trough ratio, however, is not as simple as manipulating synthetic vowels. Simply identifying the valley, and modifying the amplitude of the valley (while fixing the peak amplitude) to have a certain peak-to-valley ratio, is not sufficient, because such a change could distort the spectrum. Likewise, identifying the peak, and modifying the amplitude of the peak (while fixing the valley amplitude) to have a certain peak-to-valley ratio, will most likely alter the shape of the spectrum as well. In addition, the latter method may introduce peak clipping, i.e., the modified spectral peak amplitude may be larger than then most comfortable level (MCL), and therefore will need to be clipped to the MCL level.

A spectral contrast enhancement algorithm, which addresses the above-mentioned issues (peak clipping, spectral distortion, etc.), is proposed in this study. The algorithm is implemented in the logarithmic domain and therefore assumes that the channel amplitudes are expressed in dB units. Let E_p and E_v represent the amplitudes (in dB) of the peak and valley, respectively, of the electrical amplitudes E_i. The amplitudes E_p and E_v are estimated by finding the maximum and minimum amplitudes, respectively, of the first four channel amplitudes $20\log(E_i)$, $i=1,2,3,4$ [the first four channels cover the $F1$–$F2$ frequency region]. Then, the spectrally enhanced channel amplitudes C_i (in dB) can be obtained as

$$C_i=\frac{E_i^*-E_v}{E_p-E_v}SR+E_p-SR, \quad i=1,2,...,6, \tag{2}$$

where $E_i^*=20\log(E_i)$, and SR is the desired spectral contrast in dB. Finally, the spectrally enhanced amplitudes C_i are converted back to the linear domain using the equation: $10^{C_i/20}$. Equation (2) preserves the peak amplitude and modifies not only the valley amplitude but also the other amplitudes in order to preserve the shape of the original spectrum. Figure 2 shows examples of the spectral contrast algorithm applied to the vowel /ε/. Note that the spectrally modified amplitudes, C_i, never exceed the MCL level, since the original peak amplitude is preserved [this can be verified by setting $E_i^*=E_p$ in Eq. (2)]. By preserving the peak amplitude we avoid peak-clipping problems. There is a possibility, however, that the spectrally modified amplitudes may fall below the threshold level, and in those cases, we set the corresponding channel amplitudes to threshold. This step was necessary to ensure that the modified channel amplitudes were within the subject's dynamic range.

The above-mentioned spectral contrast algorithm was applied only to the vocalic segment of the /hVd/ words. The vocalic segment was extracted from the /hVd/ words by manually removing the first and last pitch periods of the onset and offset of the vowel. Equation (2) was applied to all sets of 6-channel amplitudes computed using the CIS strategy within the vocalic segment of the word. The channel amplitudes estimated for the remaining portion (i.e., the silence and the [h], [d] segments) of the words were set to the threshold values. To avoid possible click sensations, the new onsets and offsets of the vowels were tapered off with a half Hamming window, 20 ms in duration.

5. Procedure

A total of 6 different sets of vowels was created with different spectral contrasts (1, 2, 4, 6, 8 and 10 dB) and presented to CI listeners for identification. For comparative purposes, we also presented the vowels processed through the CIS strategy, but were not modified. There were 9 repetitions of each vowel, presented in blocks of 3 repetitions each. The 7 sets of vowels were completely randomized within each block. The test session was preceded by one practice session in which the identity of the vowel was indicated to the listeners.

The stimuli were presented directly to the subjects through our laboratory processor at a comfortable listening level. To collect responses, a graphical interface was used that allowed the subjects to identify the vowels they heard by clicking on the corresponding button on the graphical interface.

B. Results and discussion

The results, scored in percent correct, for the different spectral contrasts are shown in Fig. 3. Repeated measures analysis of variance indicated a significant main effect of peak-to-trough ratio [$F(6,30)=10.49$, $p<0.005$] on vowel recognition. Performance increased monotonically as the peak-to-trough ratio increased from 1 to 4 dB, and leveled off thereafter. Post-hoc analysis (according to Fisher's LSD) showed that the scores obtained at 4 and 6 dB were not significantly different ($p=0.784$). Neither were the scores obtained at 4 dB and the unenhanced condition significantly different ($p=0.593$). The scores obtained at 2 and 4 dB were not significantly different ($p=0.173$), but the scores obtained at 2 and 1 dB were significantly different ($p<0.05$).

The individual subjects' performance on vowel recognition is shown in Fig. 4. The subjects' performance varied considerably as a function of peak-to-trough ratio. Most sub-

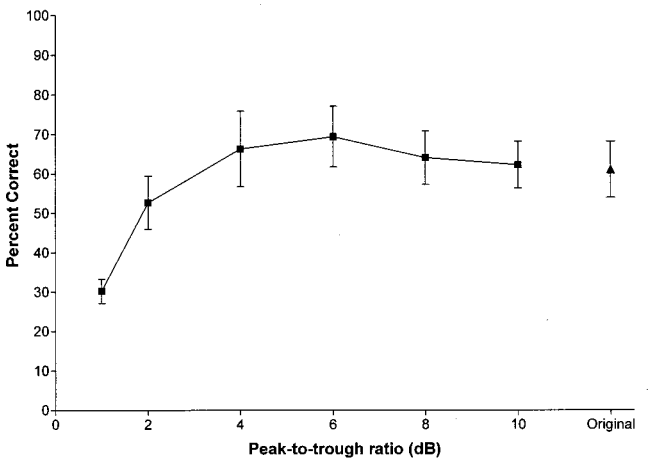


FIG. 3. Mean performance of cochlear-implant listeners on vowel recognition as a function of spectral contrast. Error bars indicate ± standard errors of the mean.

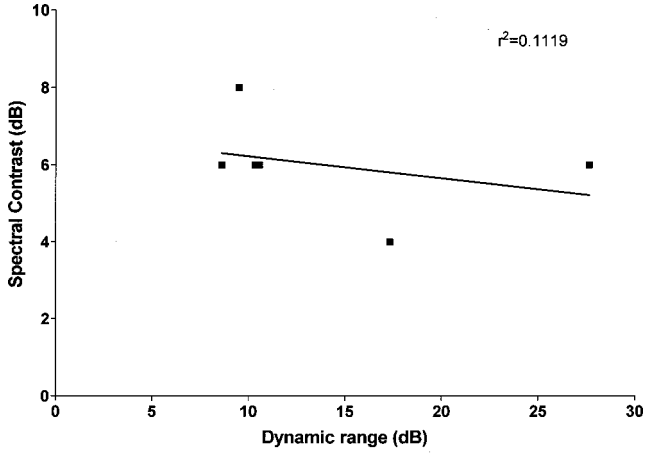


FIG. 5. Correlation between average (across all electrodes) electrical dynamic range and the amount of spectral contrast needed to achieve maximum vowel recognition performance.

jects (S2, S4, S5, S6) achieved maximum performance at 6 dB peak-to-trough ratio, one subject (S3) achieved maximum performance at 4 dB, while another subject (S1) achieved maximum performance at 8 dB peak-to-trough ratio. Vowel recognition performance declined for subjects S3 and S4 when the peak-to-trough ratio became larger than 4 dB. We suspect that this was due to the fact that the dynamic range of some electrodes was smaller than 10 dB for some subjects. For instance, the average dynamic range of electrodes 5 and 6 for subject S4 was 6 dB, i.e., it was smaller than the tested peak-to-trough ratio. In this case, over enhancing the channel amplitudes might have the same effect as turning off individual electrodes, since enhanced amplitudes smaller than the threshold levels were set to the threshold levels. Subject S1 needed an 8 dB peak-to-trough ratio to reach asymptotic performance. We suspect that this may be due to the fact that she was fitted with a 5-channel processor, compared to the other subjects who were fitted with 6-channel processors. This outcome suggests the possibility that a larger spectral contrast is needed for subjects receiving a small number of independent channels of stimulation. This hypothesis is investigated further in experiment 2.

The outcome that subjects achieved maximum vowel recognition performance at different levels of spectral contrast led us to wonder whether that was related to the subject's dynamic range, which ranged from a low of 9 dB for some subjects to a maximum of 28 dB for others. That is, were the subjects with the larger dynamic range the ones requiring larger spectral contrast to achieve maximum levels of performance? This was based on the assumption that subjects with a wide dynamic range should have a slow growth of loudness; hence they should require a larger spectral contrast for the same loudness difference. Similarly, were the subjects with the smaller dynamic range the ones requiring smaller spectral contrast? To answer these questions, we performed correlation analysis (Fig. 5) between the average (across all electrodes) dynamic range and the amount of spectral contrast needed to achieve maximum performance. The resulting correlation (Pearson's) coefficient between dynamic range and spectral contrast was very weak ($r = 0.334$) and nonsignificant ($p = 0.517$). As shown in Fig. 5, subject S6 who had a large dynamic range (28 dB) required the same amount of spectral contrast to achieve maximum performance as subject S5 who only had a 10 dB dynamic range. This outcome suggests that the amount of spectral contrast needed for vowel identification is independent of the dynamic range, and therefore may be dependent on other factors. Experiment 2 investigates the possibility that spectral resolution might be one of the factors affecting the amount of spectral contrast needed to reach asymptotic performance.

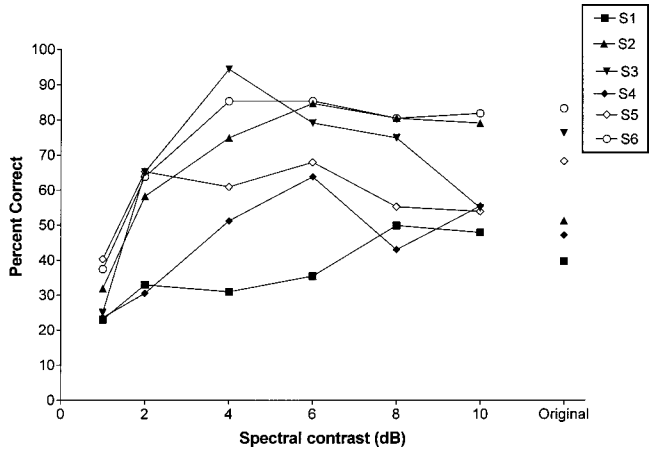


FIG. 4. Individual cochlear implant subject's performance on vowel recognition as a function of spectral contrast.

As shown in Fig. 4, not all subjects reached an asymptote in performance as the peak-to-trough ratio increased. Performance for some subjects reached a peak at 6 dB and then declined slightly thereafter. We expected that the subjects' performance would asymptote at the same level as that obtained using the original (unmodified) vowels. That was not the case, however. In fact, some of the spectrally modified vowels were more easily identified than the original vowels. Figure 6 shows the average scores for each vowel for the original, the 4 dB, and the 6 dB contrast conditions. The majority of the vowels benefited from spectral contrast modification with the largest benefit obtained for the vowels

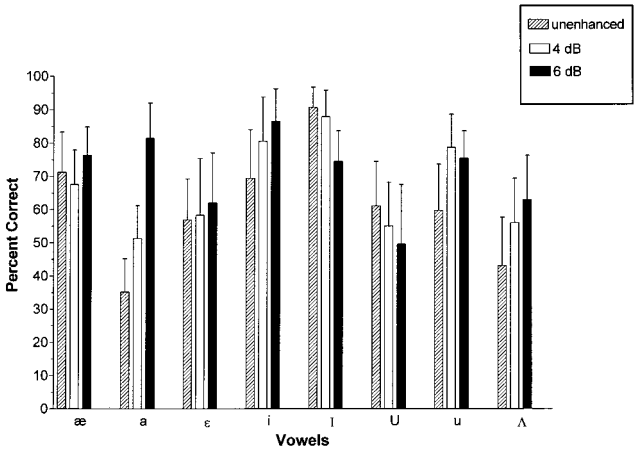


FIG. 6. Mean performance of CI listeners for the unenhanced, the 4 dB, and the 6 dB contrast conditions for each vowel. Error bars indicate standard errors of the mean.

/a i u ʌ/. The fact that the spectrally modified vowels (which had a 4 and 6 dB peak-to-trough ratio) were more easily identified than the original vowels suggests that some vowels had originally smaller spectral contrast. Indeed, we found out that the spectral contrast of some vowels was smaller than 6 dB before enhancement. Figure 7 shows, as an example, the histogram of peak-to-trough ratios of the channel amplitudes of the vowel /a/ processed through subject S2's processor, i.e., computed after bandpass filtering, envelope detection and logarithmic compression. The peak-to-trough ratio of the original (unmodified) vowel /a/ varied from a low of 0.3 dB to a high of 4 dB, with an average of 1.9 dB. It was therefore not surprising that subject S2's performance on identification of the vowel /a/ jumped from 11% correct for the original vowels to 78% correct for the vowels enhanced to 6 dB spectral contrast.

Vowel /a/ (unenhanced) was the most difficult vowel to identify (Fig. 6), consistent with previous findings by Loizou *et al.* (1998) on vowel identification by CI users. Close analysis of the well identified and the poorly identified tokens of ''hod'' in the Loizou *et al.* (1998) study showed that the poorly identified tokens lacked the distinct peak in the channel amplitude spectrum characteristic of the well-identified tokens. The poorly identified tokens of ''hod'' were characterized by a more diffuse distribution of energy across channels 4–6, and had therefore smaller spectral contrast. Increasing the spectral contrast of the vowel /a/ made the peak in the channel amplitude spectrum more distinct and perceptually more salient, leading to a significant improvement in identification. As shown in Fig. 6, not all vowels benefited from spectral contrast enhancement. This is because some vowels have inherently larger spectral contrast than others, with the front vowels having the largest spectral contrast (Fant, 1973). So, no improvements were obtained when the original (unenhanced) vowels had a spectral contrast larger than 4–6 dB.

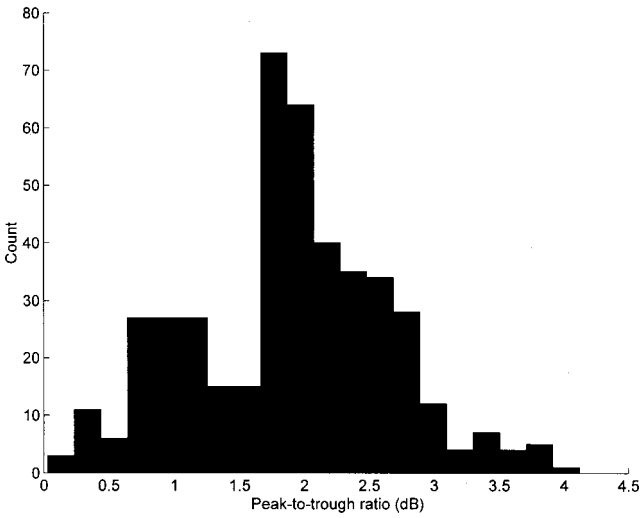


FIG. 7. Histogram of peak-to-trough ratios of the channel amplitudes of the vowel /a/ processed through subject S2's processor.

Subject S2 was not the only subject that benefited in vowel recognition from spectral contrast enhancement. As shown in Fig. 4, subjects S1, S3, S4 also benefited. Subject S3's scores improved from 76% correct using the original vowels to 94% using vowels modified to have a 4 dB spectral contrast. Subject S4's scores improved from 47% correct using unenhanced vowels to 64% using vowels enhanced to 6 dB contrast. These results are encouraging as they suggest that postprocessing the channel amplitudes (estimated using the CIS strategy) through a spectral-contrast enhancement algorithm can improve the vowel recognition performance of some CI listeners.

In addition to the improvement in vowel identification, enhancing the spectral contrast may also potentially improve consonant identification. Dorman and Loizou (1996) showed that the identification of the consonants /p t k/, which were responsible for the majority of the consonant confusion errors, can be improved by enhancing the peak of the consonant spectra at the onset. To improve the identification of /ka/ for example, Dorman and Loizou (1996) low-pass filtered the consonant using a cutoff frequency just below the frequency of channel 5. The low-pass filtering reduced the energy in channels 5 and 6, thereby emphasizing the ''midfrequency'' peak characteristic of velars. Low-pass filtering improved the spectral contrast of /k/ and consequently improved recognition, much like the spectral contrast algorithm in this study improved the contrast of the vowel /a/ and consequently improved recognition.

The results of this experiment not only tell us about the minimum spectral contrast needed for vowel identification by CI listeners, but they also tell us about the absolute minimum dynamic range needed for vowel identification. For subjects fitted with 6-channel cochlear implant processors, a minimum of 6-dB dynamic range is needed for vowel identification. And this is a very conservative estimate, because it does not account for the compression of the acoustic amplitudes to electric amplitudes. The (logarithmic) compression maps the input signal to a small portion of the output dynamic range, and it rarely, if ever, covers the whole dynamic range. It is possible, as shown in Fig. 4 in Loizou *et al.* (2000), for a signal to be mapped to a 24 dB dynamic range, and have less than 10 dB of spectral contrast. Having there-

fore a dynamic range larger than 6 dB increases the probability that the resulting spectral contrast will be at least 6 dB.

III. EXPERIMENT 2: MINIMUM VOWEL SPECTRAL CONTRAST NEEDED BY NORMAL-HEARING LISTENERS

In experiment 1 we found that most cochlear implant listeners who were fitted with a 6-channel processor needed at least a 4–6 dB peak-to-trough ratio for accurate vowel recognition. In this experiment, we investigate whether this outcome holds when speech is processed through a larger (or smaller) number of channels. We hypothesize that there is a trade-off between spectral resolution (number of spectral channels) available and spectral contrast needed. This hypothesis was partially motivated by the finding that one of our CI users (S1), who was fitted with a 5-channel CIS processor, needed a larger spectral contrast for vowel identification compared to the other CI users (see Fig. 4).

To produce speech with varying degrees of spectral resolution, speech was filtered through 4–12 frequency bands, and synthesized as a linear combination of sine waves with amplitudes extracted from the envelopes of the bandpassed wave forms, and frequencies equal to the center frequencies of the bandpass filters. The spectral contrast algorithm presented in experiment 1 was applied to the sine wave amplitudes to produce vowels with varying degrees of spectral contrast, ranging from 1 to 20 dB. The intelligibility of vowels was assessed as a function of spectral resolution and as a function of spectral contrast, using normal-hearing listeners as subjects.

A. Method

1. Subjects

Nine graduate students from the University of Arkansas at Little Rock[1] served as subjects. All of the subjects were native speakers of American English and had normal hearing. The subjects were paid for their participation.

2. Speech material

The same vowel stimuli used in experiment 1 were used.

3. Signal processing

Signals were first processed through a pre-emphasis filter (2000 Hz cutoff), with a 3 dB/octave roll-off, and then bandpassed into n frequency bands ($n=4,6,8,12$) using sixth-order Butterworth filters. Logarithmic filter spacing was used for $n<8$ and mel spacing was used for $n\geqslant 8$. The center frequencies and the 3 dB bandwidths of the filters can be found in Loizou *et al.* (1999). The envelopes of the signal were extracted by full-wave rectification, and low-pass filtering (second-order Butterworth) with a 400 Hz cutoff frequency. The envelope amplitudes were estimated by computing the rms energy of the envelopes every 4 ms. The spectral contrast algorithm presented in experiment 1 was used to modify the peak-to-trough ratio of the estimated envelope amplitudes to Q dB ($Q=1,2,4,6,8,10,15,20$). Sine waves were generated with amplitudes equal to the spectrally enhanced envelope amplitudes, and frequencies equal to the center frequencies of the bandpass filters. The phases of the sinusoids were estimated from the Fast Fourier Transform of the speech segment (Loizou *et al.*, 1999). The sinusoids of each band were finally summed and the level of the synthesized speech segment was adjusted to have the same rms value as the original speech segment.

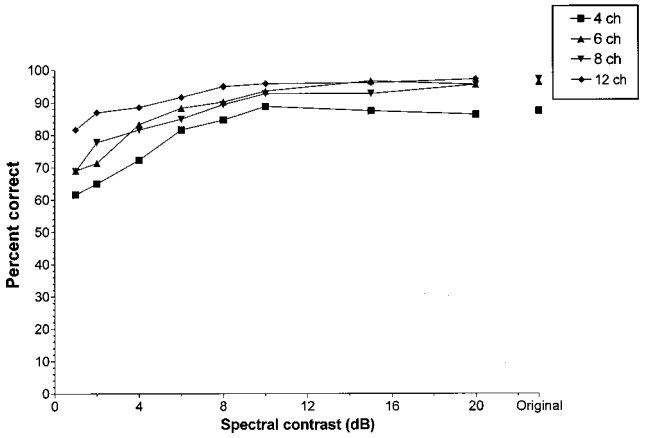


FIG. 8. Mean performance of normal-hearing listeners on vowel recognition as a function of spectral contrast and number of channels.

In addition to the spectrally enhanced vowels, we also processed vowels as described previously but without enhancing the envelope amplitudes. We used this condition for comparative reasons and refer to it as the ''unenhanced'' condition.

4. Procedure

The experiment was performed on a PC equipped with a Creative Labs SoundBlaster 16 soundcard. The subjects listened to the speech material via closed ear-cushion headphones at a comfortable level set by the subject. A graphical interface was used that allowed the subjects to select the vowel they heard using a mouse.

Before each condition, subjects were given a practice session with examples of vowels processed through the same number of channels and the same peak-to-trough ratio in that condition. A sequential test order, starting with speech material processed through a large number of channels ($n=12$) and continuing to speech material processed through a small number of channels ($n=4$), was employed. We chose this sequential test design to give the subjects time to adapt to listening to altered speech signals. The test order for the different peak-to-trough ratios in each channel condition was counterbalanced between subjects.

B. Results and discussion

The results, scored in percent correct, are shown in Fig. 8. A two-factor (channels and peak-to-trough ratio) repeated measures analysis of variance (ANOVA) showed a significant main effect of number of channels [$F(3,24)=8.73$, $p<0.0005$], a significant effect of peak-to-trough ratio [$F(8,64)=67.37$, $p<0.0005$], and a significant interaction between number of channels and peak-to-trough ratio [$F(24,192)=3.73$, $p<0.0005$].

For vowels processed through 4 channels, normal-hearing listeners needed at least a 6 dB peak-to-trough ratio

to identify vowels with greater than 80% accuracy. Post-hoc analysis, according to Tukey, showed that the vowel scores obtained at 6 dB were not significantly different ($p=0.9$) from the scores obtained at 20 dB. The scores obtained at 10 dB were not significantly different ($p=1.0$) from the scores obtained at 20 dB. For vowels processed through 6 or 8 channels, normal-hearing listeners needed a 4 dB peak-to-trough ratio to identify vowels with the same accuracy. This is consistent with our findings in experiment 1 with cochlear implant users fitted with 6-channel processors. The scores obtained with 4 dB contrast using 6 or 8 channels were not significantly different ($p>0.5$, Tukey post-hoc) from the scores obtained at 20 dB. Finally, for vowels processed through 12 channels, normal-hearing listeners needed only a 1 dB peak-to-trough ratio to identify vowels with greater than 80% accuracy. Post-hoc analysis (Tukey) showed that the score obtained at 2 dB was only marginally different ($p=0.044$) from the score obtained at 20 dB.

The above-given results obtained with 4 channels confirm our original hypothesis that when the spectral resolution is poor, a comparatively larger spectral contrast is needed for vowel identification. A larger spectral contrast is needed, because we suspect that listeners must be using amplitude differences across channels to infer the frequency content (e.g., formant locations, etc.) of the signal when the spectral resolution is poor. Conversely, when the spectral resolution is fine (12 channels), a small spectral contrast (1 dB) is sufficient. The results in experiment 1 with CI patients fitted with 6-channel processors showed that a 4–6 dB amplitude difference between the peak and the valley needs to be maintained for accurate vowel recognition. Consistent with the above-mentioned hypothesis and the findings of experiment 2, subject S1, who was fitted with a 5-channel processor, needed a larger spectral contrast (8 dB) to achieve maximum performance on vowel recognition. Judging from the subject's low scores on open set recognition (Table I), it seems likely that subject S1 may be receiving a small number (probably less than 5) of independent channels of stimulation. The results of experiment 2 suggest that if we could somehow provide at least 12 channels of stimulation to CI listeners, then a small spectral contrast (1–2 dB), and consequently, a small dynamic range (at least 2 dB) would be sufficient for vowel recognition.

The results obtained with 12 channels are consistent with those reported in the literature (Turner and Van Tassel, 1984; Leek *et al.*, 1987; Summerfield *et al.*, 1987; Alcantara and Moore, 1995) that only a 1–2 dB spectral contrast is needed to identify vowel-like harmonic complexes with 70%–75% correct accuracy. Note that the subjects in the Alcantara and Moore (1995) study needed a 3 dB contrast to achieve 75% correct accuracy (six vowel-like harmonic complexes were used in their study, whereas Leek *et al.* used four vowel-like harmonic complexes). Our study showed that high vowel recognition performance (>80% correct) can be achieved even with 1 dB spectral contrast. This vowel identification threshold is the same as the psychophysical threshold needed to detect a change in the amplitude spectrum of a complex signal. Green *et al.* (1983) showed, for instance, that normal-hearing listeners can detect 1 dB increments added to one component of a complex signal.

The mean scores obtained in this study with 1 dB contrast were considerably higher than any of the scores reported in the literature on a similar experiment. For a 1 dB contrast, the subjects of Leek *et al.* (1987) achieved 55% accuracy, the subjects of Alcantara and Moore (1995) achieved 35% accuracy, while our subjects achieved 82% accuracy. Higher performance was achieved in this study even though we represented the vowel spectra with 12 frequency components as opposed to 30 harmonics in the Leek *et al.* study, and used a larger number of vowels (8 vowels in our study versus 4 vowels in the Leek *et al.* study and 6 vowels in the Alcantara and Moore study). We believe that the higher performance in vowel recognition was obtained in our study because we used natural vowels. Our vowel stimuli contained most of the spectral cues present in naturally produced vowels, including *F*0 variation and formant movements. In addition, the listeners had access to duration cues. We do not believe that the high performance obtained with our stimuli was primarily because of duration cues, because a recent study by Hillenbrand *et al.* (2000) with normal-hearing listeners showed that the vowel duration had a small overall effect on vowel identification.

Several studies have shown that hearing-impaired listeners need a larger spectral contrast compared to normal-hearing listeners to achieve high vowel recognition performance (e.g., Leek *et al.*, 1987). This was attributed to the wider-than-normal auditory filters. The situation with CI listeners, however, is quite different, since the auditory filters are bypassed with electrical stimulation. The results from experiment 2 suggest that cochlear implant listeners need a larger spectral contrast than normal-hearing listeners not because of the limited dynamic range, but because of the reduced spectral resolution.

IV. CONCLUSIONS

(1) Cochlear implant listeners fitted with 6-channel CIS processors need at least a 4 dB spectral contrast to identify natural vowels with high accuracy. Most subjects achieved the highest performance on vowel recognition with a 6 dB spectral contrast, while one subject needed 8 dB.

(2) Increasing the vowel spectral contrast to 6 dB benefited most subjects in vowel recognition. Some subjects' vowel scores improved by about 20 percentage points when the vowels were enhanced to 6 dB. These results are encouraging as they suggest that we can improve vowel recognition for CI users, simply by postprocessing the CIS channel amplitudes through a spectral contrast enhancement algorithm. The proposed spectral contrast enhancement algorithm used in this study is relatively easy to implement and is amenable for real-time implementation.

(3) The results of experiment 2 with normal-hearing listeners indicated that the minimum spectral contrast needed for vowel identification was dependent on the spectral resolution, i.e., the number of channels of frequency information available. For vowels processed through 4 channels, normal-hearing listeners needed at least a 6 dB peak-to-trough ratio to identify vowels with greater than 80% accuracy, while for vowels processed through 6 or 8 channels, normal-hearing

listeners needed a 4 dB peak-to-trough ratio to identify vowels with the same accuracy, consistent with our findings with CI users. For vowels processed through 12 channels, normal-hearing listeners needed only a 1 dB peak-to-trough ratio to identify vowels with greater than 80% accuracy.

(4) The above-mentioned findings with normal-hearing listeners are consistent with our hypothesis that when the spectral resolution is poor, a larger spectral contrast is needed for vowel identification. Conversely, when the spectral resolution is fine, a small spectral contrast (1 dB) is sufficient.

(5) For vowels processed through 12 channels, a 1 dB contrast was sufficient to reach high performance (>80% correct) on vowel recognition. The high scores achieved with 1 dB contrast were significantly higher than the scores reported in the literature (55% correct in the Leek *et al.* study and 33% correct in the Alcantura and Moore study). The high performance obtained in our study can be attributed to the fact that we used naturally produced vowels.

(6) The outcomes of experiments 1 and 2, taken together suggest that CI listeners need a larger spectral contrast (4–6 dB) than normal-hearing listeners to achieve high recognition accuracy, not because of the limited dynamic range, but because of the limited spectral resolution.

ACKNOWLEDGMENT

This research was supported by Grant No. R01 DC03421 from the National Institute of Deafness and other Communication Disorders, NIH.

[1]The authors were previously affiliated with the University of Arkansas at Little Rock before joining the University of Texas at Dallas.

Alcantara, J., and Moore, B. (**1995**). ''The identification of vowel-like harmonic complexes: Effects of component phase, level, and fundamental frequency,'' J. Acoust. Soc. Am. **97**, 3813–3824.

Dorman, M., and Loizou, P. (**1996**). ''Improving consonant intelligibility for Ineraid patients fit with Continuous Interleaved Sampling (CIS) processors by enhancing contrast among channel outputs,'' Ear Hear. **17**, 308–313.

Fant, G. (**1973**). *Speech Sounds and Features* (MIT, Cambridge, MA).

Green, D., Kidd, G., and Picardi, M. (**1983**). ''Successive versus simultaneous comparison in auditory intensity discrimination,'' J. Acoust. Soc. Am. **73**, 639–643.

Hawks, J., Fourakis, M., Skinner, M., Holden, T., and Holden, L. (**1997**). ''Effects of formant bandwidth on the identification of synthetic vowels by cochlear implant recipients,'' Ear Hear. **18**, 479–487.

Hillenbrand, J., Clark, M., and Houde, R. (**2000**). ''Some effects of duration on vowel recognition,'' J. Acoust. Soc. Am. **108**, 3013–3022.

Hillenbrand, J., Getty, L., Clark, M., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.

Leek, M., Dorman, M., and Summerfield, Q. (**1987**). ''Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **81**, 148–154.

Leek, M., and Summers, V. (**1996**). ''Reduced frequency selectivity and the preservation of spectral contrast in noise,'' J. Acoust. Soc. Am. **100**, 1796–1806.

Loizou, P. (**1998**). ''Mimicking the human ear: An overview of signal processing techniques for converting sound to electrical signals in cochlear implants,'' IEEE Signal Process. Mag. **15**, 101–130.

Loizou, P., Dorman, M., and Fitzke, J. (**2000**). ''The effect of reduced dynamic range on speech understanding: Implications for patients with cochlear implants,'' Ear Hear. **21**, 25–31.

Loizou, P., Dorman, M., and Powell, V. (**1998**). ''The recognition of vowels produced by men, women, boys and girls by cochlear implant patients using a six-channel CIS processor,'' J. Acoust. Soc. Am. **103**, 1141–1149.

Loizou, P., Dorman, M., and Tu, Z. (**1999**). ''On the number of channels needed to understand speech,'' J. Acoust. Soc. Am. **106**, 2097–2103.

Pick, G., Evans, E., and Wilson, J. (**1977**). ''Frequency resolution of patients with hearing loss of cochlear origin,'' in *Psychophysics and Physiology of Hearing*, edited by E. Evans and J. Wilson (Academic, London).

Poroy, O., and Loizou, P. (**2000**). ''Development of a speech processor for laboratory experiments with cochlear implant patients,'' IEEE Int. Conference Acoust. Speech Signal Process. **6**, 3626–3629.

Summerfield, Q., Sidwell, A., and Nelson, T. (**1987**). ''Auditory enhancement of changes in spectral amplitude,'' J. Acoust. Soc. Am. **81**, 700–708.

Turner, C., and Holte, L. (**1987**). ''Discrimination of spectral-peak amplitude by normal and hearing-impaired subjects,'' J. Acoust. Soc. Am. **81**, 445–451.

Turner, C., and Van Tassell, D. (**1984**). ''Sensorineural hearing loss and the discrimination of vowel-like stimuli,'' J. Acoust. Soc. Am. **75**, 562–566.

Wightman, F., McGee, T., and Kramer, M. (**1977**). ''Factors influencing frequency selectivity in normal and hearing-impaired listeners,'' in *Psychophysics and Physiology of Hearing*, edited by E. Evans and J. Wilson (Academic, London).

Zeng, F-G., and Galvin, J. (**1999**). ''Amplitude mapping and phoneme recognition in cochlear implant listeners,'' Ear Hear. **20**, 60–74.

On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception[a)]

Oded Ghitza
Media Signal Processing Research, Agere Systems, Murray Hill, New Jersey 07974

(Received 8 August 2000; revised 20 February 2001; accepted 7 June 2001)

Studies in neurophysiology and in psychophysics provide evidence for the existence of temporal integration mechanisms in the auditory system. These auditory mechanisms may be viewed as ''detectors,'' parametrized by their cutoff frequencies. There is an interest in quantifying those cutoff frequencies by direct psychophysical measurement, in particular for tasks that are related to speech perception. In this study, the inherent difficulties in synthesizing speech signals with prescribed temporal envelope bandwidth *at the output of the listener's cochlea* have been identified. In order to circumvent these difficulties, a dichotic synthesis technique is suggested with interleaving critical-band envelopes. This technique is capable of producing signals which generate cochlear temporal envelopes with prescribed bandwidth. Moreover, for unsmoothed envelopes, the synthetic signal is perceptually indistinguishable from the original. With this technique established, psychophysical experiments have been conducted to quantify the upper cutoff frequency of the auditory critical-band envelope detectors at threshold, using high-quality, wideband speech signals (bandwidth of 7 kHz) as test stimuli. These experiments show that in order to preserve speech quality (i.e., for inaudible distortions), the minimum bandwidth of the envelope information for a given auditory channel is considerably smaller than a critical-band bandwidth (roughly one-half of one critical band). Difficulties encountered in using the dichotic synthesis technique to measure the cutoff frequencies relevant to intelligibility of speech signals with fair quality levels (e.g., above MOS level 3) are also discussed.
[DOI: 10.1121/1.1396325]

PACS numbers: 43.71.Pc, 43.66.Ba, 43.72.Ar [DOS]

I. INTRODUCTION

Studies in neurophysiology and in psychophysics provide evidence for the existence of temporal integration mechanisms in the auditory system (e.g., Eddins and Green, 1995). The neural circuitry that realizes these mechanisms is yet to be understood. At the least, we may view these mechanisms as ''detectors,'' characterized in part by their lower- and upper cutoff frequencies. These cutoff frequencies determine which part of the input information that is present at the auditory-nerve (AN) level is perceptually relevant. Hence, it is important to quantify these frequencies, particularly for tasks that are related to speech perception.

Two recent studies (Drullman *et al.*, 1994 and Chi *et al.*, 1999) seem to provide psychophysically based estimates of the cutoff frequencies of the auditory detectors involved in tasks related to speech intelligibility. These studies are inspired by the apparent ability of the speech transmission index (STI) to predict intelligibility scores for speech recorded in auditorium-like conditions (e.g., Steeneken and Houtgast, 1980). Recall that the STI is computed from the modulation transfer functions (MTFs) of the transmission path between the location of the speech source and that of the microphone. An MTF is specified at a given frequency as the degree to which the original intensity modulations are preserved at the microphone location. In Steeneken and Houtgast, 1980, the MTFs are measured for 7 one-octave-wide noise carriers centered at frequencies that are one octave apart (from 125 to 8000 Hz), with 14 modulation frequencies (0.63 to 12.5 Hz, in one-third-octave steps). [Note that the range of center frequencies covers the frequency range used in speech communication, and that the range of the modulation frequencies covers the time constants of the articulatory mechanisms used by the human speaker.] The high correlation of STI and speech intelligibility scores (Steeneken and Houtgast, 1980), and the fact that STI is based upon MTFs, raises the question whether auditory detectors active in the speech intelligibility task have a cutoff frequency of the order of 12.5 Hz (i.e., the maximum modulation frequency in Steeneken and Houtgast, 1980). In Drullman *et al.* (1994), an attempt was made to assess the amount by which temporal modulations can be reduced without affecting the performance in a phoneme identification task. Results showed that temporal envelope smoothing hardly affect the performance, even for cutoff frequency as low as 16 Hz. In Chi *et al.* (1999), detection thresholds were measured for spectral and temporal MTFs using broadband stimuli with sinusoidally rippled profiles that vary with time. Results showed that temporal MTFs exhibit low-pass characteristics, with cutoff frequencies similar to those of Drullman *et al.* (1994).

A question that emerges at this point is whether the psychophysical data obtained by these experiments, about the bandwidth of temporal MTFs, can also be considered as evidence of the characteristics of the relevant auditory mechanisms (i.e., that they are low-pass in nature, with cutoff frequencies of about 16 Hz). As shown in Sec. II, such a

[a)]This work was done while the author was with Bell Labs, Lucent Technologies.

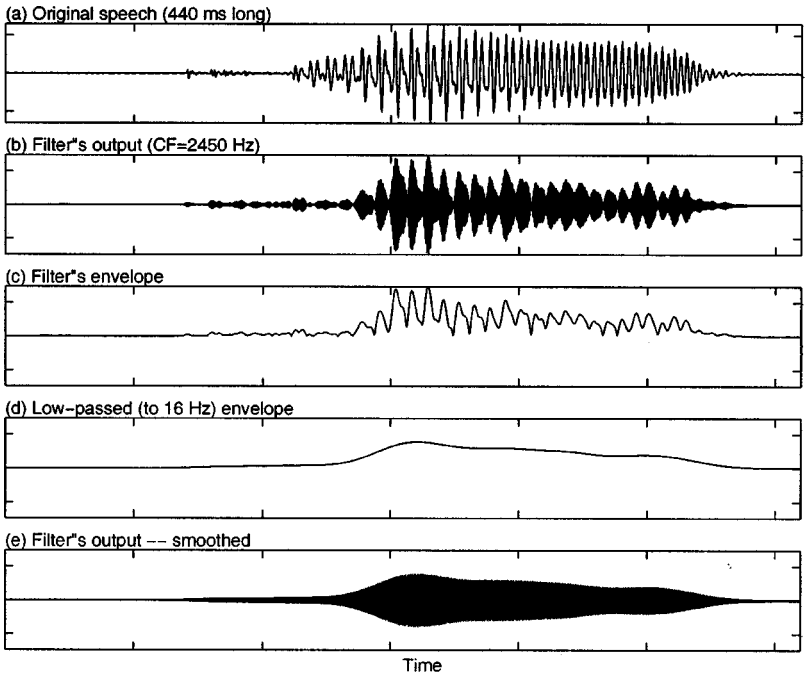


FIG. 1. From top to bottom: (a) a 440-ms-long segment of the original speech $s(t)$; (b) the output signal, $s_i(t)$, of a critical-band filter centered at 2450 Hz; (c) the envelope $a_i(t)$; (d) the smoothed envelope $\tilde{a}_i(t)$ (low-pass filtered to $B=16$ Hz); and (e) the envelope-smoothed critical-band signal $\overline{s_i}(t)$. The ordinate of panels (b) to (e) have the same scale. The ordinate of panel (a) has a different scale.

conclusion is not permissible. This is so because the observed psychophysical performance is, in part, a consequence of using signal-processing techniques which, for a prescribed envelope bandwidth, produce synthetic signals that generate internal auditory representations whose temporal envelopes are wideband signals, with envelope bandwidths as wide as one critical band. Therefore, while performing the psychophysical experiments the human observer was presented with rich temporal envelope information, with a bandwidth much beyond the nominal value prescribed at the input.

In Sec. III, the difficulties inherent in synthesizing speech signals with prescribed temporal envelope bandwidth at the output of the listener's cochlea are identified. In order to circumvent these difficulties, a dichotic[1] synthesis has been suggested with interleaving smoothed critical-band envelopes. This technique has two desired capabilities: (1) it produces synthetic signals which generate cochlear temporal envelopes with prescribed bandwidth, and (2) for unsmoothed envelopes, the synthetic signal is perceptually indistinguishable from the original. With this technique established, psychophysical experiments have been conducted to quantify the upper cutoff frequency of the auditory critical-band envelope detectors at threshold (i.e., in the context of preserving speech quality) using high-quality, wideband speech signals (bandwidth of 7 kHz) as test stimuli (Sec. IV). Finally, in Sec. V, the difficulties encountered in using the dichotic synthesis technique to measure the cutoff frequencies relevant to intelligibility of speech signals with some reasonable level of quality (say, "fair"—or 3—on the MOS scale[2]) are also discussed.

II. TEMPORAL SMOOTHING AND SPEECH INTELLIGIBILITY

It is widely accepted that a decomposition of the output of a cochlear filter into a temporal envelope and a "carrier" may be used to quantify the role of auditory mechanisms in speech perception (e.g., Flanagan, 1980). This is supported by our current understanding of the way the auditory system (the periphery, in particular) operates.

Let $s(t)$ be the original speech signal, and let $s_i(t)$ be a bandlimited signal resulting from filtering $s(t)$ through $h_i(t)$

$$s_i(t)=s(t)*h_i(t). \tag{1}$$

Here, $h_i(t)$ is the impulse response of the ith critical-band filter and the operator $*$ represents convolution. We can express $s_i(t)$ of Eq. (1) as

$$s_i(t)=a_i(t)\cos\phi_i(t), \tag{2}$$

where $a_i(t)$ is the *Hilbert envelope*[3] of $s_i(t)$, $\phi_i(t)$ is the *Hilbert instantaneous phase*[3] of $s_i(t)$, and $\cos\phi_i(t)$ is the *carrier* of $s_i(t)$. We refer to the expression of Eq. (2) as "the envelope/carrier decomposition" of $s_i(t)$.

Let $\tilde{a}_i(t)$ be a filtered version of $a_i(t)$, low-passed to some cutoff frequency B. The *envelope-smoothed* critical-band signal is defined as

$$\overline{s_i}(t)=\tilde{a}_i(t)\cos\phi_i(t), \tag{3}$$

and the envelope-smoothed speech signal is defined as

$$\overline{s}(t)=\sum_{i=1}^{N}\overline{s_i}(t)=\sum_{i=1}^{N}\tilde{a}_i(t)\cos\phi_i(t), \tag{4}$$

where N is the number of critical bands.

Figure 1 shows (from top to bottom) (a) a 440-ms-long segment of the original speech $s(t)$; (b) the output signal, $s_i(t)$, of a critical-band filter centered at 2450 Hz; (c) the envelope $a_i(t)$; (d) the smoothed envelope $\tilde{a}_i(t)$, low-pass filtered to $B=16$ Hz, and (e) the envelope-smoothed critical-band signal $\overline{s_i}(t)$.

In Drullman *et al.* (1994), the envelope-smoothed speech of Eq. (4) was used to measure human performance in a phoneme identification task as a function of the cutoff

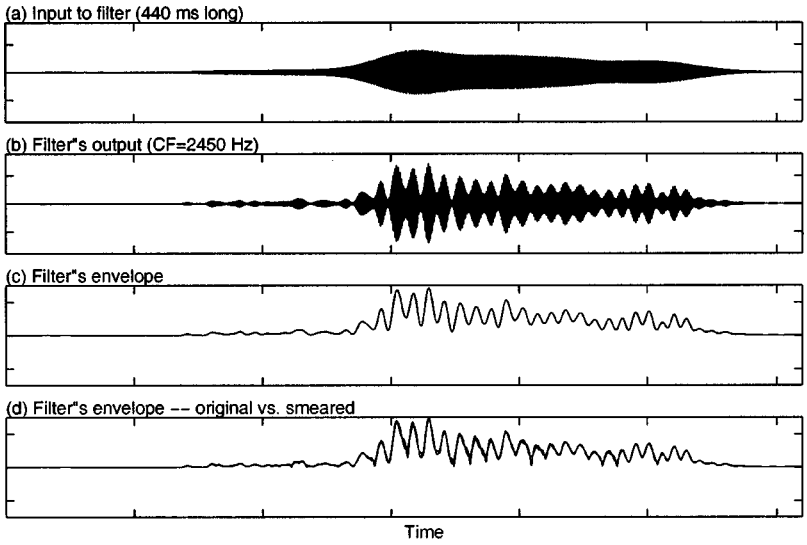


FIG. 2. From top to bottom: (a) Fig. 1(e), redrawn; (b) the output signal of a critical-band filter centered at 2450 Hz, for the input signal shown in (a); (c) the envelope signal of the critical-band signal of (b); and (d) comparison of the envelope signals of Figs. 1(c) and 2(c). Ordinate of all panels have the same scale.

frequency B of a low-pass filter representing the temporal smoothing. Results showed that performance was hardly affected by temporal envelope smoothing characterized by cutoff frequencies higher than 16 Hz.

A question that emerges at this point is whether these findings can be considered as evidence that relevant auditory mechanisms are low-pass in nature, with cutoff frequency of about 16 Hz. This question stems from our current understanding of the relationship between the envelope $a_i(t)$ of the driving signal and the properties of the auditory-nerve firing patterns they stimulate. This understanding is better, in particular, for AN fibers with high characteristic frequencies (CFs),[4] where the synchrony of neural discharges to frequencies near the CF is greatly reduced, due to the physiological limitations of the inner hair cell (IHC) in following the carrier information. At these frequencies, temporal information is preserved by the instantaneous average rate of the neural firings, which is related to the temporal envelope of the underlying driving cochlear signal.[5] Is it correct to assume that, by presenting the listener with the envelope-smoothed signal $\tilde{a}_i(t)\cos\phi_i(t)$, the instantaneous average rate of the corresponding stimulated AN fibers is also smoothed, limiting the bandwidth of the information available to the upper auditory stages to B?

A. The role of interaction between temporal envelope and phase

Such a conclusion would be justified if the processing of the speech signal would result in the signal of Fig. 1(e) *at the output of the listener's cochlear filter*. This, however, is not the case as illustrated in Fig. 2. Figure 2(b) shows the output signal of a critical-band filter, identical to the one used in Fig. 1, for the input signal shown in Fig. 1(e). [For pictorial clarity, Fig. 1(e) is redrawn as Fig. 2(a).] Figure 2(c) shows its envelope. Clearly, these signals [of Figs. 2(b) and (c)] do not look at all like the smooth signals of Figs. 1(e) and (d), respectively. Indeed, they look very much like the original (nonsmoothed) signals of Figs. 1(b) and (c), respectively. [To highlight this point, a comparison of the envelope signals, Fig. 1(c) and Fig. 2(c), is shown in Fig. 2(d).] The implication of this finding is that the envelope-smoothed speech signal $\overline{s(t)}$ of Eq. (4) is inappropriate for the purpose of measuring the cutoff frequency of the auditory envelope detector. This is so because, when listening to $\overline{s(t)}$, the human observer is presented with rich envelope information, much beyond the nominal cutoff frequency of the smoothing filter.

The fact that filtering the smooth signal restores much of the nonsmoothed envelope appears to be somewhat unexpected. However, two theorems, one in the field of signal processing and one in the field of communications, provide analytic support to this finding. These theorems determine that: (1) For a bandlimited signal $s_i(t)=a_i(t)\cos\phi_i(t)$, the envelope signal $a_i(t)$ and the phase signal $\phi_i(t)$ are related (e.g., Voelcker, 1966), and (2) If $\phi(t)$ is a bandlimited signal, and if $\cos\phi(t)$ is the input to a bandpass filter [note that the envelope of the input signal is a constant, i.e., $a_i(t)=1$], then the filter's output has an envelope that is related to $\phi(t)$ (e.g., Rice, 1973). A corollary to these theorems is that if we pass the envelope-smoothed signal $\overline{s_i(t)}=\tilde{a}_i(t)\cos\phi_i(t)$ through a bandpass filter, the bandwidth of the output envelope is larger than the bandwidth of $\tilde{a}_i(t)$ [where the extra information is regenerated from $\phi_i(t)$]. If the bandpass filter represents a cochlear filter, the bandwidth of the temporal envelope information available to the listener is greater than the nominal smoothing cutoff frequency, B!

One clarification is noteworthy. The envelope signal of Fig. 2(c) (representing the envelope at the listener's cochlear output) exhibits both pitch modulations and articulatory modulations. Recall that the articulatory modulations (the main carrier of speech intelligibility) of the input envelope signal were low-pass filtered to B (e.g., 16 Hz). A question arises whether the envelope signal shown in Fig. 2(c) is mainly composed of pitch modulations (i.e., a secondary carrier of speech intelligibility), while the articulatory modulations are bandlimited to B, as intended. To answer this question, recall that the *phase information* of the input signal is unsmoothed, comprising the unsmoothed articulatory modulations and the unsmoothed pitch modulations. It is impossible to use the analytic expressions derived by Rice to isolate the response of the filter to the articulatory modulations from its response to the pitch modulations. (This is so be-

cause of the complexity of these expressions.) Suffice it to say that even though the articulatory information of the input envelope signal was appropriately smoothed (e.g., to 16 Hz), it still exists in its entirety in the input phase signal and, therefore, will be regenerated as part of the envelope signal at the filter's output.

III. DICHOTIC SYNTHESIS WITH INTERLEAVING CHANNELS

For a direct psychophysical measurement of the cutoff frequency of the auditory envelope detector, we have to ensure bandlimited envelope information at the listener's AN. This requirement can be elaborated as follows. Recall that information is conveyed to the AN by a large number of highly overlapped cochlear filters, with a density and location determined by the discrete distribution of the IHCs along the continuous cochlear partition. When the source signal $s(t)$ is passed through this cochlear filter bank, the resulting envelopes change gradually with CF as we move across the filter bank. The signal-processing method we seek should enable us to generate a signal that, when passed through the cochlear filter bank, will result in smoothed envelopes that are the envelopes generated by the source signal $s(t)$, low-pass filtered to the prescribed cutoff frequency B. This requirement, termed "the globally smoothed cochlear envelopes criterion," is formulated in Sec. III A.

In Sec. III B we consider a signal-processing technique based on diotic[6] speech synthesis, using pure cosine carriers. We shall demonstrate that this technique indeed generates smoothed envelopes at the output of the listener's cochlea, but only at the locations that correspond to the frequencies of the cosine carriers. At all other locations, distortions are generated that are perceptually noticeable. In Sec. III C we suggest a signal-processing technique designed to circumvent this problem. The technique is based upon dichotic speech synthesis with interleaving smoothed critical-band envelopes, and is based on the assumption that when the two streams are presented to the left and the right ears, the auditory system produces a single fused image (e.g., Durlach and Colburn, 1978). By using this procedure, perceivable distortions are greatly reduced.

Finally, we note that the present study is limited to measuring the cutoff frequency of the auditory envelope detectors only at the high CF region (i.e., frequencies above 1500 Hz). As mentioned before, ascending information at this frequency range is conveyed mainly via the temporal envelope of the cochlear signals (while the carrier information is lost). The lower frequency range (i.e., below 1500 Hz) was not addressed here since we lack understanding of the post-AN mechanisms that are active at the low CFs (and are sensitive to synchrony).

A. The globally smoothed cochlear envelopes criterion

Let $s(t)$ be processed by a filter bank consisting of the cochlear-shape filters H_1, H_2, and H_x (realized, for example, as gammatone filters, Slaney, 1993), where H_1 and H_2 are one critical band apart, and H_x is located in between H_1 and H_2 (Fig. 3). Let the envelope signals of Fig. 3, $a_1(t)$, $a_2(t)$,

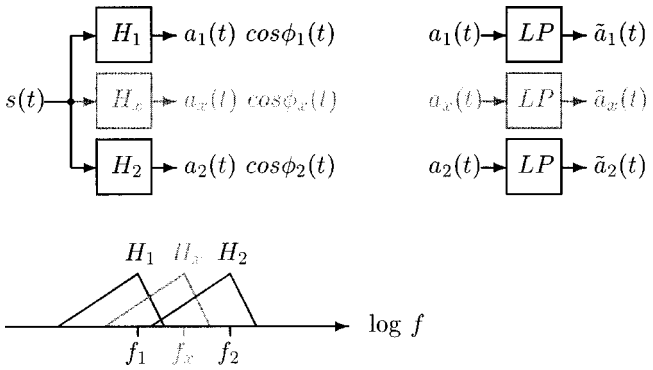


FIG. 3. Passing $s(t)$ through cochlear-shape filters H_1, H_2, and H_x. The spacing between H_1 and H_2 is one critical band. H_x represents one of the many overlapping cochlear filters located in between H_1 and H_2. The envelope signals $a_i(t)$ are temporally smoothed to $\tilde{a}_i(t)$, using a low-pass filter.

and $a_x(t)$, be temporally smoothed to $\tilde{a}_1(t)$, $\tilde{a}_2(t)$, and $\tilde{a}_x(t)$, respectively, and let

$$\tilde{s}(t)=F(\tilde{a}_1(t),\tilde{a}_2(t)), \tag{5}$$

where $F(\cdot,\cdot)$ stands for the desired signal-processing method. Let this $\tilde{s}(t)$ be fed to the filter bank of Fig. 3, as shown in Fig. 4. The resulting output signals, $b_i(t)\cos\psi_i(t)$, $i=1,2,x$, have envelope signals $b_1(t)$, $b_2(t)$, and $b_x(t)$ [and carrier signals $\cos\psi_1(t)$, $\cos\psi_2(t)$, and $\cos\psi_x(t)$]. For filters located at the high-frequency range (say, above 1500 Hz), the desired signal-processing method $F(\cdot,\cdot)$ should be designed to produce $\tilde{s}(t)$ such that

$$b_i(t)=\tilde{a}_i(t), \quad i=1,2,x. \tag{6}$$

Note that the properties of the signal carriers $\cos\psi_i(t)$ are being ignored since, at this frequency range, they are considered irrelevant due to the inability of the inner hair cell to follow the carrier information.

B. Diotic synthesis with pure cosine carriers

Reiterating Eqs. (1) and (2), let

$$s_i(t)=s(t)*h_i(t)=a_i(t)\cos\phi_i(t), \tag{7}$$

where $s(t)$ is the input signal, $h_i(t)$ is the impulse response of a gammatone filter centered at frequency f_i (above 1500 Hz), the operator $*$ represents convolution, and $a_i(t)$ and $\cos\phi_i(t)$ are, respectively, the envelope and the carrier of the filtered signal $s_i(t)$. Motivated by the observation that neural firings of AN fibers originating at this frequency range

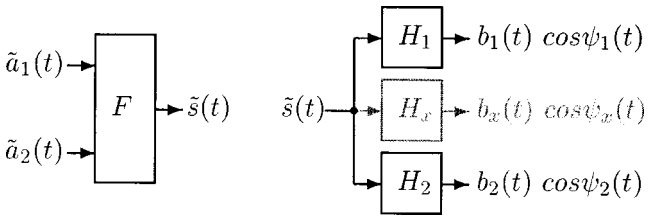


FIG. 4. Passing $\tilde{s}(t)$ through H_1, H_2, and H_x of Fig. 3. The desired signal processing method $F(\cdot,\cdot)$ should be designed to produce $\tilde{s}(t)$, which satisfies Eq. (6).

mainly transmit the envelope information $a_i(t)$, let us consider the signal

$$\hat{s}_i(t) = a_i(t)\cos 2\pi f_i t = a_i(t)\cos \omega_i t, \tag{8}$$

that is, $s_i(t)$, with the original carrier $\cos \phi_i(t)$ of Eq. (7) replaced by a cosine carrier $\cos \omega_i t$. Let $a_i(t)$ be low-pass filtered to $\tilde{a}_i(t)$, and let

$$\tilde{s}_i(t) = \tilde{a}_i(t)\cos \omega_i t. \tag{9}$$

Note that $\tilde{s}_i(t)$ is a bandlimited signal centered at frequency f_i. If $\tilde{s}_i(t)$ is presented to the listener's ear, the resulting envelope signal at the place along the cochlear partition that corresponds to frequency f_i will be the smoothed envelope $\tilde{a}_i(t)$. One possible signal-processing strategy could, therefore, be to generate a signal

$$\tilde{s}(t) = s_{\text{baseband}}(t) + \sum_{i=1}^{N} \tilde{a}_i(t)\cos \omega_i t, \tag{10}$$

where $s_{\text{baseband}}(t)$ represents the low-frequency range (i.e., below 1500 Hz), and $\tilde{a}_i(t)$, $i=1,\ldots,N$ are the smoothed-envelope signals of N gammatone filters equally spaced along the critical-band scale, with a spacing of one critical band, above 1500 Hz.

Let $\tilde{s}(t)$ of Eq. (10) be presented diotically to the listener's ear. The envelope at the output of the listener's cochlear filter located at frequency f_i is (ideally) $\tilde{a}_i(t)$, for each i, $i=1,\ldots,N$. However, the output of a cochlear filter located in between two successive cosine carrier frequencies f_i and f_{i+1} will reflect "beating" of the two modulated cosine carrier signals passing through the filter. This will result in a perceptually noticeable distortion. [Using the terminology of Sec. III A, if $F(\cdot,\cdot)$ is the diotic synthesis technique, i.e., $\tilde{s}(t) = \tilde{a}_1(t)\cos \omega_1 t + \tilde{a}_2(t)\cos \omega_2 t$, then $b_1(t) \cong \tilde{a}_1(t)$ and $b_2(t) \cong \tilde{a}_2(t)$. However, $b_x(t) \neq \tilde{a}_x(t)$, and such will be the case (to a different degree of dissimilarity) for every filter H_x located in between filters H_1 and H_2.]

C. Dichotic synthesis with interleaving critical-band envelopes

1. Principle

To reduce the amount of distortion due to beating, a dichotic synthesis with interleaving critical-band envelopes is proposed. As we shall see, this synthesis procedure is not perfect [i.e., it produces synthetic speech which does not satisfy Eq. (6) in a perfect way]. However, it allows us to circumvent the difficulties encountered in the diotic synthesis procedure and significantly reduce distortions.

Let $\tilde{s}_{\text{odd}}(t)$ and $\tilde{s}_{\text{even}}(t)$ be the summation of the odd components and even components of $\tilde{s}(t)$ of Eq. (10), respectively, i.e.,

$$\tilde{s}_{\text{odd}}(t) = s_{\text{baseband}}(t) + \sum_{i \in \text{odd}} \tilde{a}_i(t)\cos \omega_i t, \tag{11}$$

$$\tilde{s}_{\text{even}}(t) = s_{\text{baseband}}(t) + \sum_{i \in \text{even}} \tilde{a}_i(t)\cos \omega_i t. \tag{12}$$

The distance between two successive cosine carriers in each of these signals is two critical bands, resulting in a reduction of distortion due to carrier beating. When $\tilde{s}_{\text{odd}}(t)$ and $\tilde{s}_{\text{even}}(t)$ are presented to the left and the right ears, respectively, the auditory system produces a single fused image. In Secs. III D and III E, we shall examine the extent to which the fused auditory image achieves the property of Eq. (6).

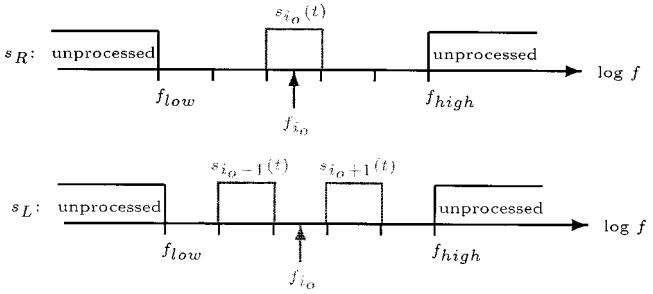


FIG. 5. Dichotic synthesis with interleaving channels. For pictorial clarity, the critical-band spectra are sketched as "flat" spectra.

2. Stimuli for the psychophysical experiments

Let us assume that, for a given input signal $s(t)$, we want to generate a fused auditory image with a range of smoothed-envelope representations that are one critical-band wide and that are centered at frequency f_{i_o}. To achieve this goal, we generate two signals, $\tilde{s}_R(t)$ and $\tilde{s}_L(t)$, as sketched in Fig. 5. More specifically, let the original signal $s(t)$ be divided into three regions: (1) the "low-frequency range," up to frequency f_{low}, denoted as $s_{\text{low}}(t)$; (2) the "high-frequency range," from frequency f_{high}, denoted as $s_{\text{high}}(t)$; and (3) the "middle-frequency range," five successive critical bands wide, located in between frequencies f_{low} and f_{high} and centered at the "target" frequency f_{i_o}. The critical-band signals are $s_i(t) = s(t) * h_i(t) = a_i(t)\cos \phi_i(t)$, where $h_i(t)$ is a gammatone filter centered at frequency f_i, $i = i_o - 2, i_o - 1, i_o, i_o + 1, i_o + 2$. Note that in Figs. 5 and 6 these critical-band spectra are sketched as "flat" spectra, for pictorial clarity.

We define $s_R(t)$ and $s_L(t)$ as

$$s_R(t) = s_{\text{low}}(t) + s_{i_o}(t) + s_{\text{high}}(t), \tag{13}$$

$$s_L(t) = s_{\text{low}}(t) + s_{i_o-1}(t) + s_{i_o+1}(t) + s_{\text{high}}(t). \tag{14}$$

Thus, $s_R(t)$ and $s_L(t)$ are obtained by adding the unprocessed outputs of the filters as illustrated in Fig. 5. Similarly,

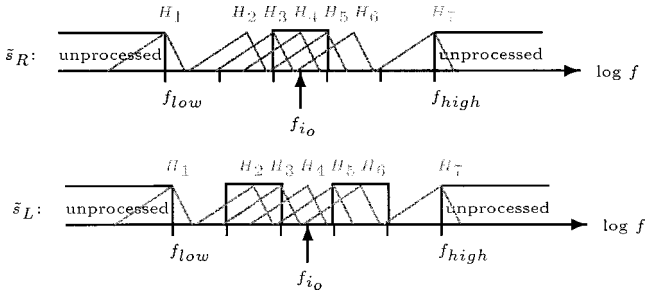


FIG. 6. Overlapping cochlear filters (in gray) superimposed over the spectral representation of $\tilde{s}_R(t)$ (top) and $\tilde{s}_L(t)$ (bottom). For pictorial clarity, the critical-band spectra are sketched as "flat" spectra.

the right- and the left smoothed-envelope signals are defined as

$$\tilde{s}_R(t) = s_{\text{low}}(t) + \tilde{a}_{i_o}(t)\cos\omega_{i_o}t + s_{\text{high}}(t), \tag{15}$$

$$\tilde{s}_L(t) = s_{\text{low}}(t) + \tilde{a}_{i_o-1}(t)\cos\omega_{i_o-1}t + \tilde{a}_{i_o+1}(t)\cos\omega_{i_o+1}t + s_{\text{high}}(t), \tag{16}$$

where $\tilde{a}_i(t)$, $i = i_o-1, i_o, i_o+1$, are the smoothed envelopes of the critical-band signals, and f_i, $i = i_o-1, i_o, i_o+1$, are the center frequencies of the critical bands in the middle frequency range (the gray-colored bands in Fig. 5), respectively. Compared to diotic synthesis, the distance between two successive occupied frequency bands in each of these signals is at least one critical band, resulting in a reduction of distortion due to carrier beating. At CF$=f_{i_o}$ and its one-critical-band neighborhood, the resulting fused auditory image contains smooth-envelope information in accordance with the prescribed bandwidth. This will be demonstrated in the remainder of the section.

D. Properties of the simulated cochlear signals

Figure 6 illustrates the filtering of the signals $\tilde{s}_R(t)$ of Eq. (15) (Fig. 6, top) and $\tilde{s}_L(t)$ of Eq. (16) (Fig. 6, bottom) by a simulated cochlea. In both figures, a sketch of seven (overlapping) cochlear filters is superimposed (in gray) over the spectral description of the signals.

Figure 6, top, illustrates the processing of $\tilde{s}_R(t)$ by the filters. All cochlear filters located to the left of filter H_1 (i.e., filters with lower CFs), and all the filters located to the right of filter H_7 (i.e., filters with higher CFs) will produce envelope signals with unsmoothed temporal structure. Filters H_2 to H_6 will produce temporally smoothed envelopes which are merely filtered versions of $\tilde{a}_{i_o}(t)$, with the response of H_4 being the strongest [and the most similar to $\tilde{a}_{i_o}(t)$]. The responses of filters H_2 and H_6 are negligible, since they are located at the energy gaps of the input signal. The amount of distortion due to beating is negligible since, for any CF, only one occupied frequency band is passing through the corresponding cochlear filter. (This is due to the wide gap, two critical-bands wide, between any adjacent occupied channels.)

Figure 6, bottom, illustrates the processing of $\tilde{s}_L(t)$ by the filters H_1 to H_7 of Fig. 6, top. Since $\tilde{s}_R(t)$ and $\tilde{s}_L(t)$ are identical for $f < f_{\text{low}}$ and for $f > f_{\text{high}}$, so is the response of all cochlear filters located in these frequency ranges. However, the response of cochlear filters in the midfrequency range is different. In contrast to their response to $\tilde{s}_R(t)$, the response of filter H_4 to $\tilde{s}_L(t)$ is the weakest while the envelope signals at the outputs of H_2 and H_6 are the strongest, similar in shape to $\tilde{a}_{i_o-1}(t)$ and $\tilde{a}_{i_o+1}(t)$, respectively [see Fig. 6, bottom, and Eq. (16)]. Also, compared to Fig. 6, top, the gap between adjacent occupied frequency bands is only one critical-band wide, resulting in some distortion due to beating.

Figure 7 shows simulated IHC response at 20 successive CFs to a 70-ms-long segment of the vowel /U/, cut from diphone /m_U/, starting at the transition point of /m/ into /U/. The top section shows the response to $s_R(t)$ and $\tilde{s}_R(t)$; bottom section is for $s_L(t)$ and $\tilde{s}_L(t)$. The channels' CFs (indicated in the upper-left corner of each panel) are equally spaced along the critical-band scale with a spacing of one-fourth critical band, from $f_{\text{low}} = 1722$ Hz to $f_{\text{high}} = 2958$ Hz, i.e., every column (four successive channels) covers one critical band. Each cochlear channel is realized as a gammatone filter, followed by an IHC model.[7] In this example, the target frequency is $f_{i_o} = 2227$ Hz, and the parameters of the dichotic synthesizer are set to $f_{\text{low}} = 1722$ Hz, $f_{\text{high}} = 2958$ Hz, $f_{i_o-1} = 1988$ Hz, $f_{i_o} = 2227$ Hz, and $f_{i_o+1} = 2494$ Hz [see Fig. 5 and Eqs. (13)–(16)]. Each panel in the figure shows the output of the IHC model to the following input signals: Black lines show the output for the signals with unprocessed critical bands, $s_R(t)$ of Eq. (13) (top) and $s_L(t)$ of Eq. (14) (bottom); gray lines show the output for the signals with the envelope-smoothed critical bands, $\tilde{s}_R(t)$ of Eq. (15) (top) and $\tilde{s}_L(t)$ of Eq. (16) (bottom), where a smoothed envelope $\tilde{a}_i(t)$ is the envelope $a_i(t)$, low-pass filtered to 64 Hz. The panel labeled 1722 Hz represents channel H_1 of Fig. 6, panel 2958 Hz represents channel H_7, and panels 1988, 2227, and 2494 Hz represent channels H_2, H_4, and H_6, respectively.

The response shown in Fig. 7 is in accordance with the observations made in Fig. 6. As we see in the top section, the IHCs' response to $s_R(t)$ of Eq. (13) (i.e., black lines) is rich in temporal structure. The overall energy changes with CF, with a stronger response by filters located in occupied frequency regions. The IHCs' response to $\tilde{s}_R(t)$ of Eq. (15) (superimposed gray lines) is rich in temporal structure for CFs below f_{low} and for CFs above f_{high}. However, the response gradually changes with CF, becoming temporally smoothed (and similar to the envelope signal $\tilde{a}_{i_o}(t)$). The output energy peaks at CF$=f_{i_o}$, then slowly decays for filters located at the frequency gap of Fig. 6, top. Note that distortion due to beating is negligible. Analogous behavior is illustrated in the bottom section of Fig. 7. Here, minimum response is produced at CF$=f_{i_o}$ while maximum response is produced at CF values near f_{i_o-1} and f_{i_o+1}. Note also the distortion produced by beating which, for this particular vowel, is most noticeable at CFs in the left energy gap of Fig. 6, bottom (i.e., CF$\approx$1900 Hz).

E. Properties of the fused auditory image

1. Integration of left and right channels

During listening, the subject's response is based upon the information contained in the fused auditory image. The "low-frequency range" and the "high-frequency range" [$s_{\text{low}}(t)$ and $s_{\text{high}}(t)$ of Eqs. (13)–(16)] are presented to the listener diotically, creating an auditory image with conventional properties. However, the midfrequency range is presented dichotically, with interleaving critical bands. This raises a question about the properties of the resulting fused (internal) auditory image. It is reasonable to assume that information from left and right ears originating at similar CFs will be integrated to generate a fused image. The use of

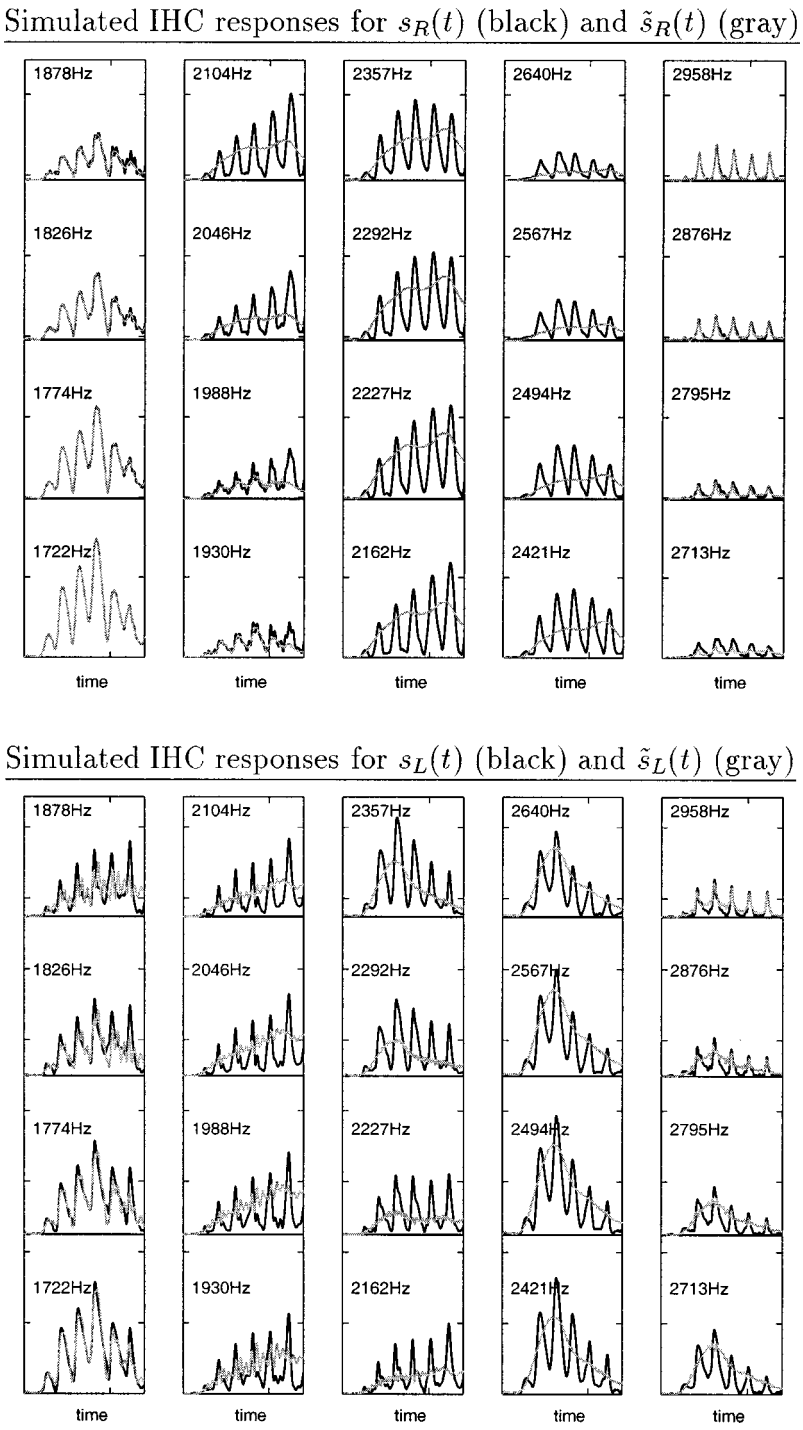


FIG. 7. Simulated IHC response at 20 successive CFs to a dichotically synthesized speech. The figure shows the response to a 70-ms-long segment of the vowel /U/, cut from diphone /m_U/, starting at the transition point of /m/ into /U/. The channels are located one-fourth of one critical band apart, with every column (four successive channels) covers one critical band. Black lines show the output for the input signals with unprocessed critical bands, $s_R(t)$ of Eq. (13) (top) and $s_L(t)$ of Eq. (14) (bottom). Gray lines show the output for the input signals with envelope-smoothed critical bands, $\tilde{s}_R(t)$ of Eq. (15) (top) and $\tilde{s}_L(t)$ of Eq. (16) (bottom), where the envelopes are low-pass filtered to 64 Hz. See the text for details.

dichotic stimulus with interleaving critical bands ensures that, at any CF, when one ear is stimulated, the opposite ear is not. Nevertheless (as illustrated in Figs. 6 and 7), cochlear channels located at the energy gaps of the input signal produce a nonzero output. The proposed synthesis procedure, therefore, only ensures that, at any CF, information from the stimulated ear is stronger than the information from the opposite ear. In Fig. 7, at any given CF, the panel from the top section (say, right ear) is assumed to be combined with the corresponding panel from the bottom section (left ear). In particular, for CFs near f_{i_o}, the signals from the stimulated ear are stronger than the signals from the other ear.

2. Coarse variation of IHC responses with CF

The proposed dichotic synthesis technique produces an inherent distortion due to undersampling (in CF) of the IHC response. Recall that information is conveyed to the AN by a large number of highly overlapped cochlear channels, with a density and location determined by the discrete distribution of the IHCs along the continuous cochlear partition. When a signal with unprocessed critical bands [e.g., $s_R(t)$ or $s_L(t)$] is passed through this cochlear filter bank, the resulting IHC responses change gradually with CF. Passing a signal with envelope-smoothed critical bands [$\tilde{s}_R(t)$ or $\tilde{s}_L(t)$] through

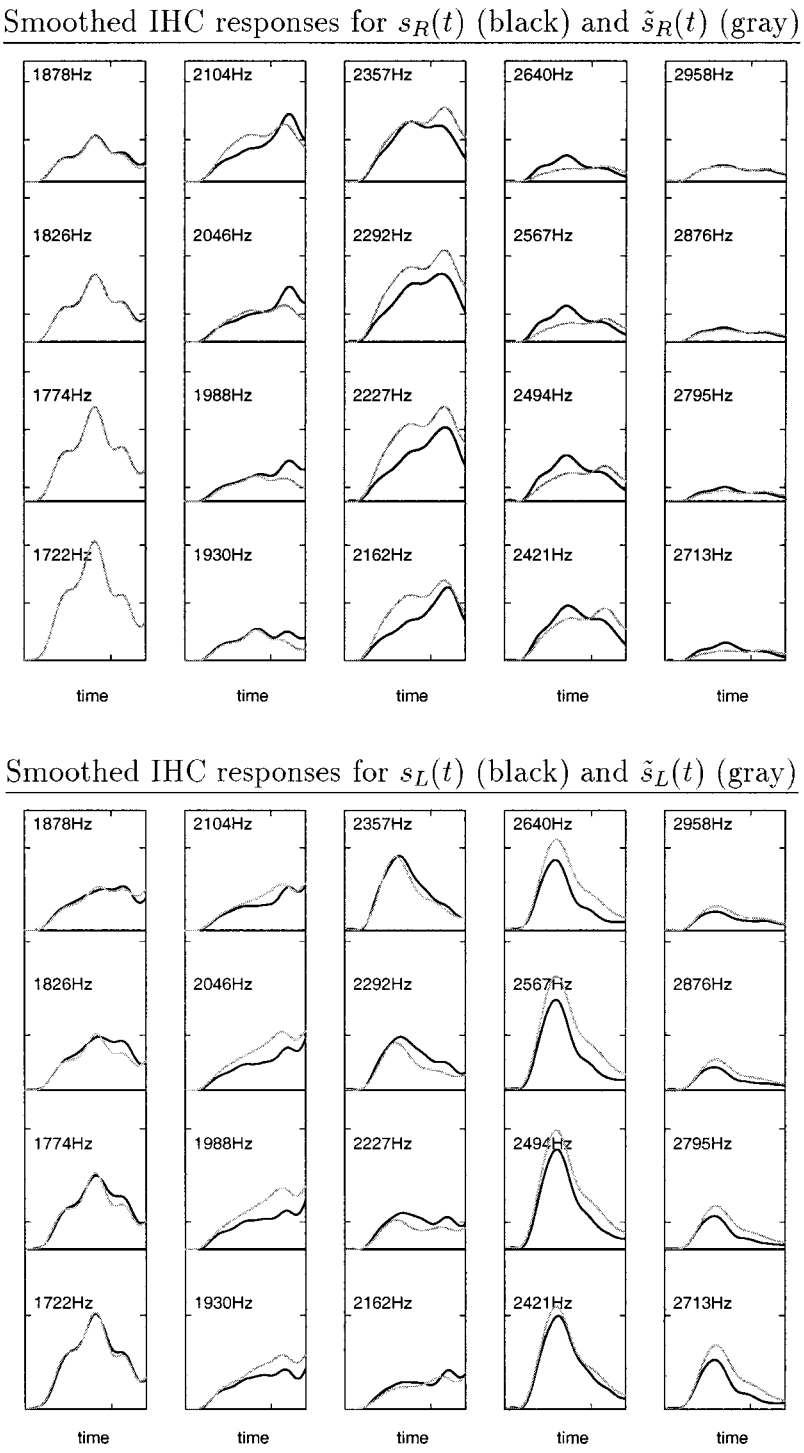


FIG. 8. Illustrating the coarse variation of IHC response with CF, due to the undersampling of the auditory channels (an inherent property of the dichotic synthesis technique). The figure shows the simulated IHC response of Fig. 7 smoothed to 64 Hz, for the input signals with unprocessed critical bands (black), and for the input signals with the envelope-smoothed critical bands (gray). Note the richer variation with CF for the unprocessed input signals (black). Notations are same as in Fig. 7. See the text for details.

the same filter bank will result in much coarser change. This is so because, in synthesizing $\tilde{s}_R(t)$ and $\tilde{s}_L(t)$, pure cosine carriers are used to place a few smoothed-envelope samples (sampled with a frequency resolution of two critical bands) at the appropriate locations along the basilar membrane. This is illustrated in Fig. 8, which is similar to Fig. 7 with the exception that, at each panel, the signals are the corresponding signals of Fig. 7 low-pass filtered to 64 Hz. The figure shows the change in envelope as a function of CF for the input signals with unprocessed critical bands (black) and for the input signals with envelope-smoothed critical bands (gray). With $\tilde{s}_R(t)$ as an input (top section), all overlapping cochlear channels located in the center column are fed with the same amplitude-modulated (AM) signal $\tilde{a}_{i_o}(t)\cos\omega_{i_o}t$, with $f_{i_o}=2227$ Hz. Therefore, the simulated IHC responses of these channels (in gray) are merely filtered versions of $\tilde{a}_{i_o}(t)$, and their similarity to $\tilde{a}_{i_o}(t)$ depends on the frequency response of the corresponding gammatone filter. In contrast, with $s_R(t)$ as an input, the variation in the simulated IHC responses of the corresponding channels (in black) is richer, reflecting the detailed information of the signal with the unprocessed critical bands. Analogous behavior will occur for $s_L(t)$ and $\tilde{s}_L(t)$ as inputs (bottom section). Note that

the coarse variation of the IHC responses with CF limits the extent to which the fused auditory image achieves the property of Eq. (6).

3. Sparse IHC responses for excessive envelope smoothing

Due to the undersampling of the IHC responses (Sec. III E 2) the coarse representation with CF becomes sparse for an excessive envelope smoothing, causing a significant perceivable distortion. If the bandwidth of $\tilde{a}_i(t)$ is B, the bandwidth of the AM signal $\tilde{a}_i(t)\cos\omega_i t$ is $2\times B$. Hence, for $\tilde{s}_{\text{odd}}(t)$ and $\tilde{s}_{\text{even}}(t)$ of Eqs. (11) and (12), each defined as a sum of AM signals for $f>1500$ Hz, the energy gap between two successive occupied frequency bands increases as B decreases. Consequently, more cochlear channels located in between successive cosine carriers will have a weak response, resulting in a sparse fused image. Illustratively, if $B\rightarrow 0$, the upper frequency band of $\tilde{s}_{\text{odd}}(t)$ and $\tilde{s}_{\text{even}}(t)$ becomes a sum of sinusoids. The perceived distortion sounds as an additive monotonic ''musical note.''

4. Spacing between successive cosine carriers

Recall that the dichotic synthesis technique was introduced to reduce perceivable distortions rising from the beating of two modulated cosine carriers passing through a cochlear filter located in between the carriers' frequencies. For the signals $\tilde{s}_{\text{odd}}(t)$ and $\tilde{s}_{\text{even}}(t)$ of Eqs. (11) and (12), the spacing between successive cosine carriers was set to be two critical bands wide. This choice was somewhat arbitrary. Obviously, the greater the spacing is, the smaller the beating-induced distortions are. However, increase in spacing will result in a coarser variation of IHC responses with CF (Sec. III E 2). Analogously, decreasing the spacing, e.g., to reduce sparse envelope representation for small values of B (Sec. III E 3), will reintroduce a perceptible amount of beating-induced distortions. This trade-off between beating-induced distortion and distortions due to sparse envelope representation is inevitable.

IV. DICHOTIC SYNTHESIS AND SPEECH QUALITY—EXPERIMENTS

In this section we use the dichotic synthesis technique to conduct two separate experiments in the context of preserving speech quality. In experiment I (described in Sec. IV B) we examine how speech quality is affected by replacing the carrier information of the critical-band signal by a cosine carrier [i.e., replacing $\cos\phi_{i_o}(t)$ by $\cos\omega_{i_o}t$], while keeping the envelope information untouched. In experiment II (Sec. IV C) we measure how speech quality deteriorates as the envelope bandwidth at the listener's cochlear output is gradually reduced.

A. Database, psychophysical procedure, subjects

The stimuli for the experiments were generated by implementing the dichotic synthesis technique [Eqs. (13)–(16)]. Twelve speech sentences were used, spoken by three female speakers and three male speakers (each speaker contributed two sentences). Since the experiments were conducted in the context of preserving speech quality, wideband speech signals were used, with a bandwidth of 7000 Hz. The speech intensity was set to 75 dB SPL. The stimuli are characterized by the center frequency of the middle frequency range [i.e., f_{i_o} of Eqs. (13)–(16)] and by the processing condition. We used five center frequencies, equally spaced on the critical-band scale and separated by (roughly) two critical bands (1600, 2000, 2500, 3200, and 4000 Hz). We used six processing conditions: one condition representing the signals with *unprocessed critical bands* [where the right and left signals are $s_R(t)$ and $s_L(t)$ of Eqs. (13) and (14), respectively], four conditions representing signals with *envelope-smoothed critical bands* [where the right and left signals are $\tilde{s}_R(t)$ and $\tilde{s}_L(t)$ of Eqs. (15) and (16), respectively], with envelope bandwidths of 512, 256, 128, and 64 Hz, and a *control condition*, termed the *null* condition, where the five successive critical bands centered at f_{i_o} are set to zero.[8]

In both experiments, we used the ABX psychophysical procedure. In this procedure, two sets of stimuli, the ''reference set'' and the ''test set,'' are defined. A stimulus in the reference set has a counterpart in the test set; both stimuli differ only by their processing condition. At each trial, a stimulus from the reference set and its counterpart from the test set are assigned to be the A stimulus and the B stimulus, at random. Then, the X stimulus is randomly chosen to be either the A or the B stimulus. The listener is presented with the A, B, and X stimuli (in this order), and must decide whether X is A or B. In our version, there is no ''repeat'' option. Note that if the listener makes his decisions at random (this may occur if the reference set and the test set are perceptually indistinguishable), the probability of correct decision is 50%.

Five subjects participated in each experiment (same subjects for both experiments). All subjects are well experienced in listening to high-quality audio signals (speech and music).

B. Experiment I—Carrier information

In this experiment we validate the hypothesis that at high CFs the auditory system is insensitive to the carrier information of the critical-band signals and that ascending auditory information in this frequency range is conveyed mainly via the temporal envelope of the cochlear signals. Towards this goal, we measure the probability of correct response in an ABX psychophysical procedure, using a reference set and a test set as defined in Table I. A stimulus in the reference set and its counterpart in the test set differ in the characteristics of the carrier information of the critical-band signals at the middle-frequency range (Fig. 5). As indicated in the middle column of Table I (processing condition), a reference stimulus is comprised of the signals $\tilde{s}_R(t)$ and $\tilde{s}_L(t)$ of Eqs. (15) and (16), respectively, with the envelopes low-pass filtered to 512 Hz (i.e., zero carrier information but full envelope information[9]). The corresponding test stimulus is composed of the signals $s_R(t)$ and $s_L(t)$ of Eqs. (13) and (14), respectively (i.e., containing the full carrier and the full envelope information).

TABLE I. Stimuli for experiment I (Sec. IV B) and experiment II (Sec. IV C). Each entry denoted by * contains 12 sentences, spoken by three female and three male speakers (two sentences each).

	Processing condition		Center frequency f_{i_o}, in Hz				
	Carrier	Envelope bandwidth	1600	2000	2500	3200	4000
Reference	$\cos \omega_{i_o} t$	512 Hz	*	*	*	*	*
Test—Experiment I	$\cos \phi(t)$	full	*	*	*	*	*
	$\cos \omega_{i_o} t$	256 Hz	...	...	*	*	*
Test—Experiment II	$\cos \omega_{i_o} t$	128 Hz	*	*	*	*	*
	$\cos \omega_{i_o} t$	64 Hz	*	*	...	...	...
Test—Control	null	null	*	*	*	*	*

C. Experiment II—Envelope bandwidth

In this experiment we measure the upper cutoff frequency of the auditory critical-band envelope detector, in terms of the minimal bandwidth of the critical-band envelope that ensures transparent speech quality. Towards this goal, we measure the probability of correct response in an ABX psychophysical procedure, using a reference set and a test set as defined in Table I. A reference stimulus and the corresponding test stimulus are composed of the signals $\tilde{s}_R(t)$ and $\tilde{s}_L(t)$ of Eqs. (15) and (16), respectively. They differ only in the bandwidth of the critical-band envelopes, with the bandwidth of a reference stimulus being 512 Hz. In the test set, only two smoothing conditions were used at each center frequency (to reduce the overall number of trials, and hence the experimental load on the subjects). For $f_{i_o} = 1600$ Hz and $f_{i_o} = 2000$ Hz, the envelope bandwidths were 64 and 128 Hz. (Note that the bandwidth of critical bands located at these center frequencies are 180 and 250 Hz, respectively.) For $f_{i_o} = 2500$ Hz, $f_{i_o} = 3200$ Hz, and $f_{i_o} = 4000$ Hz, the envelope bandwidths were 128 and 256 Hz (where the corresponding bandwidth of critical bands are 300, 360, and 440 Hz).

D. Results

In conducting the experiment, all test stimuli of experiment I, experiment II, and the control experiment were combined into one set ([5 center frequencies]×[4 test processing conditions]×[12 sentences]=240 sentences–see Table I). These sentences were randomly shuffled, then divided into four groups of 60 sentences each. The counterpart reference stimuli were arranged in the same order. Each subject participated in four sessions (a group of 60 sentences per session), lasting about 10 min each ([60 ABX trials]×[3 sentences]×[≈3 seconds]=≈600 seconds).

The results are presented in Fig. 9. Each panel represents performance at the center frequency specified at the upper-right corner of the panel. The bandwidth of a critical band[10] centered at that frequency is also indicated in parentheses. The abscissa of each panel indicates the processing condition of the test set stimuli. The entry $s_i(t)$ represents the condition with unprocessed critical bands (experiment I), the entries 256, 128, and 64 Hz represent the conditions with envelope-smoothed critical bands (experiment II), and the entry *null* represents the control experiment. (We chose to display all conditions in the same panel since a test set, in all experiments, is always contrasted with the same reference set—see Table I.) The ordinate is the probability of correct identification of the identity of the X stimuli (during the ABX procedure), in percent. The proportion of correct response for each subject was computed from 12 binary responses (one binary response for each sentence in the experiment). Each entry shows the mean and the standard deviation

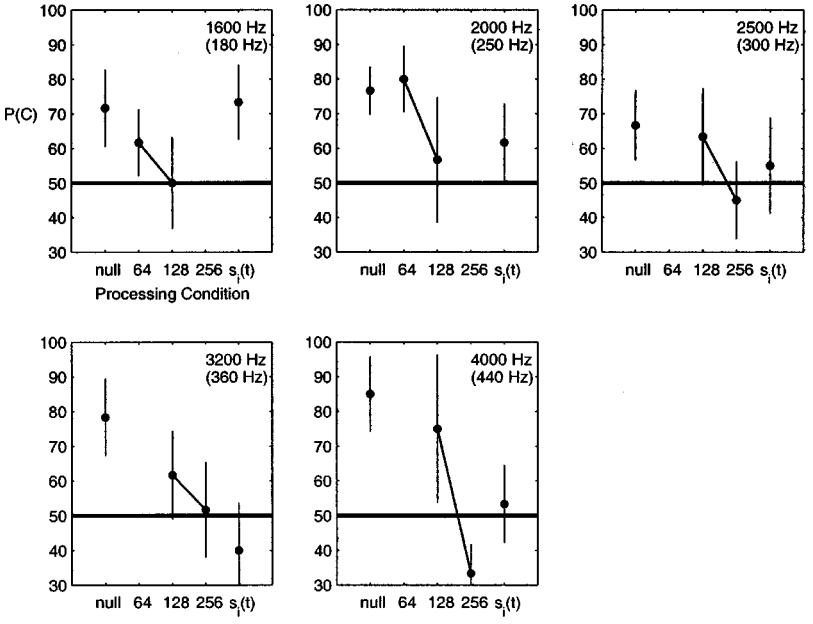


FIG. 9. Probability of correct response as a function of processing condition, with the center frequency as a parameter. Center frequencies are specified at the upper-right corner of the panel (the bandwidth of the corresponding critical bands is also indicated, in parentheses). The abscissa of each panel indicates the processing condition of the test set stimuli. The ordinate is the probability of correct identification of the identity of the X stimuli (during the ABX procedure), in percent. Each entry shows the mean percentage of correct response and the standard deviation among the five subjects. See the text for details.

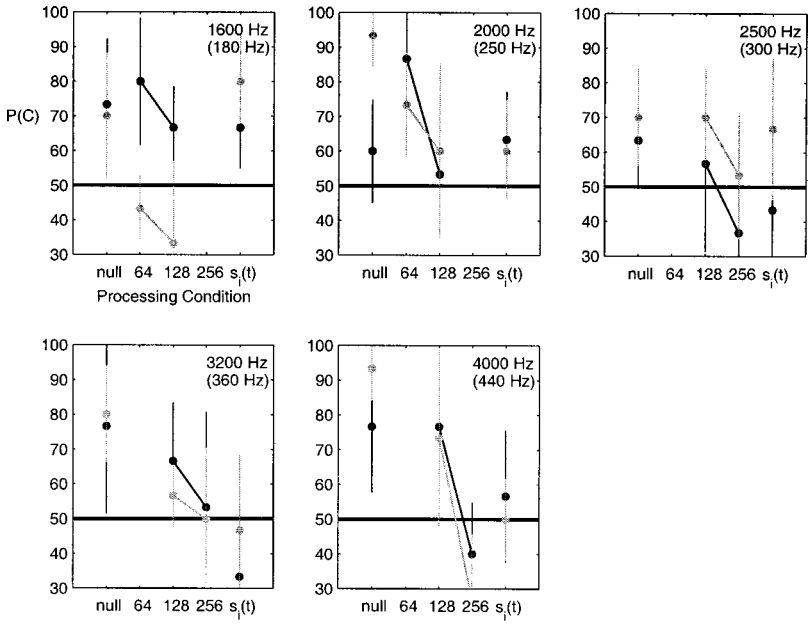


FIG. 10. Experimental results of Fig. 9, broken into two groups according to speaker gender, male speakers in black, female speakers in gray. Differences may be attributed to the interaction between the spectral contents of the stimulus (location of formants, pitch) and the center frequency under consideration.

of these five numbers. (A simple analysis of variance demonstrated that the interaction between subject and processing condition was not significant, so that it is legitimate to pool results from the five subjects.)

The control experiment (indicated *null* on the abscissa) confirms the assumption that a removal of a frequency band five-critical-bands wide results in a perceivable degradation in quality. This is so because for all center frequencies we considered, the mean probability of correct response is significantly above 50%.

For experiment I [indicated as $s_i(t)$ on the abscissa], the mean probability of correct response is about 50% for the higher center frequencies (i.e., 2500, 3200, and 4000 Hz). As the center frequency decreases, the mean probability of correct response increases (62% for $f_{i_o}=2000\,\mathrm{Hz}$, and 74% for $f_{i_o}=1600\,\mathrm{Hz}$). This result confirms the hypothesis that at high center frequencies (above $\approx 1800\,\mathrm{Hz}$) the auditory system is insensitive to the temporal details of the carrier information, and that the full carrier $\cos\phi(t)$ can be replaced with a cosine carrier $\cos\omega t$.

For experiment II (indicated as 64, 128, and 256 Hz on the abscissa), at higher center frequencies (i.e., 2500, 3200, and 4000 Hz) the mean probability of correct response is about 50% for an envelope bandwidth of 256 Hz.[11] For the other two center frequencies (1600 and 2000 Hz), a 50% mean probability of correct response is measured for an envelope bandwidth of 128 Hz. Note that these bandwidth values are considerably smaller than the bandwidth of the critical bands centered at the corresponding center frequencies (indicated in the upper-right corner, in parentheses), and are roughly one-half of one critical band.

Finally, Fig. 10 shows the experimental results of Fig. 9, broken into two groups according to speaker gender, male speakers in black, female speakers in gray. (Obviously, the number of observations per entry per subject is now only six.) The figure shows that at most center frequencies and for most processing conditions, performance is not affected much by the speaker gender. Differences may be attributed to the interaction between the spectral contents of the stimulus (location of formants, pitch) and the center frequency under consideration.

V. DICHOTIC SYNTHESIS AND SPEECH INTELLIGIBILITY

In Sec. IV, the dichotic synthesis technique was used to measure the cutoff frequencies of the auditory envelope detectors at threshold (i.e., the cutoff frequencies which maintain the quality of the original speech). A question arises whether the technique can also be used to measure the cutoff frequencies in the context of speech intelligibility, for speech signals that maintain some reasonable level of speech quality (say, above MOS level 3). In the following, it will be argued that speech stimuli produced by dichotic synthesis for intelligibility-related experiments are of poor quality, with MOS readings well below 3.

Suppose that we want to repeat the phoneme identification experiment reported by Drullman *et al.* (1994), by using a dichotically synthesized speech, with temporal envelopes that are low-pass filtered to a cutoff frequency B. Which values of B are reasonable for such an experiment? Expressing temporal envelope information in terms of the amplitude-modulation spectrum, two kinds of modulations may be considered as information carriers of speech intelligibility—the articulatory modulations and the pitch modulations. Of these, the pitch modulations convey only a limited amount of phonemic information (this is so because for speech signals, the salient mechanism for pitch perception is based on resolved harmonics at the lower frequency range[12]). The major carriers of phonemic information are, therefore, the articulatory modulations. [Indeed, the STI method is aimed at measuring these MTFs (Steeneken and Houtgast, 1980).] Hence, the B values for a phoneme identification experiment should be on the order of a few tens of Hz, determined by the mechanical properties of the articulators. Recall the properties of the speech signals generated by the dichotic synthesis technique

(Secs. III D and III E). For an appropriate spacing between successive cosine carriers (Sec. III E 4), and for B values of a few tens of Hz, the resulting speech stimuli generate fused auditory images that are too sparse (Sec. III E 3), and suffer severe degradation in speech quality (to MOS levels well below 3) due mainly to an overriding monotonic tonal accent. The speech signals produced by the dichotic synthesis technique are, therefore, inadequate for experiments intended to measure intelligibility-related Bs while maintaining fair quality levels. The appropriate signal-processing method is yet to be found.

VI. DISCUSSION

This study was motivated by the need to quantify the minimum amount of information, at the auditory-nerve level, that is necessary for maintaining human performance in tasks related to speech perception (e.g., threshold measurements for speech quality, phoneme classification for speech intelligibility). Such data are needed, for example, for a quantitative formulation of a perception-based distance measure between speech segments (e.g., Ghitza and Sondhi, 1997). The study was restricted to the frequency range above 1500 Hz, where the information conveyed by the auditory nerve is mainly the temporal envelopes of the critical-band signals. From the outset, it was assumed that these envelopes are processed by distinct, albeit unknown, auditory detectors characterized by their upper cutoff frequencies which, in turn, determine the perceptually relevant information of the envelope signals in terms of their effective bandwidth. The main contribution of this study is the establishment of a framework that allows the direct psychophysical measurement of this bandwidth, using speech signals as the test stimuli.

Measuring the perceptually relevant content of temporal envelopes was the subject of numerous studies, most of which were aimed at measuring the amplitude-modulations spectra using threshold-of-detection criteria. These studies (e.g., Viemeister, 1979; Dau *et al.*, 1997a, 1997b, 1999; Kohlrausch *et al.*, 2000) used nonspeech signals as test stimuli—mostly signals with a bandwidth of one critical band.[13] The present study extends the scope of previous studies by providing threshold measurements of the cochlear temporal envelope bandwidth (which may be regarded as the bandwidth of the amplitude-modulation spectrum) for speech signals, hence providing an estimate of the threshold bandwidth of a target auditory channel *while all other channels are active simultaneously*.

In order to conduct these experiments, a signal-processing framework had to be formulated that would be capable of producing speech signals with appropriate temporal envelope properties. As was shown in Sec. II, if the envelope of a critical-band signal is temporally smoothed while the instantaneous phase information remains untouched (e.g., Drullman *et al.*, 1994), the resulting synthetic speech signal evokes cochlear envelope signals that are not necessarily smoothed. This rather counterintuitive behavior (which is theoretically founded, as discussed in Sec. II A) suggests that a different criterion should be used for signal synthesis, such that the resulting speech signal will evoke temporal envelopes with a prescribed bandwidth *at the output of the listener's cochlea* (Sec. III A). Such a signal-processing technique is yet to be found. However, in Sec. III C, an approximate solution has been introduced based upon dichotic speech synthesis with interleaving smoothed critical-band envelopes.[14]

With this technique established, psychophysical measurements were conducted using high-quality, wideband, speech signals (bandwidth of 7 kHz) as the test stimuli. The measurements show that in order to maintain the quality of the original speech signal (1) there is no need to preserve the detailed timing information of the critical-band signal (experiment I, Sec. IV B); (2) the perceptually relevant information in this frequency range is mainly the temporal envelope of this signal, and (3) the minimum bandwidth of the temporal envelope of the critical-band signal is, roughly, one-half of one critical-band (experiment II, Sec. IV C). These results are in line with the widely accepted observation that at higher center frequencies, due to the physiological limitations of the inner hair cells to follow detailed timing information, neural firings at the auditory nerve mainly represent the temporal envelope information of the critical-band signal.

The data obtained here can be compared to previously published data only qualitatively, because of the marked difference in the underlying frameworks. As discussed by others (e.g., Dau *et al.*, 1999; Kohlrausch *et al.*, 2000), a reliable measurement of amplitude-modulation spectra can be obtained when the stimulus bandwidth is sufficiently narrower than the critical band of the target auditory channel. Previous studies that meet this requirement provide tight estimates of the envelope bandwidth at threshold, since the measurements for the target auditory channel are obtained with zero external stimulation of all other channels. In contrast, the measurements in the present study are taken with all auditory channel simultaneously active (the test stimuli are wideband speech signals), allowing interaction across channels (e.g., due to spread of masking). A qualitative comparison shows that estimates of envelope bandwidths obtained in this study are indeed lower than those published earlier. For example, for an auditory channel at CF of 3000 Hz, the estimate of the envelope bandwidth using a cosine carrier is roughly one critical band (i.e., about 350 Hz, Kohlrausch *et al.*, 2000). For speech stimuli at similar CFs, the envelope bandwidth is about 250 Hz (Fig. 9).

The methodology presented in this study provides a framework for the design of transparent coding systems[15] with a substantial information reduction (due to the use of fixed cosine carriers, modulated by smoothed critical-band envelopes, Ghitza and Kroon, 2000). One desirable property of this coding paradigm is that it performs equally well for speech, noisy speech, music signals, etc. This is so since the coding paradigm is based solely on the properties of the auditory system and does not assume any specific properties of the input source.

Finally, the dichotic synthesis technique is inadequate for the purpose of measuring the cutoff frequencies relevant to intelligibility of speech signals with fair quality levels (say, above MOS=3). Recall that the main information car-

riers of speech intelligibility are the articulatory modulations (e.g., Sec. V). Following a reasoning similar to the one used in measuring the cutoff frequencies at threshold, the appropriate speech stimuli should satisfy the criterion of generating temporal envelopes with *smoothed articulatory modulations* at the output of the listener's cochlea. In view of the discussion in Sec. II, a speech signal produced by smoothing the envelope signal alone (while keeping the original instantaneous phase information untouched) is inadequate because it will regenerate, at the cochlear output, most of the original envelope information, including the articulatory modulations and the pitch modulations. Indeed, the dichotic synthesis technique is capable of producing speech stimuli that generate cochlear temporal envelopes with smoothed articulatory modulations as desired. Alas, the quality of these signals is well below MOS=3 (Sec. V). We still lack the knowledge of how to synthesize speech stimuli which simultaneously satisfy both requirements (i.e., cochlear temporal envelopes with smoothed articulatory modulations *and* a prescribed level of speech quality).

ACKNOWLEDGMENTS

I wish to thank M. M. Sondhi and Y. Shoham for stimulating discussions throughout this work, and S. Colburn and two anonymous reviewers for reviewing earlier versions of the paper.

[1]Signals presented to left and right ears are different.

[2]The Mean-Opinion-Score, or MOS, is a test which is widely used to assess quality of speech coders. It is a subjective test that can be categorized as a rating procedure. Subjects are presented, once, with a speech sentence and are requested to score its quality using a scale of five grades. The grades (and their numerical aliases) are Excellent (5), Good (4), Fair (3), Poor (2), and Bad (1). The MOS is the mean score, averaged over the database and the subjects.

[3]The *Hilbert envelope* and the *Hilbert instantaneous phase* are defined as follows: Let $z_i(t)$ be the *analytic signal* of $s_i(t)$, i.e., $z_i(t)=s_i(t)+j\hat{s}_i(t)$, where $\hat{s}_i(t)$ is the Hilbert transform of $s_i(t)$. We express $s_i(t)$ in terms of $z_i(t)$ as $s_i(t)=\Re(z_i(t))=a_i(t)\cos\phi_i(t)$, where $a_i(t)=\sqrt{s_i^2(t)+\hat{s}_i^2(t)}$ is the *envelope* of $s_i(t)$, and $\phi_i(t)=\arctan[\hat{s}_i(t)]/[s_i(t)]$ is the *instantaneous phase* of $s_i(t)$.

[4]CF, for *Characteristic Frequency*, indicates the place of origin of a nerve fiber along the basilar membrane in frequency units.

[5]Obviously, there is no distinct boundary between the low-CF and high-CF AN regions. Rather, the change in properties is gradual. Our working hypothesis is that the region of transition is around 1500 Hz.

[6]The same signal is presented to both ears.

[7]The IHC model is comprised of a half-wave rectifier, followed by a low-pass filter with the amplitude transfer function $\|H(f)\|=1/\sqrt{(1+(f/600)^2)(1+(f/3000)^2)}$, reflecting the synchrony roll-off in AN firings (e.g., Johnson, 1980).

[8]The null condition is for control purposes, to validate the assumption that a removal of a frequency band five-critical-bands wide indeed causes perceivable degradation in quality.

[9]Note that the bandwidth of the critical band centered at the highest center frequency considered in this experiment (i.e., $f_{i_o}=4000$ Hz) is about 440 Hz.

[10]We follow the ERB definition of a critical band, according to Moore and Glasberg (1983).

[11]Note that at center frequency of 4000 Hz the mean probability of correct response, for an envelope bandwidth of 256 Hz, is about 33%. This indicates that the two conditions are being distinguished somehow, but that the response is consistently incorrect.

[12]Recall the existence of two competing mechanisms for pitch perception. One is based upon resolved harmonics and, for speech signals in particular, operates at the lower frequency range (say, below 1500 Hz);, the other is based on temporal envelope periodicities and operates at the higher frequency range. When both mechanisms are active (as in the case of speech signals) the salient mechanism is the former one (e.g., Goldstein, 2000).

[13]The study by Drullman *et al.* (1994) belongs to a different category since it used a threshold criterion related to speech intelligibility (i.e., percent correct in a phoneme classification task). Obviously, Drullman *et al.* had to use speech signals as test stimuli.

[14]See Secs. III D and III E for a discussion on the properties and the shortcomings of this approximate solution.

[15]That is, at the receiving end, the system produces speech signals that are perceptually indistinguishable from the original speech.

Chi, T., Gao, Y., Guyton, M. C., Ru, P., and Shamma, S. (**1999**). "Spectro-temporal modulation transfer functions and speech intelligibility," J. Acoust. Soc. Am. **106**, 2719–2732.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997a**). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band 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," J. Acoust. Soc. Am. **102**, 2906–2919.

Dau, T., Verhey, J., and Kohlrausch, A. (**1999**). "Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band noise carriers," J. Acoust. Soc. Am. **106**, 2752–2760.

Drullman, R., Festen, J. M., and Plomp, R. (**1994**). "Effect of temporal envelope smearing on speech reception," J. Acoust. Soc. Am. **95**, 1053–1064.

Durlach, I. N., and Colburn, S. (**1978**). "Binaural phenomena," in *Handbook of Perception, Volume IV: Hearing*, edited by E. C. Carterette and M. P. Friedman (Academic, New York), pp. 365–466.

Eddins, D. A., and Green, D. M. (**1995**). "Temporal integration and temporal resolution," in *Hearing*, edited by B. C. J. Moore (Academic, New York), pp. 207–242.

Flanagan, J. L. (**1980**). "Parametric coding of speech spectra," J. Acoust. Soc. Am. **68**, 412–430.

Ghitza, O., and Kroon, P. (**2000**). "Dichotic presentation of interleaving critical-band envelopes: An application to multi-descriptive coding," in *Proceedings of the IEEE Workshop on Speech Coding*, Delavan, Wisconsin (September), pp. 72–74.

Ghitza, O., and Sondhi, M. M. (**1997**). "On the perceptual distance between speech segments," J. Acoust. Soc. Am. **101**, 522–529.

Goldstein, J. L. (**2000**). "Pitch perception," in *Encyclopedia of Psychology*, edited by A. E. Kazdin (American Psychological Association, Washington, D.C.), Vol. VI, pp. 201–210.

Johnson, D. H. (**1980**). "The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones," J. Acoust. Soc. Am. **68**, 1115–1122.

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. **108**, 723–734.

Moore, B. C. J., and Glasberg, B. R. (**1983**). "Suggested formula for calculating auditory-filter bandwidth and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.

Rice, S. O. (**1973**). "Distortion produced by band limitation of an FM wave," Bell Syst. Tech. J. **52**, 605–626.

Slaney, M. (**1993**). "An efficient implementation of the Patterson-Holdsworth auditory filter bank," Technical Report 33, Apple Computer.

Steeneken, H. J. M., and Houtgast, T. (**1980**). "A physical method for measuring speech-transmission quality," J. Acoust. Soc. Am. **67**, 318–326.

Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.

Voelcker, H. B. (**1966**). "Towards a unified theory of modulation. I. Phase-envelope relationships," Proc. IEEE **54**(3), 340–354.

A probabilistic union model with automatic order selection for noisy speech recognition

Peter Jančovič[a] and Ji Ming
School of Computer Science, The Queen's University of Belfast, Belfast BT7 1NN, United Kingdom

(Received 4 January 2001; accepted for publication 23 May 2001)

A critical issue in exploiting the potential of the sub-band-based approach to robust speech recognition is the method of combining the sub-band observations, for selecting the bands unaffected by noise. A new method for this purpose, i.e., the probabilistic union model, was recently introduced. This model has been shown to be capable of dealing with band-limited corruption, requiring no knowledge about the band position and statistical distribution of the noise. A parameter within the model, which we call its order, gives the best results when it equals the number of noisy bands. Since this information may not be available in practice, in this paper we introduce an automatic algorithm for selecting the order, based on the state duration pattern generated by the hidden Markov model (HMM). The algorithm has been tested on the TIDIGITS database corrupted by various types of additive band-limited noise with unknown noisy bands. The results have shown that the union model equipped with the new algorithm can achieve a recognition performance similar to that achieved when the number of noisy bands is known. The results show a very significant improvement over the traditional full-band model, without requiring prior information on either the position or the number of noisy bands. The principle of the algorithm for selecting the order based on state duration may also be applied to other sub-band combination methods. [DOI: 10.1121/1.1387083]

PACS numbers: 43.72.Ne, 43.72.Dv, 43.72.Ar [DOS]

I. INTRODUCTION

In the sub-band approach to robust speech recognition in noise, the entire speech frequency band is divided into several sub-bands, each sub-band being modeled as an independent feature stream. In recognition, the probabilities associated with the individual sub-bands are combined to generate an overall probability. The sub-band approach has been shown to be capable of achieving better recognition performance than the full-band approach for dealing with speech corrupted by band-limited or band-selective noise (Hermansky *et al.*, 1996; Bourlard and Dupont, 1996, 1997; Tibrewala and Hermansky, 1997a, 1997b; Cerisara *et al.*, 1998; Okawa *et al.*, 1998; Bourlard, 1999; Morris, Hagen, and Bourlard, 1999). Because these types of noise may only cause a partial corruption within the sub-bands, the other sub-bands that are not affected by the noise can still provide correct information about the speech utterance.

To exploit this potential, we face the problem of how to select from the given set of sub-bands the bands that are unaffected or affected only slightly by the noise. This is related to a more general problem: missing feature theory (Cooke, Morris, and Green, 1997; Lippmann and Carlson, 1997). This selection is difficult if there is no prior information about the noise. In previous studies, several strategies have been used to produce the overall probability given the sub-band probabilities. These typically include the weighted-average method (Bourlard and Dupont, 1996, 1997; Cerisara *et al.*, 1998; Okawa *et al.*, 1998; Morris, Hagen, and Bourlard, 1999) and the neural-network method (Hermansky *et al.*, 1996; Tibrewala and Hermansky, 1997a, 1997b). The weighted-average method produces the overall probability based on an arithmetic or geometric average of different combinations of different sub-band probabilities, and the contribution of each sub-band is weighted by its signal-to-noise ratio (SNR) (e.g., Okawa *et al.*, 1998), or by its relative reliability (e.g., Morris, Hagen, and Bourlard, 1999). In the neural-nets approach, independent networks may be trained to estimate the probabilities of all possible combinations of the sub-bands. This approach faces the same problem of how to select the best combination from all the combinations given no knowledge about the noisy bands.

Recently, a new approach, i.e., the probabilistic union model, has been studied as an alternative solution to the above sub-band combination problem (Ming and Smith, 1999, 2000, 2001). The new model combines the sub-bands based on the probability theory for the union of random events (i.e., subsets of the sub-band observations). This model requires no prior knowledge or an estimation of the noisy bands. Experiments have shown that the union model is capable of dealing with unknown, time-varying band-limited noise (Ming and Smith, 1999, 2000, 2001).

While the union model assumes no knowledge on the noisy bands, its performance can be maximized when the number of noisy bands is known (Ming and Smith, 1999, 2000, 2001). This performance maximization is achieved by matching the order of the model to the actual number of the noisy bands. To obtain this match, we may directly estimate the number of corrupted bands, based on which an appropriate model order can be decided. However, although this estimation may be easier than the estimation of the exact position of corrupted bands (which actually includes estimating

[a]Electronic mail: p.jancovic@qub.ac.uk

both the number *and* position of the noisy bands), a method for selecting the model order without requiring any knowledge on the noise is of greater practical significance. Ming and Smith (1999, 2000, 2001) discussed this problem and suggested that if there is no knowledge about the environment (e.g., clean, or noisy with unknown noisy bands), we may select an order to accommodate as many noisy bands as possible, subject to an acceptable performance for clean speech recognition. For convenience, we call this a balanced fixed-order method, which represents a compromise between accuracy and robustness.

This study is deepened by introducing an *automatic* algorithm for order selection for the probabilistic union model, incorporated into a HMM framework. This algorithm selects the order based on the state duration pattern generated by the HMM built on the union model. In recognition, the state-duration patterns produced by the union model of varying orders are recorded, and then compared to the duration pattern generated on the clean training data. The best order is defined as the order that produces the closest match between the training and testing state-duration patterns. In comparison to the above fixed-order method, the new algorithm operates on the utterance basis and produces an order that can closely match the actual number of noisy bands (including the no noise case). This new algorithm has been tested on the TIDIGITS database corrupted by various types of noise with unknown noisy bands. The results have shown a further significant improvement over the balanced fixed-order method, which had been shown previously to be a significant improvement over the full-band method.

This paper is organized as follows. In Sec. II we provide a brief review of the probabilistic union model. In Sec. III we present the recognition results of the experiments performed for all the orders of the union model. In Sec. IV, we describe the automatic algorithm for order selection. Experimental results based on the automatic algorithm are presented in Sec. V. In Sec. VI we give a summary of this work.

II. PROBABILISTIC UNION MODEL

A. Theory

Assume a recognition system with N sub-bands, each sub-band being characterized by a feature stream o_n, $n=1,\ldots,N$. In recognition, we are given a set of sub-band features $o=(o_1,o_2,\ldots,o_N)$ in which some of the o_n's may be noisy, due to some unknown band-limited or band-selective noise. When there is no noise, the traditional approach for extracting information from this feature set is to combine the individual sub-band features by using the ''and'' (i.e., conjunction) operator $\wedge$. Assuming that the sub-band features are independent of one another, the overall likelihood of this feature set then equals the product of the individual likelihoods $p(o_n)$'s associated with each o_n, i.e.,

$$p(o)=p(o_1\wedge o_2\wedge\cdots\wedge o_N)=p(o_1)\cdot p(o_2)\cdots p(o_N). \quad (1)$$

We call the model in Eq. (1) the product model. For this model, if the probability density functions of the individual sub-bands, $p(x_n)$'s, are trained on clean speech and used for modeling an utterance with some noisy sub-bands, then the corresponding $p(o_n)$'s for the noisy o_n's will be highly inaccurate—on the correct model they may become very small, because of the mismatch between the model and data, particularly if the noise is strong. This can destroy the model's ability to produce high likelihoods for the correct word class. To overcome this problem, we may alternatively assume that, in a given feature set $o=(o_1,o_2,\ldots,o_N)$, the useful features that characterize the speech utterance may be any of the o_n's, $n=1,\ldots,N$, or any combinations of the o_n's, including the complete feature set. This can be expressed, based on the inclusive ''or'' (i.e., disjunction) operator $\vee$, as

$$o_\vee=o_1\vee o_2\vee\cdots\vee o_N, \quad (2)$$

where $o_\vee$ is a combined observation based on $\vee$. If we assume that the o_n's are discrete random vectors, then the probability $P(o_\vee)$ of $o_\vee$ can be computed based on the rules of probability for the union of random events. Specifically, assuming that o_n's are mutually independent and note that $\vee_{n=1}^{m} o_n=(\vee_{n=1}^{m-1} o_n)\vee o_m$, then the probability $P(o_\vee)$ can be computed using the recursion

$$\begin{aligned} P(o_\vee)&=P(\vee_{n=1}^{m} o_n)\\ &=P(\vee_{n=1}^{m-1} o_n)+P(o_m)-P(\vee_{n=1}^{m-1} o_n)P(o_m). \end{aligned} \quad (3)$$

Equations (2) and (3) are called the probabilistic union model (Ming and Smith, 1999, 2000, 2001). The advantage of Eq. (3) over Eq. (1) (i.e., the product model) for noisy speech is that, because Eq. (3) is effectively the sum of the individual probabilities, a small probability $P(o_n)$ makes only a small contribution to Eq. (3), and as such will have little effect on $P(o_\vee)$. This effectively reduces the influence of noisy bands, which typically produce small $P(o_n)$'s, on the overall probability associated with the correct model. In other words, the probability of the correct model is dominated by noiseless bands. As long as there is one noiseless band and the bandwidth is not too small, Eq. (3) should be able to produce high probability for the correct word classes with more success than the product model. However, because Eq. (3) effectively averages the probabilities of individual sub-band features, it is not an effective model for a feature set in which there are two or more clean sub-bands. This can be improved if there is a knowledge on the number (not the position) of the noisy bands. Specifically, for an N-band system, if we know that the noise affects M bands ($M<N$), then we should combine the remaining $(N-M)$ clean bands by using the ''and'' operator to accumulate—rather than to average—the information of each clean band. Because we assume no knowledge about the position of the noisy bands, these $(N-M)$ clean bands can be any of the combinations of $(N-M)$ bands. This can be modeled by using the ''or'' operator. So, we obtain a model in which the ''and,'' and ''or'' operators are combined, i.e.,

$$o_{\vee_M}=\bigvee_{n_1 n_2\cdots n_{N-M}} (o_{n_1}\wedge o_{n_2}\wedge\cdots\wedge o_{n_{N-M}}), \quad (4)$$

where the ''or'' operator $\vee$ is taken over all possible combinations of N different values $(1,2,\ldots,N)$ taken $(N-M)$ at a time, resulting in a total of ${}^{N}C_{N-M}$ combinations. Equation (4) is called the union model with order M (Ming and Smith, 1999, 2000, 2001). This model is best suited to the situation

where the number of noisy bands equals M, in which case Eq. (4) will include a conjunction of all the remaining $(N-M)$ clean bands. This conjunction leads to a joint probability over the clean bands, which should dominate the probability associated with the correct model. The probability expression for Eq. (4) can be derived by replacing the o_n in Eq. (3) by the appropriate conjunctions of sub-band features $o_{n_1}o_{n_2}\cdots o_{n_{N-M}}$. Equation (4) is reduced to the product model of Eq. (1) when $M=0$, and to the model of Eq. (3) when $M=N-1$.

As indicated above, the value of the order M corresponds to the maximum number of noisy bands that can be accommodated by the union model. In the previous work by Ming and Smith (1999, 2000, 2001), the order was decided based on a compromise between the maximum clean performance and robustness, i.e., select an order that can accommodate as many noisy bands as possible, subject to an acceptable performance for clean speech. As such, a fixed model order is used for all situations. The main topic of this paper is to develop an algorithm to choose the order automatically on each utterance basis. This ensures that the union model can always approach an optimum condition in terms of a match to the actual noise. The algorithm is introduced in detail in Sec. IV. Because it is based on the state duration patterns produced by the HMM, in the next section we shortly describe the incorporation of the union model into the HMM.

B. Incorporation into HMM

Let us assume an N sub-bands recognition system, and that the speech in each sub-band is represented by a sequence of frame vectors $o_n(1)$, $o_n(2),\ldots,o_n(T)$, $n=1,...,N$. Consider the combination of the individual sub-band observations by the union model at the frame level. This combination is performed within a HMM, by using an observation probability distribution, in state i, of a form $B_i(o_{\vee_M})$, where $o_{\vee_M}$ is defined by Eq. (4), with each o_n corresponding to a frame from each of the sub-bands. Denote by $o_{\vee_M}(t)$ the combined observation for bands $o_1(t)$, $o_2(t),\ldots,o_N(t)$ at time t. Then the HMM can be expressed as

$$P(o|\lambda)=\sum_s \pi_{s_0}\prod_{t=1}^{T} a_{s_{t-1}s_t}B_{s_t}(o_{\vee_M}(t)), \tag{5}$$

where o represents the frame sequences for all the sub-bands, $s=s_0s_1,\ldots,s_T$ is the state sequence of the observation, and λ represents the model parameter set including $\{\pi_i\}$, $\{a_{ij}\}$ and parameters of $B_i(o_{\vee_M})$. As discussed in the previous section, if we assume that the frames across the sub-bands are statistically independent, then the probability $B_i(o_{\vee_M})$ is only a function of the individual frame probabilities $B_i(o_n)$'s. In the case of discrete-observation HMMs these probabilities are readily available. In the case of continuous-observation HMM these probabilities can be approximated by using the corresponding likelihood values (Ming and Smith, 1999, 2000, 2001). Two methods for this approximation have been studied. In the first method, the probability of each individual frame $B_i(o_n)$ is approximated from the likelihood $b_i(o_n)$ by using a sigmoid function. An alternative way is to leave out the product term in Eq. (3), assuming that it is small and can be neglected in comparison to the other two additive terms. This permits the replacement of the probabilities in Eq. (3) by the likelihoods with a negligible error. The two methods have been found to produce almost identical results (Ming and Smith, 1999, 2000, 2001). In the latter case, the union-based observation probability $B_i(o_{\vee_M})$ is approximated as

$$B_i(o_{\vee_M})=\sum_{n_1n_2\cdots n_{N-M}} b_i(o_{n_1})\cdot b_i(o_{n_2})\cdots b_i(o_{n_{N-M}}), \tag{6}$$

when the summation is over all possible combinations of N different values $(1,2,...,N)$ taken $(N-M)$ at a time.

III. EXPERIMENTS PERFORMED FOR ALL ORDERS IN THE UNION MODEL

In order to show the influence of the order on the performance of the union model, in this section we first present the experiments performed for all possible orders. The experiments in this paper are performed on the TIDIGITS database, unlike Ming and Smith (1999, 2000, 2001), who tested the union model on an E-set database.

A. The experimental conditions

The TIDIGITS database (Leonard, 1984) contains the speech data of connected digit sequences from 225 adult speakers (111 male and 114 female) for speaker-independent recognition. From this database the isolated-digit parts were extracted for the tests. This includes 11 isolated-digit words: ''one''–''nine,'' ''zero,'' and ''oh,'' each digit surrounded by silence parts. This database was subdivided into training and testing sets based on the recommendations by NIST. The training set contains 1232 utterances from 112 speakers, 55 male and 57 female, each speaker producing one token for each of the eleven digits; the test set contains 2486 utterances from 113 speakers, 56 male and 57 female, each speaker contributing with two utterances of each digit.

The speech signal, sampled at 8 kHz, was divided into frames of 30 ms, with an overlap 10 ms between frames. The pre-emphasis was used and the Hamming window applied on each frame. For each frame, a multichannel, Mel-scaled filter bank analysis with 35 channels is used to estimate the log-amplitude spectra of the speech. These filter channels are then grouped uniformly into five sub-bands, each sub-band consisting of information from seven channels. A DCT is applied within each sub-band and the first four MFCCs coefficients form the sub-band feature vector. In order to include dynamic spectral information, the first-order delta parameters were calculated and added to each sub-band feature vector. The probabilities of the individual sub-band feature streams are then combined at the frame level using the probabilistic union model. A 12-state HMM is estimated for each digit, with the first and the last states being tied among all the digits to account for the silence parts of each utterance at the beginning and end. For comparison, we also implemented a full-band model, which used 18 MFCCs plus 18 delta MFCCs for each frame calculated across the full spectrum band of the speech.

TABLE I. The narrow-band noises, with a bandwidth of 100 Hz and different central frequencies, used in the experiments, with the corresponding numbers of affected sub-bands.

Noise central frequency (Hz)	Number of affected sub-bands
600 1000 2100 2800	1
1500 600;1000 600;2100 600;2800 1000;1500 1000;2100 1000;2800 1500;2100 2100;2800	2
600;1500 1500;2800	3

TABLE II. Recognition accuracy results by union model of different orders with five streams, and by the full-band model (static+delta/delta), assuming that the noise exists only in the speech part.

SNR (dB)	Number of affected bands	Union model with order M					Full-band model
		0	1	2	3	4	
Clean	0	99.2	99.1	98.4	96.1	83.5	98.0/95.5
10	1	88.3	97.4	96.5	91.3	75.6	88.2/88.4
	2	75.6	89.6	93.6	88.3	72.0	80.6/81.8
	3	72.3	79.0	83.1	83.9	69.1	74.1/78.0
0	1	67.7	91.7	89.4	82.2	67.2	68.2/79.1
	2	50.7	69.9	83.0	77.4	62.2	55.8/66.6
	3	52.2	63.0	70.1	77.8	62.8	50.6/61.6

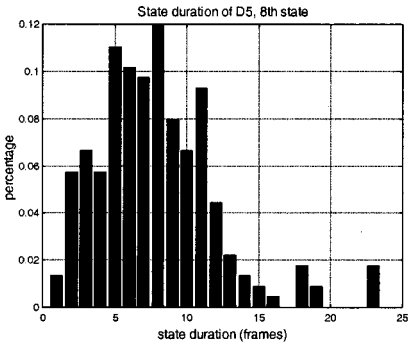
The training was performed on clean utterances from the training set. In recognition the testing set was corrupted by narrow-band noise of different band positions. The noise component was additive, which was generated by passing the Gaussian white noise through a bandpass filter. The 3 dB cutoff bandwidth of the filter was fixed at 100 Hz and the central frequencies were varied across the sub-bands. In particular, five different central frequencies are chosen, which are 600, 1000, 1500, 2100, and 2800 Hz, respectively. The noise, consisting of *one* or *two* such narrow-band components with different central frequencies, was added to each test utterance, resulting in the utterances with various numbers of corrupted sub-bands. The noises used in the experiments with corresponding numbers of affected sub-bands are shown in Table I (for example, the noise with a central frequency of 1500 Hz was located around the border of two adjacent sub-bands thereby affecting both bands). Initially, we assumed that the noise exists only in the speech part within each utterance. Later on we will look at a more realistic situation where the noise lasts across the whole signal (i.e., affecting both the speech and surrounding silence parts). The SNR is calculated based on the averaged energy of the speech parts of all the test utterances, so that the noise has a constant loudness in all the utterances, regardless of the actual energy of the speech in each specific utterance. The experiments are performed for SNR=10 dB and SNR=0 dB, respectively.

B. Tests of the union model with all possible orders

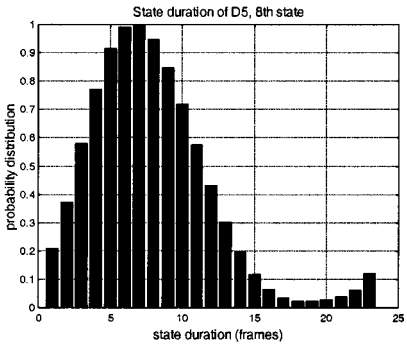
Based on Eq. (4), for a system with N sub-bands, recognition can be performed with different orders (i.e., M) from 0 to $N-1$. Table II presents the recognition results for all the orders (i.e., $M=0,...,4$) within our five-band (i.e., $N=5$) system. For comparison, Table II also includes the results achieved by the full-band model using both static and delta parameters and using only delta parameters (i.e., assuming that the static parameters are strongly affected by noise), respectively. The results are shown for both the clean and noisy speech, as a function of the SNR and the number of corrupted bands. From Table II we can see the influence of the order on the recognition accuracy.

First, we look at the recognition results for clean speech. Because there is no band corruption, every sub-band provides useful information. Therefore the union model with order $M=0$, which takes a full conjunction of the sub-band observations, produced the best result. As shown in Table II, as the order increases, the performance for clean speech recognition decreases. This is because a higher-order model bases the recognition on the conjunction of fewer sub-bands, which carry less discriminative information. For example, the model with order $M=2$, containing conjunctions of three sub-bands, produced better results than the model with order $M=3$, which is based on conjunctions of only two sub-bands. The union model with order $M=4$, containing conjunctions of only single bands (each covers about 20% of the entire frequency band), achieved an accuracy of only 83.5%, which is significantly lower than 99.2% achieved by the model with a matched order (i.e., $M=0$).

Next, we look at the recognition results for noisy speech. The second column of Table II shows the number of noisy sub-bands in the speech signal. As indicated in Table II, the union model achieved its maximum performance when the model order matched the number of noisy bands. This is found to be true for all the cases with one, two, and three noisy sub-bands and with different SNRs. As explained earlier, this match causes the union model to include a conjunction of all the remaining $(N-M)$ clean bands, thereby catching maximum discriminative information. The improvement in performance due to this match can be very significant, for example in Table II, in the case with two corrupted bands and SNR=0 dB, the union model with a matched order $M=2$ achieved an accuracy of 83.0%; but the model with *a lower* order, e.g., $M=1$, and *a higher* order, e.g., $M=3$, only obtained an accuracy of 69.9% and 77.4%, respectively. Table II also indicates that in the case of no knowledge about the noisy bands, the model with a higher order (e.g., $M=2$, or 3) can provide higher robustness against the uncertainty of the number of noisy bands than a model with a lower order (e.g., $M=1$). This principle has been exploited previously for the selection of a balanced fixed order for the union model; see (Ming and Smith, 1999, 2000, 2001). Our objective in this paper is to develop an algorithm to automatically



(a)



(b)

FIG. 1. Histogram of the state duration of the eighth state for digit "five" (a) and the corresponding probability density obtained by the polynomial approximation (b).

select the order for the union model, to improve upon the previous balanced fixed-order selection.

IV. AN ALGORITHM FOR AUTOMATIC ORDER SELECTION IN THE UNION MODELS

The new algorithm for automatic order selection is based on the state duration patterns generated by the HMM within which the union model is incorporated.

As explained earlier, when the union model has an order lower than the matched order, some noisy bands are included in each conjunction for recognition. On the other hand, if the model uses an order higher than the matched order, some discriminative information may be lost due to the average between the clean sub-band probabilities. Only the match can capture the maximum correct information. Therefore, it is reasonable to assume that the state duration pattern produced by the matched model should be the most similar to the duration pattern obtained for clean speech. So in recognition we can run the union model for a range of orders in parallel, and decide the best order by comparing the state duration pattern produced by each of these models with the pattern obtained from the training stage for clean speech utterances. The following gives the details of this algorithm.

A. Order selection algorithm

Assume that the state duration pattern of the HMM for each vocabulary word w is expressed by the probability density functions $p_i^w(d)$'s, where d represents the occupancy count and i the state, $i=1,...,I$. These functions are assumed to be estimated in the training stage based on clean speech data. In recognition, we are given an utterance with an unknown number of corrupted bands. First, we perform the recognition for a range of different orders. For each order, we obtain a recognition result, together with the corresponding counts of occupancy of each state. Assuming that L orders are used, then we have L recognized words $w(l)$ and L state duration counts $D_i(l)$ for state i, $l=1,...,L$, respectively. The best order, l^*, is defined as the order whose duration pattern has the maximum probability, i.e.,

$$l^* = \arg\max_{1\leq l\leq L} \prod_{i=1}^{I} p_i^{w(l)}(D_i(l)). \tag{7}$$

Based on Eq. (7), we obtain the recognition result $w(l^*)$. In the following we describe the estimation of the duration probability density functions, $p_i^w(d)$'s.

B. Modeling of state duration pattern

After training the HMMs on clean speech utterances, for each utterance we can obtain the state occupancy count for each state. Then, these counts based on the whole training set lead to a histogram $f_i^w(d)$, for the occupancy of each state i for each vocabulary word w. These state duration histograms are then used to derive the state duration probability density functions, $p_i^w(d)$'s. For example, the $p_i^w(d)$ can be expressed as a Gaussian, a gamma, or a polynomial function to fit the histogram $f_i^w(d)$ (e.g., Russell and Cook, 1987; Gu, Tseng, and Lee, 1991; Vaseghi, 1995; Burshtein, 1996). In this paper we use a sixth-order polynomial function for each $p_i^w(d)$. An example of a state duration histogram and the corresponding probability density is shown in Fig. 1.

The above model considers the duration pattern within each individual state. This can be modified by considering the duration pattern within several adjacent states. In particular, we consider the use of the duration probability density $p_{i(\Delta)}^w(d)$, where $i(\Delta)$ represents all states within the range $[i-\Delta, i]$ defined by Δ. We have tested different values for Δ and found that this model can be more robust against inaccuracy in the frame-state alignment than the above single-state duration model, for a range of Δ values. We call this the multistate duration model. Both the single-state and multistate duration models have been used in our experiments.

V. RECOGNITION RESULTS USING AN AUTOMATIC ORDER-SELECTION ALGORITHM

A. Tests with stationary narrow-band noise existing only in the speech part

First, we test the automatic order-selection algorithm under the same conditions as described in Sec. III B. We test the single-state duration model and the multistate duration model, respectively, for the order selection.

Table III presents the recognition results obtained by using the multistate duration model with the parameter Δ ranging from 0 to 3. Note that when $\Delta=0$ this is equivalent to a single-state duration model. The algorithm searches for the best order from the orders (0,1,2,3). The order $M=4$ was excluded from the search because it provides poor recognition accuracy for clean speech recognition (see Table II) and thus is not considered to be useful to the system for speech recognition. (However, the experimental results indicate that a search including order 4 causes only very little loss of

TABLE III. Recognition accuracy results by the union model with five streams based on automatic order selection, as a function of the state-range parameter Δ, with a comparison to the matched order, assuming that the noise exists only in the speech part.

SNR (dB)	Number of affected bands	Union model				
		Matched order	Automatic order algorithm			
			Δ=0	Δ=1	Δ=2	Δ=3
Clean	0	99.2	99.4	99.1	98.9	98.8
10	1	97.4	96.6	97.0	96.6	96.0
	2	93.6	92.2	93.6	93.1	91.7
	3	83.9	85.2	86.9	86.7	85.5
0	1	91.7	88.8	91.2	90.5	89.1
	2	83.0	80.0	83.9	83.6	81.4
	3	77.8	75.3	78.0	77.8	75.0

recognition accuracy.) For comparison, Table III also includes the highest recognition accuracy in Table II, which is obtained by using an order matching the noise condition.

Table III indicates that the automatic order-selection algorithm can produce equally good results as the order-matched model, for all the Δ values. We can see that in some cases, with or without noise, the automatic algorithm can produce a recognition accuracy that is higher than that obtained by the order-matched model. This is because this algorithm can also take some information from the noisy bands, as long as the SNR is high. In the case of noisy speech, we can see that the multistate duration model with $\Delta=1$ and $\Delta=2$ outperformed the single-state duration model (i.e., $\Delta=0$). However, the recognition performance becomes worse when $\Delta>2$, due to excessive averaging over the state duration. Based on these results, the following experiments were performed by using $\Delta=1$.

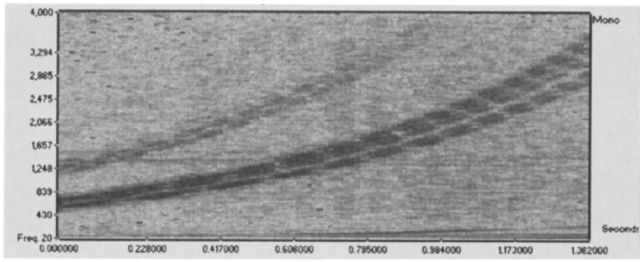
B. Tests with stationary narrow-band noise existing in both the speech and surrounding silence parts

In the previous experiments we have assumed that the noise exists only in the speech part. However, in real applications, this information may not be available. Therefore, in the following experiments noise affects all parts of the signal, including the speech part and the surrounding silence parts. The SNR is calculated in the same way as in the previous experiments. The results are presented in Table IV. As can be seen, the automatically selected order produced similar or better results in comparison to the matched order. However, compared to Table III, the recognition performance becomes poorer. This may be because if the noise is not real band-limited (as in our case), its energy can spread over a wide range of frequencies and therefore may dominate in all the sub-bands for the silence parts without a speech signal. This problem may be reduced by combining the static and dynamic features using the union model principle, as will be described below.

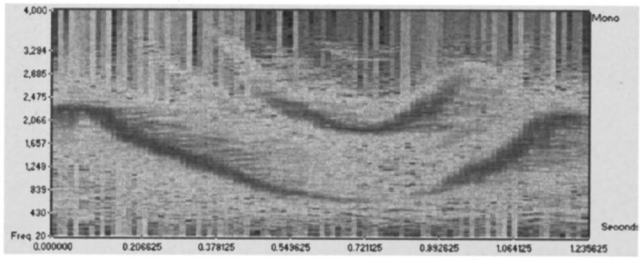
TABLE IV. Recognition accuracy results by the union model with five streams based on automatic order selection with a comparison to the matched order, assuming that the noise exists in both the speech and the surrounding silence parts.

SNR (dB)	Number of affected bands	Union model	
		Matched order	Automatic order algorithm, Δ=1
10	1	76.7	80.4
	2	63.3	67.5
	3	61.5	64.1
0	1	29.7	31.8
	2	30.8	30.2
	3	29.0	31.3

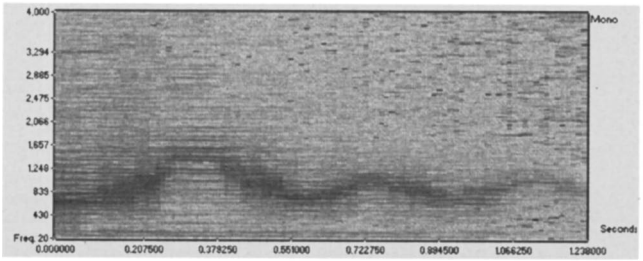
In the system used above, the probabilities associated with the static and dynamic features in each sub-band are combined using a product. Therefore if one stream is corrupted by noise the entire probability will be affected. In the above situation, we can assume that the static parameters of all the sub-bands for the silence parts will be strongly affected by the noise, but the dynamic parameters (i.e., delta features) may be affected less, as they are more robust for slowly changing noise. Therefore, combining the static and delta parameters using the union model rather than the product model should increase the robustness for dealing with the



(a) Noise 1



(b) Noise 2



(c) Noise 3

FIG. 2. Spectrograms of nonstationary real-world noise used in the experiments.

TABLE V. Recognition accuracy results by the union model with 10 streams based on automatic order selection with comparison to the matched order and full-band model (static+delta/delta), assuming that the noise exists in both the speech and the surrounding silence parts.

SNR (dB)	Number of affected bands	Union model: Matched order	Union model: Automatic order algorithm, $\Delta=1$	Full-band model
Clean	0	99.4	99.3	98.0/95.5
10	1	91.5	93.6	21.6/84.8
	2	84.2	87.9	14.9/76.2
	3	79.8	85.6	12.5/72.3
0	1	78.1	75.5	16.7/75.3
	2	63.2	62.4	11.0/62.9
	3	55.9	58.0	10.9/57.4

noise in the silence parts. In our recognition system with five sub-bands this leads into ten streams (i.e., five static and five dynamic), combined using the probabilistic union model. This is used for recognizing the noisy utterances described above, with noise lasting over the whole signal. The system uses the automatic order-selection algorithm and searches for the best order among the orders from 0 to 6. The recognition results are presented in Table V. Comparing Table V to Table IV, we can see significant improvement in the recognition performance. For example, in the case with one noisy band and SNR=10 dB, the recognition performance is improved from 80.4% to 93.6%, using the automatic order selection algorithm.

C. Tests with nonstationary real-world noise

Finally, we performed experiments for some nonstationary real-world noises, which have a dominant frequency-localization characteristic. The noise data used in these experiments are sound files obtained from program "ICQ," a largely used Internet tool that enables the user for easy online contact, chat, send messages, files and URLs. The spectrograms of these sounds are shown in Fig. 2, in which a Noise 1, Noise 2, and Noise 3 refers to the sounds produced during establishing a connection, maintaining a connection and a request for chat, respectively. As can be seen, all of the noises are highly nonstationary. Each of these noises was added to the whole utterance (i.e., including the silence parts). Table VI presents the recognition results. We can see that the results achieved by the automatic order-selection algorithm are similar to the best order results selected from the orders from 0 to 6.

TABLE VI. Recognition accuracy results for real-world noise by the union model with ten streams based on automatic order selection with comparison to the best order and full-band model (static+delta/delta).

Type of noisy speech	Union model: The best order	Union model: Automatic order algorithm, $\Delta=1$	Full-band model
Clean	99.4	99.3	98.0/95.5
Noise 1	70.2	65.2	18.2/66.4
Noise 2	85.3	88.8	20.8/36.2
Noise 3	66.0	62.7	21.9/69.9

D. Which orders did the automatic order algorithm select?

As was described earlier, the automatic order-selection algorithm selects the model order on a per utterance basis, i.e., the selected order may differ from one utterance to another, given the same type of noise. Figure 3 presents the distribution of the selected orders by the automatic order algorithm for some of the noisy speech utterances described in Sec. V A. Figure 3 indicates that the algorithm selected, with the highest frequency, the order that gives the best recognition accuracy, i.e., the matched order. However, we can also see that the automatic order algorithm also selected some orders different from the matched order, for example, order 0 for two noisy sub-bands with a frequency of 12%, as shown in Fig. 3(c).

VI. CONCLUSION

In this paper we described an algorithm for automatic order selection in the probabilistic union model for noisy speech recognition. The probabilistic union model is an al-

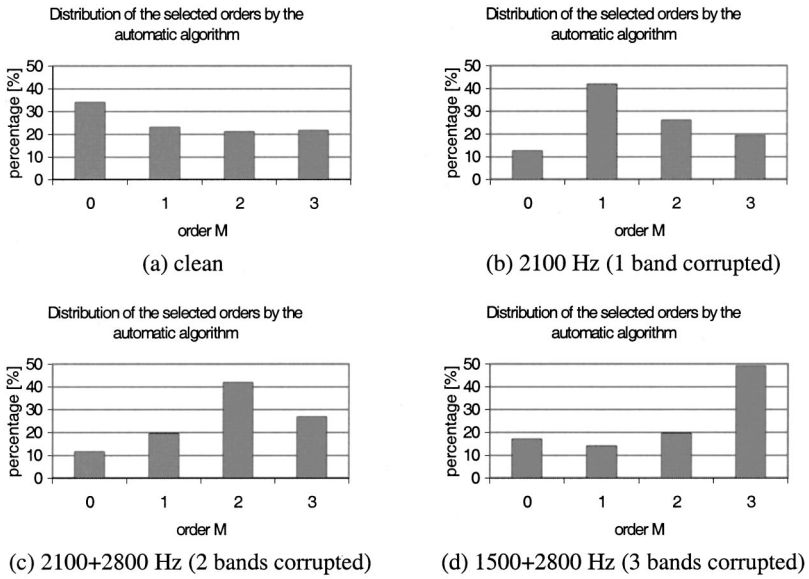


FIG. 3. Histogram of the automatically selected orders versus the number of corrupted bands.

ternative statistical approach for combining sub-band observations for dealing with band-limited noise, which assumes no knowledge about the band position of the noise. However, the number of bands affected by noise is in the union model equal to the order of the model and its choice is critical. In this paper we have introduced an automatic algorithm for order selection. This algorithm selects the order based on the state duration pattern generated by the HMM built on the union model. The computational cost for the selection is similar to a normal *n*-best rescoring algorithm. Experiments have been conducted on the speaker-independent TIDIGITS database, corrupted by various frequency-selective noises. The results have shown that the union model based on the order selected by the automatic algorithm has achieved a recognition performance similar to that achieved by using the matched-order (i.e., when the number of noisy bands is known *a priori*). In the case of isolated digit recognition with various types of band-limited noise, the resulting extended union model gives very significant improvements over the full-band model.

The principle of the algorithm, i.e., using the state duration for order selection in the union model, can be applied more generally in the sub-band recognition system for the detection of sub-bands of signals corrupted by noise. Other applications of the probabilistic union model, for example, for speech recognition subjected to partial temporal duration corruption, can be found in Ming, Stewart, Hanna, and Smith (1999); Ming, Jančovič, Hanna, Stewart, and Smith (2000).

ACKNOWLEDGMENTS

This work was supported by UK EPSRC Grant No. GR/M93734. The authors thank the three reviewers for their helpful comments.

Bourlard, H. (**1999**). ''Non-stationary multi-channel (multi-stream) processing towards robust and adaptive ASR,'' in *Proceedings of the Workshop on Robust Methods for Speech Recognition in Adverse Conditions*, Tampere, Finland, pp. 1–10.

Bourlard, H., and Dupont, S. (**1996**). ''A new ASR approach based on independent processing and recombination of partial frequency bands,'' in *Proceedings of the International Conference on Spoken Language Processing*, Philadelphia, pp. 426–429.

Bourlard, H., and Dupont, S. (**1997**). ''Sub-band based speech recognition,'' in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, Munich, Germany, pp. 1251–1254.

Burshtein, D. (**1996**). ''Robust parametric modeling of durations in hidden Markov models,'' IEEE Trans. Speech Audio Process. **4**, No. 3, 240–242.

Cerisara, C., Haton, J.-P., Mari, J.-F., and Fohr, D. (**1998**). ''A recombination model for multi-band speech recognition,'' in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, Seattle, pp. 717–720.

Cooke, M., Morris, A., and Green, P. (**1997**). ''Missing data techniques for robust speech recognition,'' in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, Munich, Germany, pp. 803–806.

Gu, H., Tseng, Ch., and Lee, L. (**1991**). ''Isolated-utterance speech recognition using hidden markov models with bounded state durations,'' IEEE Trans. Signal Process. **39**, 1743–1751.

Hermansky, H., Tibrewala, S., and Pavel, M. (**1996**). ''Towards ASR on partially corrupted speech,'' in *Proceedings of the International Conference on Spoken Language Processing*, Philadelphia, pp. 462–465.

Leonard, R. G. (**1984**). ''A database for speaker-independent digit recognition,'' in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, San Diego, California, pp. 42.11/1–4.

Lippman, R. P., and Carlson, B. A. (**1997**). ''Using missing feature theory to actively select features for robust speech recognition with interruptions, filtering, and noise,'' in *Proceedings of Eurospeech*, Rhodes, Greece, pp. KN37–40.

Ming, J., Jančovič, P., Hanna, P., Stewart, D., and Smith, F. J. (**2000**). ''Robust feature selection using probabilistic union models,'' in *Proceedings of the International Conference on Spoken Language Processing*, Beijing, China, pp. 546–549.

Ming, J., and Smith, F. J. (**1999**). ''Union: a new approach for combining sub-band observations for noisy speech recognition,'' in *Proceedings of the Workshop on Robust Methods for Speech Recognition in Adverse Conditions*, Tampere, Finland, pp. 175–178.

Ming, J., and Smith, F. J. (**2000**). ''A probabilistic union model for sub-band based noisy speech recognition,'' in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, Istanbul, Turkey, pp. 1787–1790.

Ming, J., and Smith, F. J. (**2001**). ''Union: A new approach for combining sub-band observations for noisy speech recognition,'' Speech Commun. **34**, Issue 1-2, 41–55.

Ming, J., Stewart, D., Hanna, P., and Smith, F. J. (**1999**). ''A probabilistic union model for partial and temporal corruption of speech,'' in *Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding*, Keystone, Colorado, pp. 43–46.

Morris, A. C., Hagen, A., and Bourlard, H. (**1999**). ''The full-combination sub-bands approach to noise robust HMM/ANN based ASR,'' in *Proceedings of Eurospeech*, Budapest, Hungary, pp. 599–602.

Okawa, S., Eorico, B., and Potamianos, A. (**1998**). ''Multi-band speech recognition in noisy environments,'' in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, Seattle, pp. 641–644.

Russell, M. J., and Cook, A. E. (**1987**). ''Experimental evaluation of duration modeling techniques for automatic speech recognition,'' in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, Dallas, Texas, pp. 2376–2379.

Tibrewala, S., and Hermansky, H. (**1997a**). ''Sub-band based recognition of noisy speech,'' in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, Munich, Germany, pp. 1255–1258.

Tibrewala, S., and Hermansky, H. (**1997b**). ''Multi-band and adaptation approaches to robust speech recognition,'' in *Proceedings of Eurospeech*, Rhodes, Greece, pp. 2619–2622.

Vaseghi, S. V. (**1995**). ''State duration modeling in hidden Markov models,'' Signal Process. **41**, 31–41.

Effects of relative phases on pitch and timbre in the piano bass range

Alexander Galembo
Setchenov Institute of Evolutionary Physiology and Biochemistry, Russian Academy of Sciences, Thorez Prospect 44, St. Petersburg 194223, Russia

Anders Askenfelt[a)]
Department of Speech, Music and Hearing, Royal Institute of Technology, SE-100 44 Stockholm, Sweden

Lola L. Cuddy and Frank A. Russo
Department of Psychology, Queen's University, Kingston, Ontario K7L 3N6, Canada

(Received 15 December 1999; accepted for publication 21 June 2001)

Piano bass tones raise questions related to the perception of multicomponent, inharmonic tones. In this study, the influence of the relative phases among partials on pitch and timbre was investigated for synthesized bass tones with piano-like inharmonicity. Three sets of bass tones (A0 = 27.5 Hz, 100 partials, flat spectral envelope) were generated; harmonic, low inharmonic, and high inharmonic. For each set, five starting phase relations among partials were applied; sine phases, alternate (sine/cosine) phases, random phases, Schroeder phases, and negative Schroeder phases. The pitch and timbre of the tones were influenced markedly by the starting phases. Listening tests showed that listeners are able to discriminate between tones having different starting phase relations, and also that the pitch could be changed by manipulating the relative phases (octave, fifth, major third). A piano-like inharmonicity gives a characteristic randomizing effect of the phase relations over time in tones starting with nonrandom phase relations. A measure of the regularity of the phase differences between adjacent partials is suggested for quantifying this randomization process. The observed phase effects might be of importance in synthesizing, recording, and reproducing piano music. [DOI: 10.1121/1.1391246]

PACS numbers: 43.75.Cd, 43.75.Mn, 43.66.Hg [ADP]

I. INTRODUCTION

Piano bass tones connect instrument design and psychoacoustics in a challenging way. In the bass, the piano design is forced to its limits, and the piano engineer encounters a variety of problems. The produced sound defines a stimulus of extraordinary complexity, of which many aspects have not yet been explored by the psychoacoustician.

A. The piano designer's problem

Differences in tone quality between pianos of different size, model, and make are due to differences in design, material, and the production process. This truism is best illustrated in the bass range. Above the mid-range (say C4), stringing scales of most pianos are rather similar. The speaking length of the strings is derived iteratively from the length of the C8 string (which always is close to 5 cm), via a scaling factor, which always is close to the 12th root of two [see, e.g., Conklin (1996)]. Design limitations, however, become clearly exposed in the bass range, in particular, in the two lowest octaves A0–A1 covering 27.5–110 Hz [Galembo and Cuddy (1997); Galembo, Askenfelt, and Cuddy (1998)]. Some of the most important difficulties which the piano designer has to tackle in the bass are: (1) the perceptual importance of inharmonicity due to string stiffness increases; (2) the longitudinal string modes become a significant factor of the timbre [Conklin (1996)]; (3) due to physical space limitations it is no longer possible to use plain strings and wrapping is introduced that influences the tone quality; (4) an inevitable discontinuity in the string scale and corresponding change in timbre occurs at the transition from the treble bridge to the bass bridge; and (5) the radiation efficiency of the soundboard decreases rapidly in the lowest bass.

Mainly by trial and error, piano designers have developed a variety of strategies to overcome or compensate for these limitations in order to achieve the best tone quality. Further refinements of the design may still be possible. A deeper understanding of the physical processes involved in tone generation in the bass range, as well as their perceptual relevance, are thus of immediate interest to the piano manufacturer.

B. Psychoacoustic perspective

From a psychoacoustician's point of view the piano bass range brings several questions to the surface, in particular, how perception is influenced by the temporal evolution in magnitude and phase over the duration of an inharmonic tone.

A piano bass tone is an intriguing stimulus [see Figs. 1(a) and (b)]. Viewed in the frequency domain, it is very rich. The spectra of the lowest bass tones may contain more than 100 partials, and generally extend up to 4–5 kHz within a 60-dB amplitude drop. The magnitude spectrum has a characteristic formant-like envelope with groups of six to eight partials between spectral minima, determined by the striking

a)Electronic mail: andersa@speech.kth.se

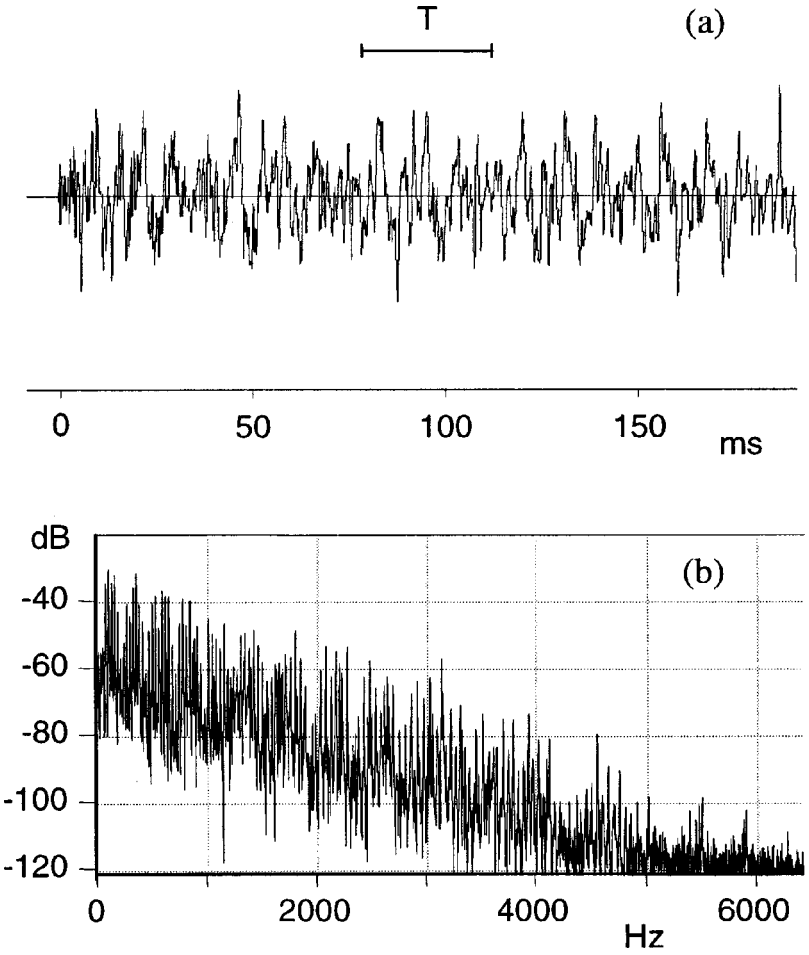


FIG. 1. (a) Waveform of the initial 200 ms of a grand piano bass tone (*Steinway C*, 224 cm, A0=27.5 Hz), recorded by a microphone at a distance of about 1 m. The fundamental period *T* is indicated by a horizontal bar line. Note that the peak structure in the waveform gives no clear evidence of the period time because of aperiodicity due to string stiffness and a weakly radiated fundamental. (b) Spectrum of the same note showing more than 100 partials extending to 4–5 kHz. The spectrum envelope has a characteristic formant-like shape with groups of about eight partials between spectral minima, determined by the striking position of the hammer. The partials are progressively stretched due to string inharmonicity.

position of the hammer on the string. The partials are progressively stretched due to the inherent inharmonicity of the stiff strings [Rayleigh (1894); Fletcher (1964)], as well as a motion of the string terminations [Schuck and Young (1943); Rimski-Korsakov and Diakonov (1952); Exley (1969); Benade (1976)]. The decay rate generally increases toward higher partials [Martin (1947); Hundley, Benioff and Martin (1978)].

In the time domain, the inharmonicity is manifested as aperiodicity, giving a marked evolution in the waveform when viewed over several periods. Among other things, this means that the phase relations among partials (''relative phases'') change continuously. The presence of aperiodicity suggests that temporal processes might be of perceptual importance in the bass range. Psychoacoustical findings support this view; for complex tones in the bass range, no partials are resolved by critical bands [see, e.g., Gelfand (1998)]. As a consequence, temporal pitch coding can be supposed to dominate. Further, as we deal with a high number of audible partials, perceptual effects on pitch and timbre from the relative phases among partials seem quite probable. In particular, it is natural to suppose that *audible effects due to the dynamically changing relative phases in piano bass tones may occur*.

This is the topic addressed in this study. The perception of synthesized multicomponent bass tones with a piano-like inharmonicity is explored in three experiments. The design of the experiments focused on two main control parameters: (1) the relative phases at note onset (starting phases), and (2) the level of inharmonicity, and its influence on stimulus waveforms, pitch and timbre. The starting phases will control the initial waveform (peak structure), while the inharmonicity will influence the temporal evolution of the phase relations, and hence also the evolution of the waveform.

II. BACKGROUND

A. Phase effects

The classic statement [Helmholtz (1877)] that the auditory system is insensitive to the phase relations among partials is known to be a first-order approximation only. Several studies have demonstrated an influence of relative phases of partials on timbre and pitch perception in experimental situations [see, e.g., Schroeder (1975); Nordmark (1978); Patterson (1987); Moore (1997); Hartmann (1998)]. However, the reported experiments usually involve synthesized complex tones consisting of a few harmonics, restricted to a specific range of fundamental frequencies. In contrast, tones produced by real musical instruments usually contain numerous partials and show a complex dynamic evolution of the spectral envelope. In many cases, such as the piano, musical tones also have a noticeable inharmonicity. In view of these differences, it is not straightforward to apply the psychoacoustical findings on the perception of musical tones. Phase

effects found under specific experimental conditions with relatively simple stimuli may not be noticeable in tones produced by real musical instruments.

The bass range, roughly defined as tones with fundamentals below 100–150 Hz, has typically been outside the scope of most psychoacoustical experiments. This situation is unfortunate since the low-frequency range, where temporal processes are known to dominate the auditory analysis, seems to provide a promising field for studying the effects of relative phase on pitch and timbre perception. In fact, one of the few reported studies including stimuli below 100 Hz shows convincingly that phase effects on timbre are strongest in the bass range [Patterson (1987)].

A primary aim of most psychoacoustical experiments on phase effects in complex tones has been to model the pitch and timbre perception mechanisms, particularly to argue for or against temporal versus spectral models of hearing [Nordmark (1978); Moore (1997)]. Attempts to apply the results to the perception of pitch and timbre of tones of musical instruments are scarce [Mathes and Miller (1947); Hansen and Madsen (1974)]. However, some important results from psychoacoustical experiments on phase relations among partials are clearly of relevance for some aspects of perception of musical tones dealt with in this study. These relate, in particular, to two features of the auditory system.

(1) *Selective attention to a single partial or a spectral range of a tone*. It is well known that manipulation of the *magnitude* spectrum may draw attention to an individual partial or group of partials and change the judgement of the pitch of the tone [Schouten (1940)]. Manipulation of the *phase* spectrum can give similar results [see, e.g., Moore (1977); Hartmann (1988); Shackleton and Carlyon (1994)].

(2) *Influence on pitch strength*. When the partials of a complex harmonic tone are in phase, the fundamental periodicity in the waveform is enhanced (peak structure), and a pitch corresponding to the fundamental frequency is promoted. The *pitch strength*, which is a measure of how clearly the pitch is perceived [Fastl and Stoll (1979)], is expected to be higher for a tone with in-phase relations among partials than for other combinations of phases [see, e.g., Mathes and Miller (1947); Licklider (1957)].

B. Inharmonicity

Inharmonicity is a common term in psychoacoustics referring to spectra in which one or several partials are shifted away from their harmonic positions. The purpose of most experiments with such controlled inharmonicity has been to gain a better understanding of the perception of *harmonic* complex tones [see, e.g., Schouten, Ritsma, and Cardozo (1962); Terhardt (1972); Goldstein (1973); Wightman (1973a); Moore, Glasberg, and Peters (1985)]. Moore, Peters, and Glasberg (1985) connected the results of psychoacoustical experiments on inharmonicity with the perception of piano tones.

The main feature of an inharmonic musical tone is that it lacks strict periodicity. This property may affect both pitch and timbre. In particular, the inharmonicity may influence the pitch strength. A harmonic spectrum is known to evoke a more pronounced sensation of pitch than does an inharmonic spectrum [Wightman (1973b)].

The inharmonicity in piano tones is caused by several physical phenomenon, influencing all partial frequencies. A major contribution comes from the stiffness of the piano string, which causes its transverse modes of vibration to be higher than those of an ideal string (''stretched'' partials). The classical formula describing the inharmonicity in piano strings is

$$f_n = n f_1^0 \sqrt{1 + n^2 B}, \tag{1}$$

where f_n is frequency of the nth partial, n is partial number, f_1^0 is fundamental frequency of the ideal, flexible string, and B is an inharmonicity coefficient, set by the string dimensions, material, and design. Fletcher (1964) provides a good survey of earlier work together with derivations of Eq. (1) for different types of boundary conditions and experimental validations.

In the time domain, the inharmonicity in piano tones is manifested as dispersion, the high-frequency components propagating on the string with a higher velocity than the lower components. This leads to characteristic ripples in the string waveform, preceding each main fundamental pulse, ''pitch glides'' [Podlesak and Lee (1989)] or ''precursors'' [Askenfelt and Jansson (1993)]. The precursors are particularly prominent in the bass where the excitation pulse is short compared to the fundamental period. The influence of inharmonicity in piano tones on timbre is not at all fully explored. Perceptual studies of synthesized piano-like tones have suggested that the inharmonicity adds ''warmth'' or ''liveness'' to the decay [Fletcher, Blackham, and Stratton (1962)], as well as a ''bite'' to the attack, enhancing the percussive character [Reinhold, Jansson, and Askenfelt (1987)].

C. Overlooked factors

A serious limitation of earlier works on the perception of piano tones is that two important factors influencing the timbre in combination with the inharmonicity have been overlooked in the experiments. The first factor is the *spectral bandwidth* of a tone, which covaries with the degree of inharmonicity and influences the brightness (sharpness) markedly [Galembo and Cuddy (1997)]. For example, a threefold increase of the inharmonicity coefficient (a factor that well may be observed between two grand pianos of different size) may increase the spectral bandwidth of a bass tone by as much as 50%.

The second overlooked factor influencing piano timbre as well as pitch is the *phase relations* among partials. This factor is the focus of the three experiments included in this study. In Experiment I, the listeners' ability to discriminate between tones with five different starting phase relations under three levels of inharmonicity is investigated. In Experiment II, we test specific pitch effects in harmonic tones (ambiguous or alternative pitches). In Experiment III, temporary timbral effects in inharmonic tones (''squeals'') are examined.

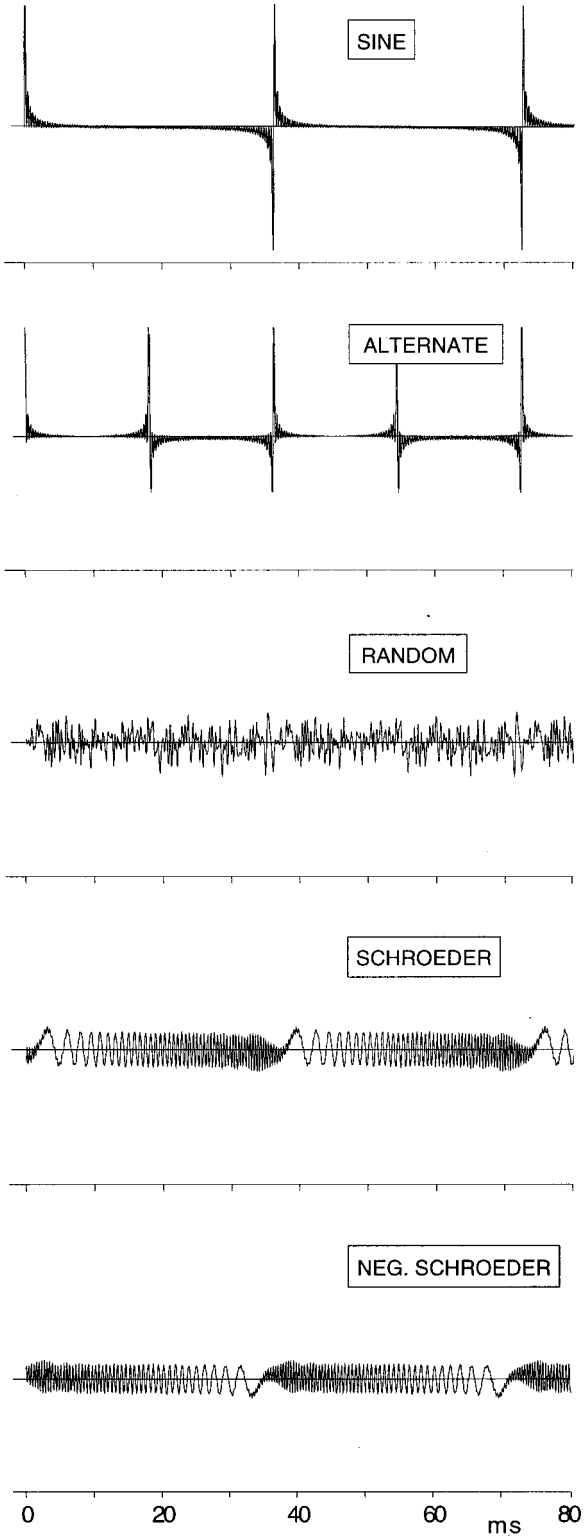


FIG. 2. Computed waveforms of five harmonic A0 tones ($f_1^0 = 27.5$ Hz) consisting of 100 consecutive harmonics of equal and constant amplitude, but with different starting phases.

III. STIMULI

A total of 15 stimulus tones were synthesized, comprising five different starting phase relations and three levels of inharmonicity. Different subsets of these tones were used in Experiments I–III.

The tones were computed as

$$tone(n,t) = \sum_{n=1}^{N} A \sin[2\pi f_1^0 n(1+Bn^2)^{0.5} t + Ph(n,0)], \tag{2}$$

where N is the number of partials in the spectrum, $Ph(n,0)$ is the starting phase of the nth partial, and B is the inharmonicity coefficient ($B = 0$, 60×10^{-6}, or 600×10^{-6}).

All tones had a steady spectrum with a flat envelope, and a common frequency of the (unstretched) fundamental, $f_1^0 = 27.5$ or 55.0 Hz, corresponding to the nominal frequency of the lowest note on the piano (A0), or an octave above (A1). Each tone consisted of 100 partials. This number is representative of a real piano bass tone at a relatively loud dynamic level. The tones were generated by additive synthesis on a Pentium PC at a sampling frequency of 16 kHz.[1]

The inharmonicity was implemented according to Eq. (1), mimicking the partial frequencies of real piano tones. Three levels of inharmonicity were used, defining three sets of tones; "harmonic" ($B=0$), "low inharmonic" ($B=60\times10^{-6}$), and "high inharmonic" ($B=600\times10^{-6}$). For the low-inharmonic set, the inharmonicity coefficient B was chosen close to the minimum value in the bass range across many observed pianos, [Galembo and Askenfelt (1999); Galembo and Cuddy (1997)]. For the high-inharmonic set, B was set close to the maximum observed value in the bass range.

For each level of inharmonicity, five versions of starting phases $Ph(n,0)$ of the partials were generated. The resulting waveforms for the harmonic set are shown in Fig. 2.

Sine phases:

$$Ph(n,0) = 0. \tag{3}$$

Sine (zero) starting phase for all partials.

Alternate phases:

$$Ph(n,0) = \pi/4[(-1)^n + 1]. \tag{4}$$

Sine phases for odd partials and cosine for even partials.

Random phases:

$$Ph(n,0) = \text{random}[0,2\pi]. \tag{5}$$

Randomized starting phases.

Schroeder phases:

$$Ph(n,0) = -\pi n^2/N. \tag{6}$$

Sine phases for all partials, but with a consecutive delay (n) ranging from zero to half the fundamental period.

Negative Schroeder phases:

$$Ph(n,0) = \pi n^2/N. \tag{7}$$

Sine phases for all partials, but with a consecutive lead ranging from zero to half the fundamental period.

The Schroeder phases originate from a derivation of the phase relations, which minimize the peak factor of a periodic signal with a flat power spectrum [Schroeder (1970)]. Partial n of a Schroeder tone is given by

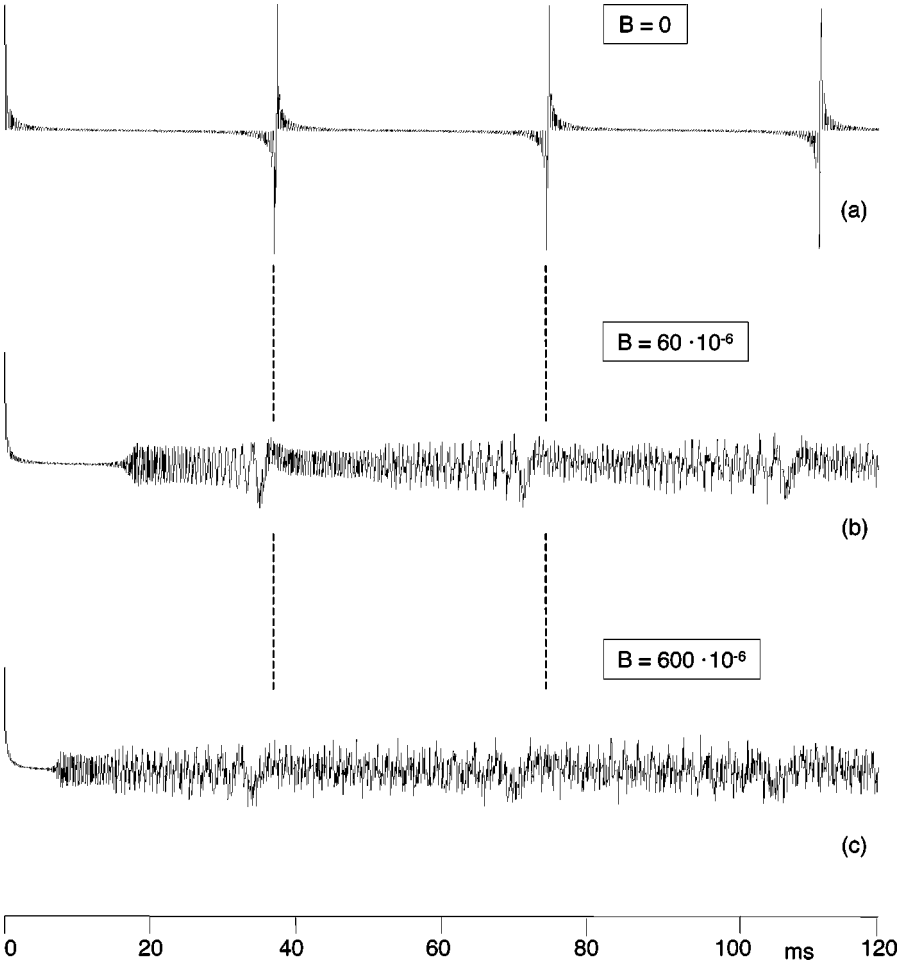


FIG. 3. Influence of the inharmonicity on the waveform during the initial periods. Waveforms of three A0 tones consisting of 100 consecutive harmonic partials of equal and constant amplitude, all with sine starting phases, but with a different amount of inharmonicity; (a) harmonic tone, $B=0$; (b) low-inharmonic tone, $B=60\times10^{-6}$; and (c) high-inharmonic tone, $B=600\times10^{-6}$. The vertical dashed lines indicate the fundamental period.

$$partial(n)_{\mathrm{SCH}}=\sin[2\pi f(n)t-\pi n^2/N]$$
$$=\sin[2\pi n f_1^0(t-delay(n))], \quad (8)$$

where

$$delay(n)=n/(2Nf_1^0)=(n/N)T_1^0/2, \quad (9)$$

and T_1^0 the fundamental period.

The starting phase relations in a Schroeder tone can thus be viewed as sine phases for all partials, but with a consecutive delay ranging from zero to half the fundamental period. Negative Schroeder tones show a corresponding consecutive lead. A (positive) Schroeder tone will resemble a chirped FM signal with monotonically increasing instantaneous frequency, while a negative Schroeder tone will give decreasing frequency within the period.

IV. EXPERIMENT I: DISCRIMINATION BETWEEN TONES WITH FIVE DIFFERENT STARTING PHASE RELATIONS ASSESSED UNDER THREE LEVELS OF INHARMONICITY

A. Overview of the stimulus waveforms

In this section we provide an overview of the stimulus waveforms and some characteristics of pitch and timbre, compiled from the spontaneous first impressions of a few listeners. The discrimination listening test is reported below in Sec. IV B.

1. Harmonic tones

The set of harmonic tones gives a clear demonstration of some features that are also of relevance for the more realistic inharmonic tones. As seen in Fig. 2, the five different relative phase conditions are clearly manifested in the waveforms. The sine-phase condition emphasizes the peak structure of the waveform and gives a chugging (''motorcycle'') timbre with a strong sensation of the fundamental periodicity. The alternate phase results in a quasihalving of the waveform period, giving an impression of an upward octave shift in pitch compared to sine-phase tones. Random phases provide a waveform with less clear indications of the fundamental period, and a smooth steady timbre. Schroeder phases result in upward pitch glides within each fundamental period, while with negative Schroeder phases the glides are downward. The Schroeder cases will be discussed in detail in connection with the inharmonic examples in the following section.

2. Inharmonic tones

The general effect of inharmonicity on the waveform is illustrated in Fig. 3, which compares the initial three periods of three tones, one from each set (harmonic, low-inharmonic and high-inharmonic), all with sine starting phases.

In the harmonic tone, the starting phases among the partials repeat each period, while in the inharmonic tones the phases will be disordered over time (''dephasing''). As seen

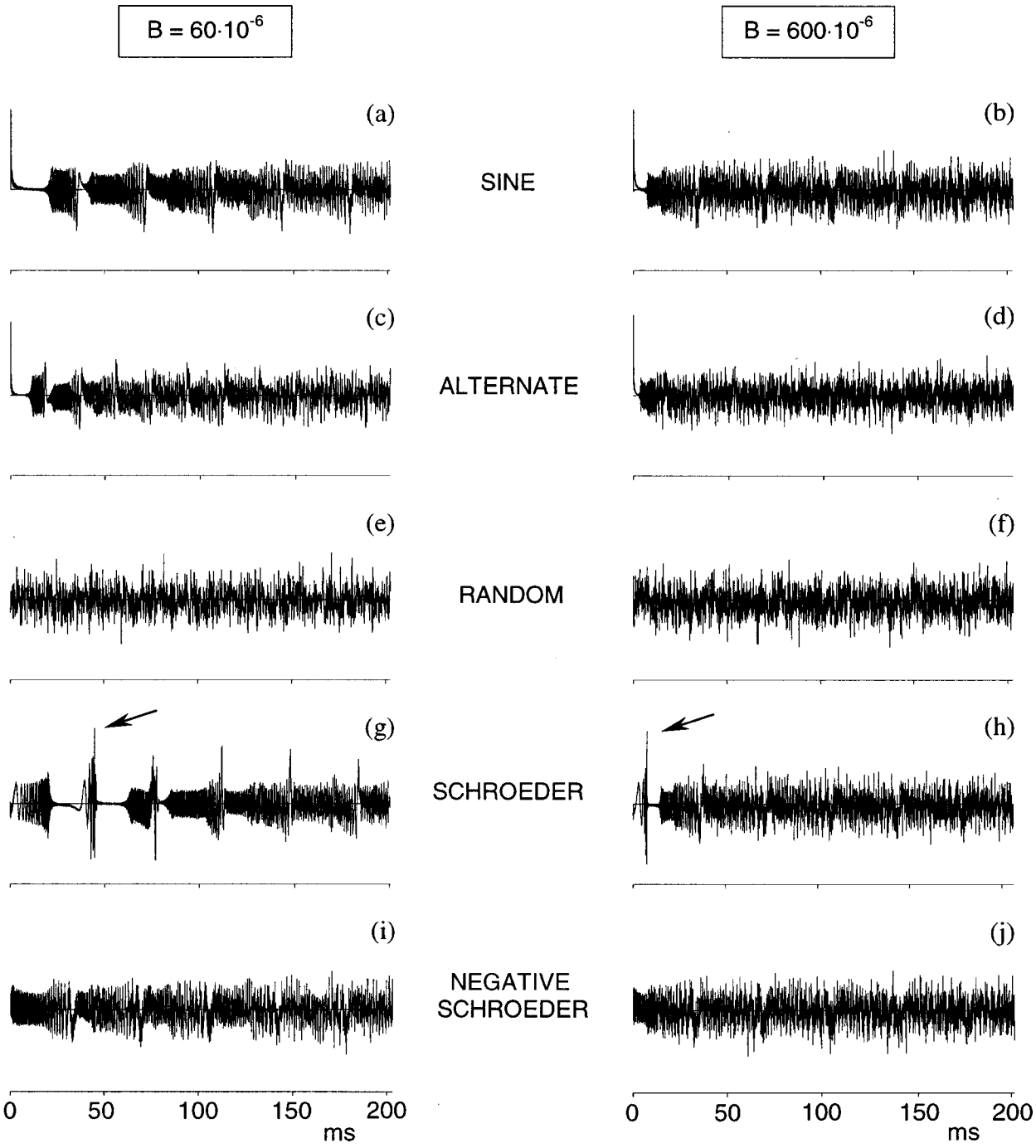


FIG. 4. Computed waveforms of the first 200 ms of ten A0 tones, consisting of 100 consecutive inharmonic partials of equal and constant amplitude. The tones differ in starting phases (*rows*) and inharmonicity (*columns*), low-inharmonic tones, $B=60\times10^{-6}$ (*left*), and high-inharmonic tones, $B=600\times10^{-6}$ (*right*). The arrows in (g) and (h) indicate moments of maximal in-phase relations in the Schroeder tones (see the text).

in the figure, the initial pulse in the two inharmonic tones is severely distorted after only a few periods, the energy contained in the initial pulse being smeared out over a substantial fraction of the period. During the first quasiperiods, this effect is manifested as characteristic downward frequency glides prior to the nominal position of the ''main pulse.'' These forerunners, or precursors, occupy about half of the first period of the low-inharmonicity tone, and about three-quarters of the first period of the high-inharmonicity tone. After a few periods, these glides disappear and the waveforms appear more stochastic.

For inharmonic tones it is thus important to distinguish between the starting phase conditions $Ph(n,0)$ and the instantaneous phase $Ph(n,t)$. The dephasing process will successively bring the starting phases out of order, randomizing the phases from period to period. The rate of the dephasing increases with increasing inharmonicity. As inharmonic tones are not periodic, the phase relations will never return to the starting conditions.

The first 200 ms of the waveforms of all ten tones included in the low-inharmonic and high-inharmonic sets are shown in Fig. 4 (the two inharmonic cases in Fig. 3 reap-

pearing at the top). A significant difference between random-phase tones [Figs. 4(e) and (f)] and the other tones can be seen. For tones with random starting phases, no characteristic changes in waveform (and timbre) with time can be observed. Due to the inharmonicity, the phases of the other tones (which all start with nonrandom phases) are successively randomized, and after some time it is difficult to attribute the waveform (and timbre) to a specific starting phase.

In the Schroeder case, the inherent upward pitch glides generate an interesting conflict with the downward glides due to inharmonicity. Comparing the waveforms in Fig. 4 it can be observed that the sine starting phases are randomizing gradually from one period to the next. Negative-Schroeder phases are also randomizing monotonically, only faster. In contrast, Schroeder tones start with an upward glide, but the cumulative effect of inharmonicity overcomes this tendency with time. Inharmonicity finally overrides, resulting in a downward glide. The turning point between these two conflicting mechanisms corresponds to a distinctive peak in the waveform, separating the initial upward glide and the following downward glide. In Fig. 4(g) this event takes place at about 45 ms after the onset of the tone (indicated by an arrow). With higher inharmonicity, the turning point occurs sooner, after about 4 ms in Fig. 4(h). The presence of a peak in the waveform indicates a ''phase concentration'' at this moment, meaning that the partials are more in phase at this moment than before or after.

B. Listening test 1

In Listening test 1, we investigated whether listeners were able to discriminate between tones with five different starting phase relations (Sine, Alternate, Random, Schroeder, and Negative Schroeder) under three levels of inharmonicity. Three sets of tones: (1) harmonic (2) low inharmonic, and (3) high inharmonic were independently examined in separate blocks of trials, using an ABX procedure. We expected that tones of the harmonic set would be easiest to discriminate because phase relations of these tones remain constant across duration.

1. Method

(a) Listeners. Eight listeners were recruited from the Queen's University community, 3 men and 5 women ranging in age from 19 to 36. All listeners were paid for their participation. All had received a minimum of seven years formal training on a musical instrument. All reported normal hearing.

(b) Materials. The stimuli consisted of the three complete sets of tones (harmonic, low inharmonic, and high inharmonic) described in Sec. III, each set including five starting phase relations. All tones had a pitch of A0 ($f_1^0 = 27.5$ Hz) and duration of 1 s. A pause of 10 ms separated all tones in a trial.

(c) Apparatus. The stimulus presentation and collection of responses were controlled by a PC Pentium computer running dedicated listening test software.[2] Tones were presented using a Sound Blaster acoustical output (32 bit), driving Sennheiser HD580 precision headphones with magnitude response within ± 3 dB across 16–30 000 Hz. Before the actual collection of data started, the listeners were asked if the loudness was comfortable or needed adjustment, but no one exercised this option. In effect, the loudness level was the same for all listeners. Consideration of the phase response of the headphones is addressed in the Appendix.

TABLE I. Mean discrimination for tone combinations within each tone set in Listening test 1. Note: Mean discrimination was collapsed across all ABX permutations.

	Sine	Random	Schroeder	Negative Schroeder	Alternate
Harmonic ($B=0$)					
Sine	···	100%	100%	98.43%	100%
Random		···	100%	100%	100%
Schroeder			···	96.88%	100%
Negative Schroeder				···	100%
Alternate					···
Low inharmonic ($B=60\times10^{-6}$)					
Sine	···	100%	93.75%	90.63%	98.44%
Random		···	100%	100%	100%
Schroeder			···	100%	100%
Negative Schroeder				···	100%
Alternate					···
High inharmonic ($B=600\times10^{-6}$)					
Sine	···	100%	89.06%	76.56%	98.44%
Random		···	100%	100%	100%
Schroeder			···	93.75%	100%
Negative Schroeder				···	92.19%
Alternate					···

(d) Procedure. Listeners heard 80 trials in each block, two presentations each of the 10 possible tone combinations in each of the four possible ABX permutations. The order of trials within blocks was independently randomized for each listener. In any given trial, listeners heard three tones; the third tone was always a repetition of the first or second tone (i.e., ABX procedure). The task was to verbally state whether the third tone (X) matched the first tone (A) or the second tone (B). Within each block, all possible ABX combinations and permutations of the five tones were tested. Listeners were able to hear the trial as often as necessary before submitting a response. A cash prize was awarded to the listener achieving the highest percentage correct; thus listeners were inclined to only submit a response once they were confident. Once a response was submitted, feedback was provided in the form of ''correct'' or ''incorrect.'' Prior to commencing each block, listeners were given approximately 15 min to familiarize themselves with the tones used in the block and to practice the discrimination task with those tones. The order of blocks was identical for all listeners: (1) harmonic, (2) low inharmonic, and (3) high-inharmonic. This block order was designed to follow the expected progression from easy to difficult discrimination.

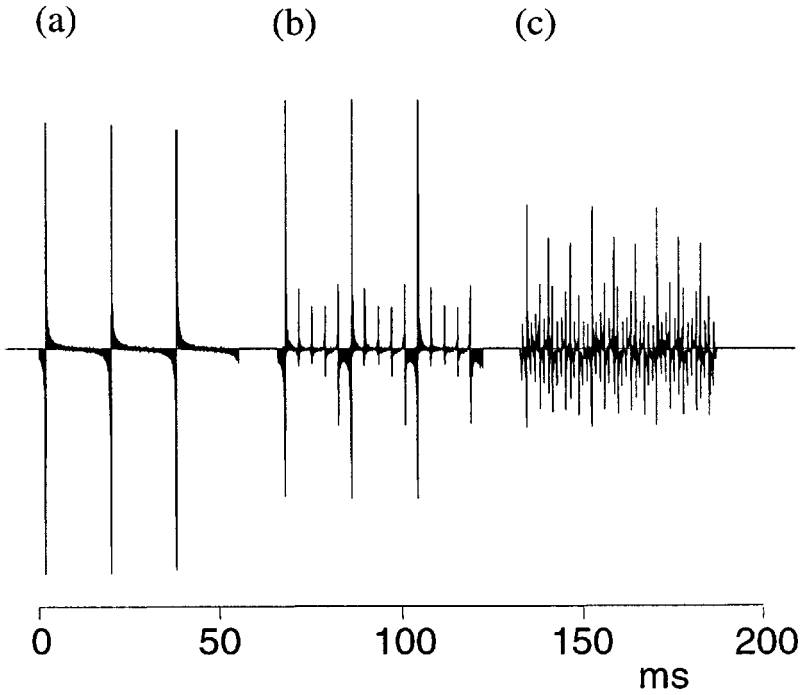


FIG. 5. Computed waveforms of the first three periods of three A1 tones ($f_1^0 = 55$ Hz) consisting of 100 consecutive harmonic partials of equal amplitude. The tones have the same magnitude spectrum, but differ in phase spectrum. In (a) all partials have sine starting phases; in (b) every fifth partial has a phase shift of π, and in (c), every third partial has a phase shift of $\pi/2$ and every fifth partial a phase shift of π. The pitch structure is different for the three tones (see the text).

2. Results and discussion

The overall mean percentage of correct discrimination was well above chance (i.e., 50%) at 97.6%. Variance in mean percentage correct across tone sets was small but significant with harmonic at 99.53% (SD=0.65), low inharmonic at 98.28% (SD=1.88), and high-inharmonic at 95% (SD=5.9), $F=(2,14)=2.8$, $p<0.05$. This variance indicates that ability to discriminate between tones having different phase is related inversely to the level of inharmonicity in those tones. Table I lists for each tone set the mean percentage of correct discrimination for each of the AB tone combinations (collapsed across all permutations). The most difficult discriminations were Sine-phase and Negative Schroeder-phase combinations, particularly in the high-inharmonic tone set. However, even for difficult tone combinations the mean discrimination was well above chance. In conclusion, listeners are able to discriminate between tones having different starting phase relations, even when those tones possess high levels of inharmonicity.

V. EXPERIMENT II: PITCH OF HARMONIC ALTERNATE-PHASE TONES

The listeners' spontaneous reactions to the stimulus waveforms (Sec. IV A) included three kinds of responses when comparing harmonic sine and alternate-phase tones: (1) The pitch of the alternate-phase tone was *an octave higher* than that of the sine-phase tone; (2) the alternate-phase tone was *brighter* than the sine-phase tone; (3) the alternate-phase tone had *ambiguous pitch*, differing by one octave. A dedicated study of the reason for these differing responses was undertaken in Experiment II. The term *pitch structure* is used as a common term for all possible pitches associated with a particular tone.

A. Pitch and timbre interactions

Response (1) reporting an octave shift is not surprising. Alternate phase relationships are known to generate an octave shift in pitch compared to sine-phase harmonics, when applied to unresolvable higher partials of middle-range fundamentals [Shackleton and Carylon (1994)]. In our bass tones, all partials are unresolvable or close to that, so the octave-shift response could be expected.

The appearance of response (2) reporting differences in brightness as an alternative to the pitch-shift (1) could be explained by the well-known interaction between pitch and timbre [Singh (1987); Crowder (1989); Iverson and Krumhansl (1989, 1993); Hesse (1982)]. In particular, the "brightness" or "sharpness" [Lichte (1941)] dimension of timbre, known to reflect the "position of energized spectral region" [Von Bismark (1974)], or the "spectral center of gravity" [Wessel, Bristow, and Settel (1987)], has been shown to interact with pitch. If other pitch cues are weak, a change in sharpness (a property that can be scaled from low to high as can pitch) might be confused with a change in pitch [Singh and Hirsh (1992); Pierce (1992)]. The pitch-shift response in (1) is thus not categorically different from the brightness-change response in (2). No clear perceptual border between changes in pitch and changes in brightness does exist.

Response (3) about the ambiguous pitch of the alternate-phase tone can be supported by rearranging the expression for the tone in even and odd terms:

$$
\begin{aligned}
tone_{\mathrm{ALT}}(n) &= \sum_{n}^{\text{odd}} \sin 2\pi n f_1^0 t + \sum_{n}^{\text{even}} \cos 2\pi n f_1^0 t \\
&= \sum_{n}^{\text{all}} \sin + \sum_{n}^{\text{even}} \cos - \sum_{n}^{\text{even}} \sin \\
&= \sum_{n}^{\text{all}} \sin + \sum_{n}^{\text{even}} (\cos - \sin) \\
&= \sum_{n=1}^{N} \sin 2\pi n f_1^0 t \\
&\quad + \sqrt{2} \sum_{n=1}^{N/2} \sin(2\pi 2n f_1^0 t + 3\pi/4). \qquad (10)
\end{aligned}
$$

In other words, an alternate-phase tone with 100 harmonics should sound the same as the sum of a sine-phase tone with 100 harmonics and a tone one octave higher with 50 harmonics having $\sqrt{2}$ times larger amplitude and different phase.

Any change of the phase of some particular partial(s) in a complex tone can be interpreted as an addition (or subtraction) of another complex tone made up of these manipulated partials. This gives a selective attention to the ''added'' or ''subtracted'' tone when comparing two consecutive tones with different phase relations. A striking demonstration of this effect was given by Schroeder (1975), who reported that it was possible to play ''Mary had a Little Lamb'' using sounds consisting of the 31 lowest harmonics of a complex tone with a fundamental of 100 Hz and differing in phase spectrum only.

B. Alternate-phase tones with additional pitches

In order to be able to discriminate between ''changes in pitch'' and ''changes in brightness'' when comparing sine and alternate-phase tones in listening tests, we decided to try to produce phase effects which would give pitch shifts other than an octave. Preliminary experiments showed that the perceived differences between sine and alternate-phase tones were more audible for tones one octave above A0. The most probable explanation is that A1 gives a stronger perception of the pitch corresponding to the fundamental. For this reason A1 tones ($f_1^0 = 55.0$ Hz) were used in Experiment II.

Two A1 tones were synthesized, each consisting of 100 harmonics. In one tone the phases of every third partial were shifted by π, and in the other tone the phases of every fifth partial were shifted by the same amount. In this way, a pitch corresponding to the third and fifth harmonic, respectively, was emphasized. The tones were generated as

$$
\begin{aligned}
tone3(k) = \sum_{k=0}^{32} &[\sin 2\pi(3k+1) f_1^0 t + \sin 2\pi(3k+2) f_1^0 t \\
&+ \sin(2\pi(3k+3) f_1^0 t + \pi)] + \sin 2\pi 100 f_1^0 t, \qquad (11)
\end{aligned}
$$

which in analogy with Eq. (10) can be rearranged as

$$
tone3(k) = \sum_{k=1}^{100} \sin 2\pi 100 f_1^0 t - 2 \sum_{k=1}^{33} \sin 2\pi 3k f_1^0 t. \qquad (12)
$$

Similarly,

$$
tone5(k) = \sum_{k=1}^{100} \sin 2\pi 100 f_1^0 t - 2 \sum_{k=1}^{20} \sin 2\pi 5k f_1^0 t. \qquad (13)
$$

For *tone3* and *tone5*, waveforms were obtained in which the period was divided by rather prominent peaks in three and five equal parts, respectively. In a pilot test it was found that when these tones were presented over headphones, a ''fundamental pitch'' corresponding to A1 was perceived, as expected. In addition, a distinctive pitch at an octave plus fifth, E3 (or just a fifth, E2, due to the common octave ambiguity) was perceived for *tone3*. For *tone5*, the additional pitch was two octaves plus a major third, $C^{\sharp}4$, with octave ambiguities. Depending on the listening conditions (headphones, loudspeakers), the $C^{\sharp}4$ pitch could sometimes be perceived as stronger than the A1 pitch. In Listening test 2 we tested whether a sequence of the phase-manipulated tones (*tone3*, *tone5*) gave rise to changes in pitch or in timbre (see Sec. V C).

In further experiments with complexes consisting of 100 equal-amplitude harmonics with a fundamental frequency of 55 Hz (A1), the phases of partials were adjusted so as to emphasize the pitches of the third and fifth harmonics simultaneously. The effect on the waveform is illustrated in Fig. 5. The figure shows the detailed waveforms of (a) an A1 tone with all partials starting with sine phases, (b) an A1 tone with every fifth harmonic phase shifted by π, and (c) an A1 tone with every third harmonic phase shifted by $\pi/2$ and every fifth harmonic phase shifted by π. The amplitude spectrum is the same throughout the sequence. The period of the second tone is divided by peaks into five parts, while the third tone is divided into 15 parts.

In a pilot test, the response of some listeners when listening over headphones was that the sequence consisted of the tone A1 followed by a major third interval ($A1–C^{\sharp}4$) spread over one or two octaves, and then by a (weak) major chord ($A–C^{\sharp}–E$). Other listeners reported a major arpeggio in a wide disposition, following the most distinctive pitch changes of the sequence. When listening over loudspeakers in a room with normal reverberation the three tones of the sequence were often perceived as differing in timbre only.

C. Listening test 2: Pitch effects in harmonic alternate-phase tones

In Listening test 2, we assessed whether listeners could perceive pitches other than the pitch corresponding to the fundamental frequency in harmonic tones with shifted phases of selected sets of harmonics.

To assess this, three-tone sequences were presented as a short melody. Each tone of a sequence had an identical harmonic spectrum, but different phase relations: (1) all sine phases; (2) all sine phases except every third harmonic phase shifted by $\pi/2$; and (3) all sine phases except every fifth harmonic phase shifted by $\pi/2$. All three tones could be predicted to possess the pitch A corresponding to the fundamental, supplemented by the additional pitches E and C in (2) and (3), respectively. Listeners were asked to match each three-tone sequence to one of 27 possible three-tone sequence alternatives. Each three-tone sequence was presented

in each of three reverberation conditions: (1) headphones, (2) speaker near, and (3) speaker far. We expected that pitch effects would be clearest under low-reverberation presentation conditions (i.e., headphones and speaker near).

1. Method

(a) Listeners. Ten listeners, five men and five women, were recruited from the Kingston and Queen's University communities. Listeners ranged in age from 19 to 51. Five listeners were paid for participation; of these, three were recruited from responses to posters on campus and two were known to the experimenters. Three listeners were recruited from a first-year psychology subject pool and received course credit for their participation. The remaining two listeners were members of the Acoustics Laboratory and did not receive any remuneration. All had received training on a musical instrument to a minimum performance level of Grade VI Royal Conservatory (or equivalent). All reported normal hearing.

(b) Materials. The stimuli were sequences of three synthesized tones, each tone with different phase relations: (1) all sine phases, (2) all sine phases except every third harmonic phase shifted by $\pi/2$, and (3) all sine phases except every fifth harmonic phase shifted by $\pi/2$. All tones were harmonic with a pitch of A1 ($f_1^0 = 55$ Hz), flat, steady magnitude spectrum, and a duration of 0.5 s.

(c) Apparatus. The tones were presented over headphones and loudspeakers. Headphone presentation was identical to that described in Experiment I (Sec. IV B 1). Speaker presentation was realized using a small computer loudspeaker (Creative CS100), specified to be within ± 3 dB across 20–20 000 Hz. The loudspeaker was used in two conditions; (1) Speaker near: The listeners were asked to hold the loudspeaker within 2–3 cm from their right ear; (2) Speaker far: The loudspeaker was positioned at a distance of about 90 cm from the listener with a reverse orientation so that the listener was not exposed to the direct sound. The listening tests were conducted in a small office room (2×3 m) with concrete walls. Consideration of the phase response of the loudspeaker presentation is addressed in the Appendix.

(d) Procedure. Each listener heard each of six tone sequences in each of three presentation conditions. The sequences were the six possible permutations of the three tones presented with a 10-ms pause between each tone. The order of presentation conditions was identical for all listeners: (1) headphones, (2) speaker near, and (3) speaker far. In each presentation condition, each of the six tone sequences was presented three times, for a total of 18 trials. The order in which trials were presented was randomized independently for each presentation condition and each listener.

Listeners were told that the three tones could be heard either as a sequence of single pitches, or as a sequence containing some, or all, of the pitches A, $C^\sharp$, and E. Thus, there were 27 possible response alternatives. A sample of the possible response alternatives was demonstrated on the piano keyboard.

Listeners were asked to enter the letter names of the pitches that they heard, in order, on the computer keyboard (e.g., AAA, $AC^\sharp E$, $C^\sharp EA$, etc). Listeners were instructed to ignore tone height and to attempt to focus on the similarities or differences among the three sounds. After completing all 18 trials, they were asked to listen to each trial again and to verify their judgements. Listeners were allowed to use the piano keyboard for assistance. They were also permitted to replay each trial as many times as they liked.

2. Results and discussion

A response was considered ''expected'' if it conformed to the theoretical predictions. A strict criterion was used; each pitch had to be labeled in the predicted order. For example, when the prediction was that the pitches of the three sounds was A E $C^\sharp$, only that order and no other permutation of the pitches, was considered ''expected.''

The majority of responses (72.4%) contained three different pitches. The mean number of expected responses across listeners, out of a possible maximum of 18, was 9.33 for the headphone condition, 8.67 for the speaker-near condition, and 3.00 for the speaker-far condition. Analysis of variance revealed a significant effect of mode of presentation, $F(2,18) = 10.35$, $p < 0.001$. Orthogonal contrasts revealed that the number of expected responses in the headphone condition was not significantly different from the speaker-near condition, $F(1,9) = 0.98$, $p = 0.35$, but that the combined mean for the headphone and speaker-near conditions surpassed that for the speaker-far condition, $F(1,9) = 12.54$, $p < 0.001$.

In conclusion, the listening test showed that the perception of alternative pitches is a relatively difficult task, in particular in the reverberant speaker-far condition. Music training may have played a role in fostering responses that conformed to our expectations. A ''high-conformance'' group with four listeners (min 17, max 17.5 expected responses) contrasted distinctively against a ''low-conformance'' group with two listeners (min 1.5, max 3 expected responses). The high-conformance group included listeners with recent or current music training, and activities such as conducting and composing.

VI. EXPERIMENT III. PHASE CONCENTRATIONS ("SQUEALS")

In general, a correspondence between the peak structure of the waveform and timbre could be observed, as discussed above (see Sec. IV A). Sometimes, however, some significant exceptions occurred. In particular, when listening over headphones to tones with high inharmonicity, two distinctive acoustical ''events'' or ''inclusions'' of short duration were distinguished within the one-second duration of the tone (except for the random-phase tones). These events, which had the character of ''squeals,'' were perceptually separated from the tone and are not visible in the waveforms. The squeals occurred at about 0.4 and 0.8 s from the onset of the tones. After such a long time, the phases of partials are normally randomized so much that the tone would sound as a random-phase tone with steady timbre. The most plausible explanation for the squeals would be that they are generated by a

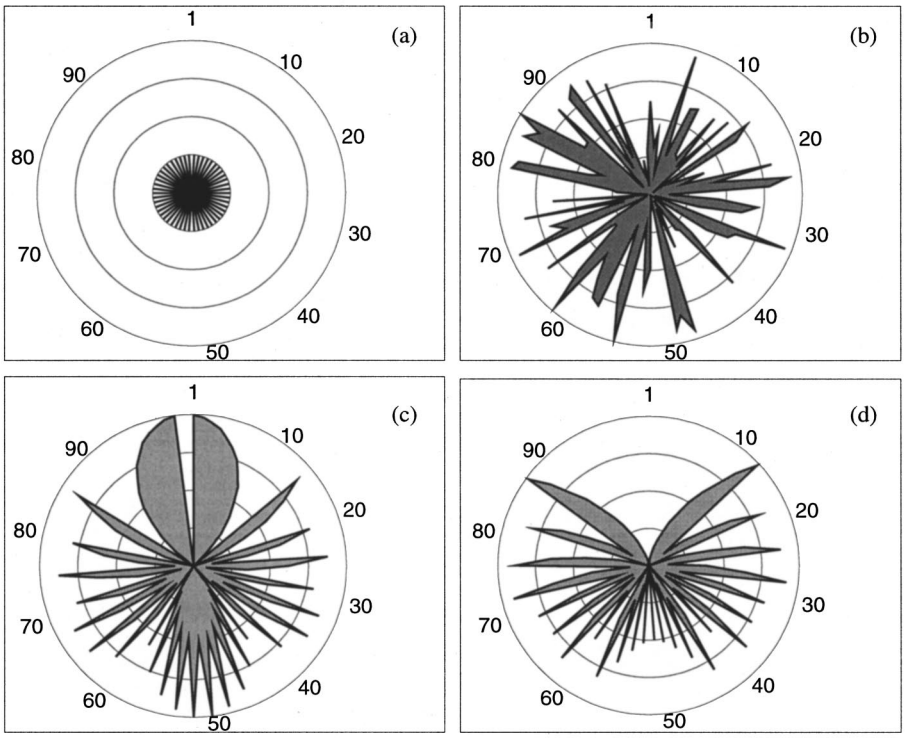


FIG. 6. Phase images of the starting phases ($t=0$) of all 100 partials for (a) alternate, (b) random, (c) Schroeder, and (d) negative Schroeder tones. The partial number is shown along the perimeter of the circle, and the phase angle along the radius (0 at the center and 2π at the perimeter). Grid line circles correspond to steps of $\pi/2$.

group of partials, temporarily fused by their phase relations (''phase concentrations''). An analysis that evaluates this hypothesis is given in Sec. VII B. First, listeners' ability to detect squeals was tested.

A. Listening test 3: Timbral effects of phase concentrations (squeals)

In Listening test 3, we assessed the consequences of random and nonrandom starting phase relations on the perception of inharmonic tones. Our prediction was that listeners would only perceive squeals in tones possessing nonrandom starting phase relations (e.g., sine, alternate, Schroeder, or negative Schroeder). As in Listening test 2, tones were presented in each of three reverberation conditions: (1) headphones, (2) speaker near, and (3) speaker far. We anticipated that the detection of squeals would be clearest in low-reverberation presentation conditions, where phase is relatively well preserved.

1. Method

(a) Listeners. Nine listeners, four men and five women, were recruited from the Kingston and Queen's University communities. Six of the nine listeners had participated in Listening test 2. Listeners ranged in age from 19 to 51. Six listeners were paid for participation; of these, five were recruited from responses to posters on campus and one was known to the experimenters. Two listeners were recruited from a first-year psychology subject pool and received course credit for their participation. The remaining listener was a member of the Acoustics Laboratory and did not receive any remuneration. All had received training on a musical instrument to a minimum performance level of Grade VI Royal Conservatory (or equivalent). All reported normal hearing.

(b) Materials. The stimuli consisted of the five tones of the high-inharmonicity set (see Sec. III), including four tones with nonrandom starting phase relations (i.e., sine, alternate, Schroeder, or negative Schroeder) and one tone with random starting phase relations (random). All tones had $B=600\times10^{-6}$, flat, steady magnitude spectrum, pitch of A0 ($f_1^0=27.5$ Hz) and a duration of 1 s. The four tones with nonrandom phase relations each occurred twice and the tone with random phase relations occurred 8 times, for a total of 16 trials.

(c) Apparatus. The tones were presented over headphones and loudspeakers, as described in Experiment II (Sec. V C 1).

(d) Procedure. The order of trials within each presentation condition was independently randomized for each listener. The order of presentation conditions was identical for all listeners: (1) headphones, (2) speaker near, and (3) speaker far. Listeners were instructed that each trial contained one of two types of sound. They were asked to classify the sound as to whether it contained squeals (i.e., if they heard ''events'' in the timbre), or not (i.e., they heard a ''smooth'' timbre). They recorded judgements on the computer keyboard, ''1'' for ''squeals'' and ''0'' for ''no squeals.'' Listeners were allowed to replay each trial as many times as they wished before recording a judgement.

2. Results and discussion

For purposes of scoring, a judgment was considered ''expected'' if it corresponded to predictions. Given the prediction that only nonrandom starting phase relations would

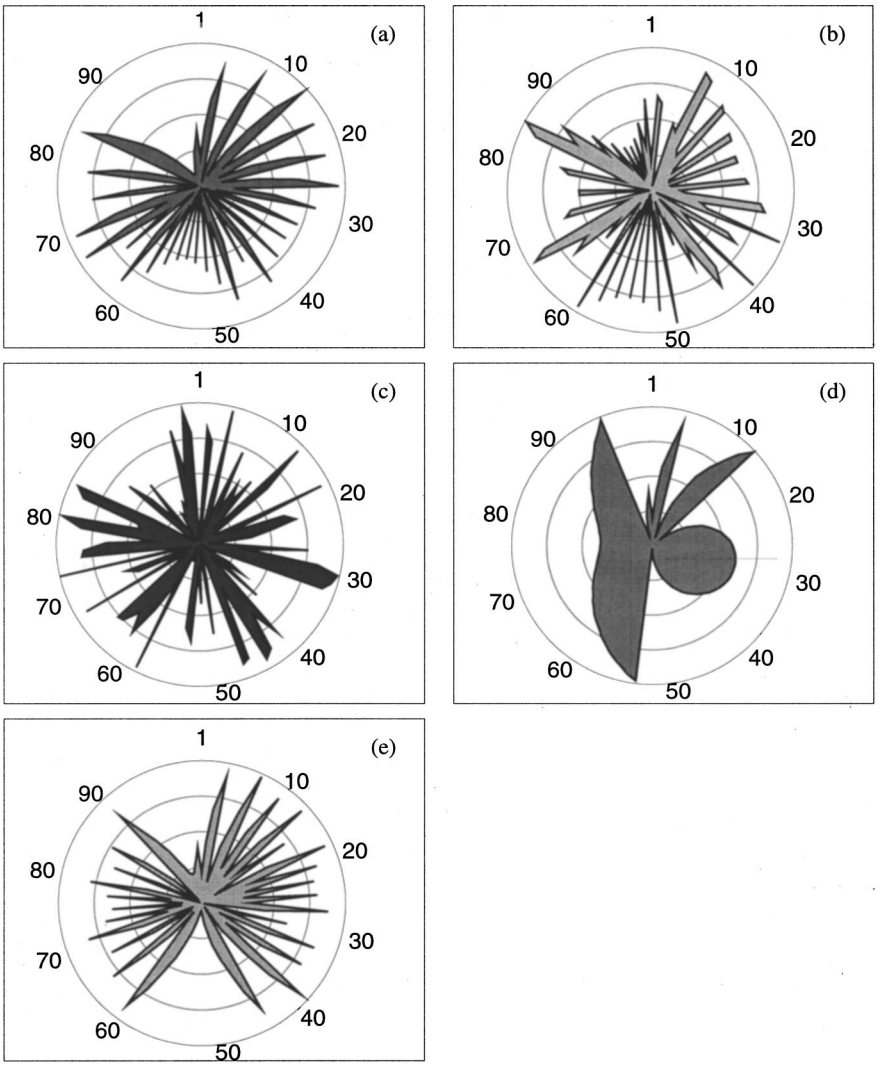


FIG. 7. Phase images for the (a) sine, (b) alternate, (c) random, (d) Schroeder, and (e) negative Schroeder tones in Fig. 4 (left column) at $t = 43.8$ ms after the onset, corresponding to the moment of the most prominent peak in the waveform in the Schroeder tone with low inharmonicity [Fig. 4(g)].

contain squeals, an assignment of ''1'' on trials with nonrandom phases and an assignment of ''0'' on trials with random phases was scored, as expected. The mean number of expected responses across listeners, out of a possible maximum of 16, was 15.89 for the headphone condition, 15.67 for the speaker-near condition, and 11.44 for the speaker-far condition. Analysis of variance revealed a significant effect of mode of presentation, $F(2,16) = 21.28$, $p < 0.001$. Orthogonal contrasts revealed that the number of expected responses in the headphone condition was not significantly different from the speaker-near condition, $F(1,8) = 0.64$, $p = 0.457$, but that the combined mean for the headphone and speaker-near conditions surpassed that for the speaker-far condition, $F(1,8) = 22.72$, $p < 0.001$.

In conclusion, the listening test showed that listeners were able to detect phase concentrations corresponding to squeals in the tones with nonrandom starting phases with high accuracy. This was true even in the reverberant speaker-far condition, where performance was better than chance, $X^2(8,n=9) = 21.63$, $p < 0.01$.

VII. INDICATORS OF PHASE REGULARITY

Inspired by the results in Listening test 3, which indicated that temporary phase concentrations can be perceived by a majority of listeners, different ways of illustrating and quantifying the instantaneous phase relations were developed.

A. Instantaneous phase images

A visual indicator of the regularity of the relative phases among partials, coined *instantaneous phase image*, was found to be useful when discussing the regularity of phase relations and the evolution of the phase relations over time. Such an image represents the phases of the partials at a particular moment of time as a contour in a radar diagram. The phase angle is shown as the distance from the center ($n.b$: not as an angle) with zero at the center and 2π at the perimeter. The partials (1–100) are arranged clockwise around the circle.

In Fig. 6 four phase images representing the starting phases for (a) alternate, (b) random, (c) Schroeder, and (d) negative Schroeder tones are shown. The sine-phase tone is not included as the phases are all concentrated at the center of the circle in that case. It is easy to distinguish between the three cases with regular patterns of the starting phases (periodicity, symmetry), and the irregularity of the random-phase case. Using the phase image representation, the influence of

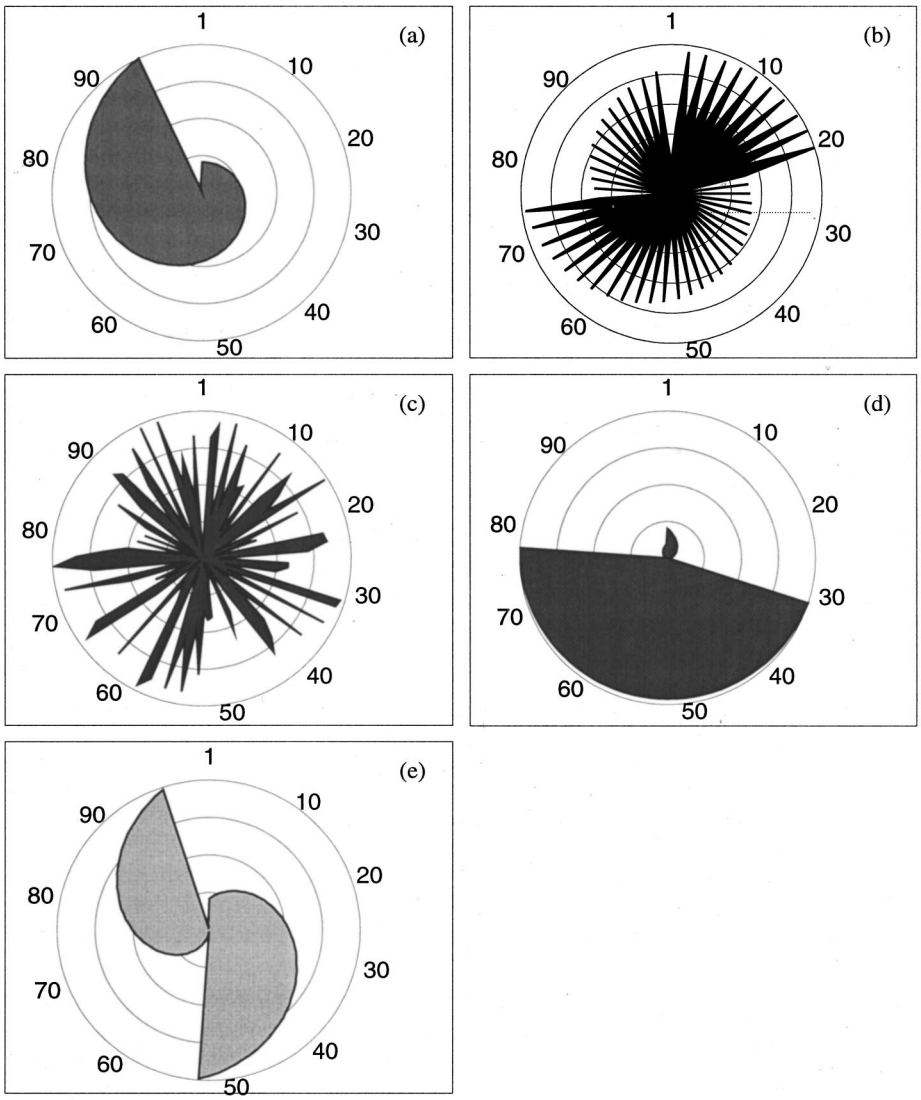


FIG. 8. Images of the phase differences between adjacent partials for the tones in the left column of Fig. 4 ($t = 43.8$ ms). (a) Sine, (b) alternate, (c) random, (d) Schroeder, and (e) negative Schroeder tones. Observe that the shaded area in the Schroeder case does not extend fully to the 2π circle.

inharmonicity (or other dephasing factors) on the phase evolution can easily be traced over time.

In Fig. 7, phase images for the five tones with low inharmonicity in Fig. 4 (left column) are shown at the moment of the peak in the Schroeder tone [$t = 43.8$ ms after the onset, Fig. 4(g)]. Unlike the other tones, the phase image of the Schroeder tone shows a clear regularity [see Fig. 7(d)]. The majority of the partials, including large groups of adjacent partials, have phases below π, thus forming a positive peak in the waveform. Also, within these groups there is very little change in phase between adjacent partials. This ''in-phase'' property between adjacent partials becomes even more evident when examining images of the *phase differences* between adjacent partials, wrapped to the 0–2π range [see Figs. 8(a)–(e)]. For the Schroeder tone in (d), almost all phase differences between adjacent partials are close to zero (2π) at the moment of the peak in the waveform. The development of the Schroeder tone after that point is consequently close to that of the sine-phase tone.

B. Quantitative approach

Listening tests 1 and 3 suggested that the pitch and timbre of the stimuli were consistent with the waveform features, in particular, the presence or absence of a clear peak structure, which in turn are dependent on the phase relations among the partials. The perceptual effects related to this observation can be summarized as follows.

(1) If the starting phases among partials are *random*, there is no evident peak structure in the waveform and the timbre of the tone is only dependent upon the degree of inharmonicity.

(2) If the starting phases are *not random*, the timbre of the tone changes with time because of the dephasing driven by inharmonicity. After some time when the phases have been randomized enough, the initial peak structure of the waveform has disappeared and the tone has reached the timbre of a random-phase tone with the same inharmonicity.

Some instantaneous measure of the regularity of the phases might thus serve as an indicator of the timbre. There is no single mathematical definition of the regularity or randomness of an arbitrary finite series of numbers, but various approaches have been suggested [for a survey, see Pincus (1991)]. Generally, a series of values within a certain interval is considered more random than another series within the same interval, if (1) the values are more uniformly distributed over the range, and/or (2) the order of values is less

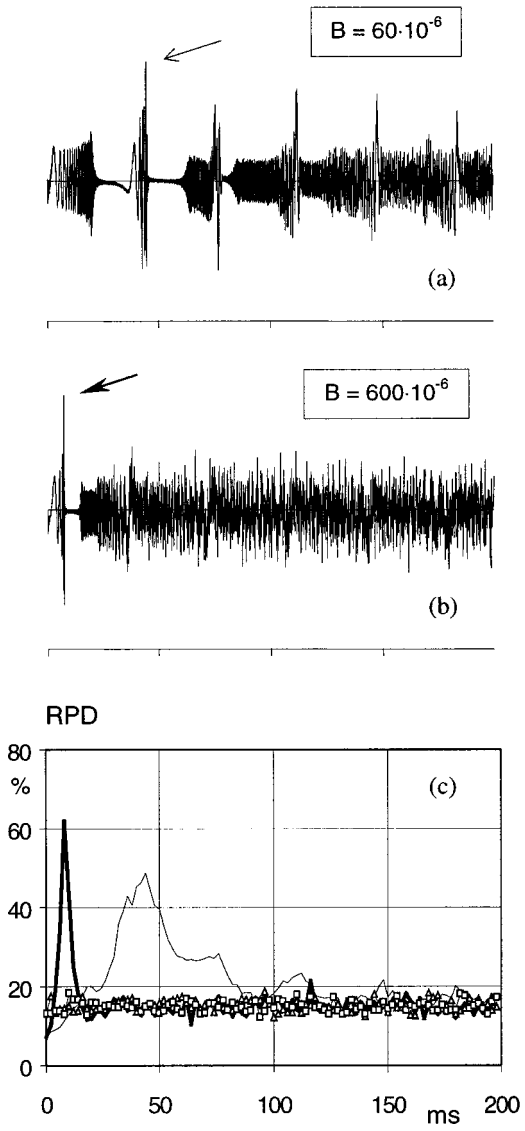


FIG. 9. Comparison between waveforms and RPD curves (redundancy of successive phase differences). Computed waveforms of the Schroeder tones in Fig. 4 with (a) low and (b) high inharmonicity, and (c) corresponding RPD curves (low *B, thin line*; high *B, thick line*). The arrows indicate moments of maximal in-phase relations in the waveforms, corresponding to the peaks in the RPD curves in (c). Also included in (c) are RPD curves for two random-phase tones with low inharmonicity (*triangles*), and high inharmonicity (*squares*).

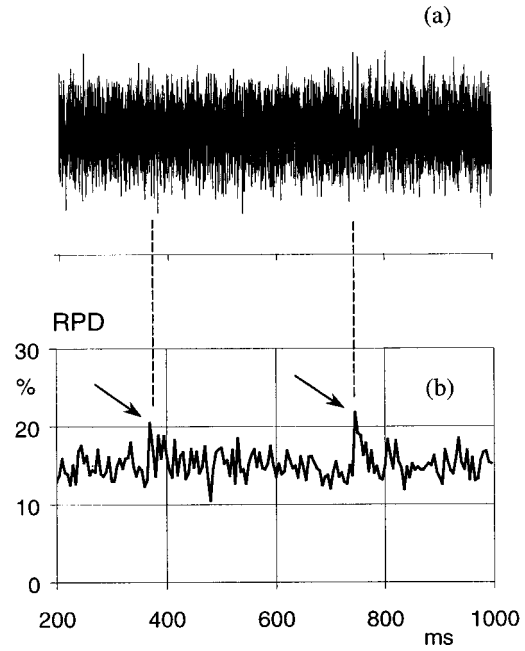


FIG. 10. A fragment of (a) the computed waveform, and (b) the RPD curve of a high-inharmonic, negative Schroeder tone (A0=27.5 Hz, $B=600\times10^{-6}$) containing "squeals." Indications of temporary phase concentrations are seen in the RPD curve (*arrows*), but they are not visible in the waveform.

predictable. As a reasonable working hypothesis it was assumed that criterion (1)—the uniformity of the distribution of the phase differences between adjacent partials—alone could serve as a first-order estimate of the peak structure in the waveform, and hence also as an indicator of timbre. Some support for this suggestion might be found in the Schroeder tones, which show a low peak factor in spite of highly ordered and predictable phase relations.

An additional assumption was made: The phase relations between *adjacent* partials were considered more important for pitch structure and timbre than the relations between other combinations of partials. As before, the term "pitch structure" is used as a common term for all possible pitches associated with a particular tone. The justification for this assumption is that the audible interaction between partials such as difference tones, beats, etc. are known to be stronger for close partials.

In a spectrum with N partials there will be $N-1$ phase differences $Pd(n,t)$ between adjacent partials with instantaneous phase $Ph(n,t)$:

$$Pd(n,t)=Ph(n,t)-Ph(n-1,t), \tag{14}$$

where $n=1,2,3,...,N$ and

$$Pd(1,t)=Ph(1,t)-0. \tag{15}$$

The phase difference range $[0, 2\pi]$ is divided into M equal subranges and the number of phase differences in each subrange is calculated, creating a sequence of occurrences o_i:

$$\{O(t)\}:\ o_1(t),o_2(t)\cdots o_M(t), \tag{16}$$

where

$$\sum_{i=1}^{M} o_i(t)=N \quad \text{for any } t. \tag{17}$$

It is further assumed that $M=N$. This means that an ideally uniform distribution would give one occurrence in each subrange. A calculation of how uniform the distribution of phase differences is across $[0, 2\pi]$ is made, using the redundancy concept [Attneave (1959)]. The redundancy of the phase differences (RPD factor) in percent is calculated as

$$RPD(t)=100[1-H(t)/H_{\max}], \tag{18}$$

where $H(t)$ is the entropy,

$$H(t)=-\sum_{i=1}^{N}[o_i(t)/N]\log[o_i(t)/N], \tag{19}$$

and

$$H_{max} = \log N; \tag{20}$$

RPD=0 corresponds to a case where the phase differences are distributed uniformly, and *RPD*=100% to a case where all phase difference values are concentrated in the same subrange (for example, as for the sine-phase tone).

In Fig. 9 the waveforms and the corresponding RPD trajectories for the two Schroeder tones in Figs. 4(g) and (h) are compared. It is evident that the RPD function gives a good description of the waveform envelope of both tones, the peaks of the waveform being perfectly matched in time with the phase concentrations indicated in the RPD curve. The RPD curves for random-phase tones included in Fig. 9(c) do not differ significantly between the two inharmonicity cases; neither do they change with time. A detailed analysis of all synthesized tones showed that the RPD curves mapped the peak structure of the waveforms accurately.

The RPD trajectory allows a rough estimation of the time elapsed before the timbre has reached its quasisteady state. According to the assumptions above, this should occur when the RPD curve had dropped to the level of the random-phase tone. This would occur after about 150 ms for the tone with low inharmonicity and much earlier for the tone with high inharmonicity [see Fig. 9(c)].

Figure 10 shows a fragment of a negative Schroeder-phase tone containing two short squeals. The waveform does not show any significant changes at the places corresponding to the squeals, but the RPD trajectory gives clear indications of the events. Our assumption that temporary phase concentrations were the reason for the squeals was thus verified.

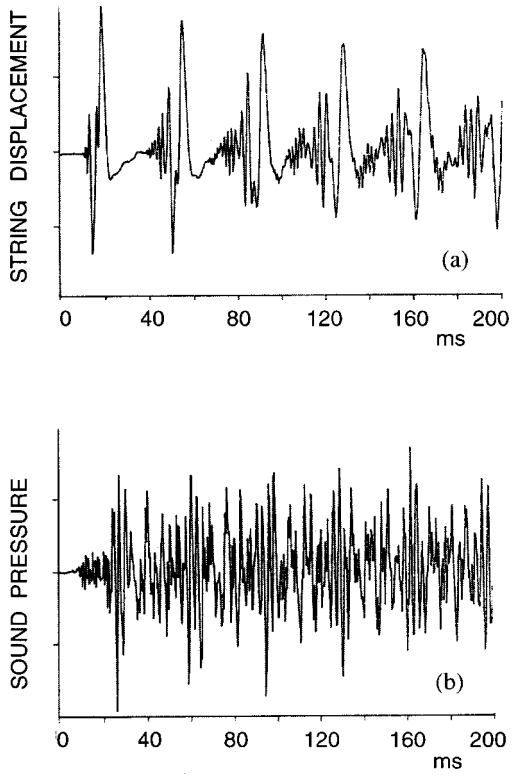


FIG. 11. Waveforms of (a) the displacement of an A0 piano string measured close to the bridge, and (b) the radiated sound. Compiled from Podlesak and Lee (1989).

VIII. DISCUSSION AND CONCLUSIONS

A. Survey of the results

In this study, we have explored some effects of phase on pitch and timbre that may be of relevance for musical tones in the bass range. The evolution of the waveform and phase relations in synthesized bass tones consisting of 100 partials was analyzed (A0=27.5 Hz and A1=55 Hz), including harmonic tones and inharmonic tones with string-like inharmonicity. Listening tests showed a clear effect of the relative phases of partials on both pitch structure and timbre.

For harmonic tones with flat steady magnitude spectra, it was shown that it was possible to transform a chugging motor cycle-like sounding sequence of impulses (sine phases) into a smooth timbre with weaker pitch strength (random phases), or into a sequence of periodic downward or upward pitch glides (Schroeder phases), only by changing the phase relations. Further, by changing the phase of selected groups of partials, it was possible to change the brightness of the tone, as illustrated by the differences between alternate-phase and sine-phase tones, and also to make the tone sound as an interval (octave, fifth, major third), or even a chord with two or three pitches emphasized simultaneously, with the lowest (fundamental) pitch not necessarily being the strongest.

Inharmonic tones with flat steady spectra were used to investigate the evolution of the phase spectrum over time. The changes in the phase spectrum were shown to be reflected in changes in the waveform and in the timbre over the duration of the tone. Only random starting phases of the partials provide steady timbre. If the starting phases are not random, the inharmonicity disorders them (''dephasing''), resulting in randomized phase conditions after some time. An ongoing process of randomization is audible only in the attack, as downward pitch glides, differently depending on the starting phase relations and the degree of inharmonicity. Higher inharmonicity gives a faster dephasing process.

A measure of the redundancy of the phase differences between adjacent partials (RPD) was calculated in order to estimate the phase regularity. High peaks in the RPD curve were shown to be consistent with the peak structure of the waveform. RPD peaks of small amplitude could be used to detect temporary phase concentrations, invisible in the waveform but audible as sudden changes in timbre (''squeals'').

The perceptual effect of different phase relations depends on listening conditions. When the tones were presented over headphones, very clear responses were reported for a majority of the listeners. When presented over loudspeakers in a normal room, the effect decreased with increasing distance to the loudspeakers. In a large and moderately reverberant room most listeners do not perceive any differences between the tones, not even in the simplest task with discrimination between harmonic tones with five different phase conditions. The reason for the large influence of the presentation conditions is that the phase relations are well preserved in the headphone presentation, but more or less deteriorated when listening over loudspeakers due to the interfering reflections in the room. This raises doubts about the significance of phase effects when listening in concert halls

or in other relatively reverberant conditions. Some specific cases in which an audible effect of the phase relations can be expected are discussed below in connection with the piano.

Regarding a possible effect of audible phase distortion introduced by the equipment in the listening test (primarily by the headphones and loudspeakers), it is unlikely that such distortions would provide cues that were both reliable and distinctive enough to enable perceptual discrimination. According to the evaluation in the Appendix there were large differences in the amount of phase deterioration for three different presentation conditions (headphones, speaker near, speaker far). For the headphone presentation it seems safe to conclude that any phase distortion introduced by the test equipment was by far outweighed by the phase differences in the computed stimuli. For the loudspeaker presentations (speaker near, speaker far) the room reverberation introduced a progressive deterioration in the stimuli with increasing listening distance, which is in concordance with the results of the listening tests.

B. Future studies and applications to the piano

In this study, only A0 and A1 tones were studied in combination with five well-defined phase relations. We consider our conclusions valid for at least the bass range below A1. The prominence of the perceived effects are dependent on the conditions of the presentation (headphones versus loudspeakers, room reverberation), and in some cases, musical training. Despite these limitations the experiments have pointed out several important perceptual phenomena, which need further research. A continuation focusing on perception should include the determination of a parameter space in which relative phases have audible effects on timbre and pitch structure. The main parameters would be fundamental frequency range, number of partials, degree of inharmonicity, and spectral envelope and its dynamic evolution. In this connection it would be appropriate to evaluate how useful the crest factor of a waveform is as a first-order approximation of the in-phase/out-of phase relations among partials [Schroeder (1970)].

Regarding applications to real piano tones, the dephasing processes at successive stages in the tone generation are of particular interest. There are several potential sources of phase randomization after the initial string excitation by the hammer. The first is *inharmonicity* due to string stiffness, as mentioned. The larger the inharmonicity coefficient and the higher the pitch, the faster the phases will be randomized. A second source of phase randomizing is the *soundboard*, a distributed radiator. An additional phase randomization might be expected in the internal volume of the case of the upright piano. A third phase randomizer is the *room*, due to multiple reflections. The listener is exposed to all three randomizing processes, while the pianist (tuner, voicer), being closer to the soundboard than the soundboard dimensions, is in a better position to experience a less dephased sound.[3]

There are few data available on the initial phase relations between the vibration modes of a struck piano string and the relative phases of the partials in the radiated tone. An indication that the string motion signal starts with nearly ordered partials, while the radiated piano tone starts with phases already disordered to some extent, can be obtained from the waveforms in Fig. 11 [Podlesak and Lee (1989)]. Audible pitch glides in the signal reported by Podlesak and Lee, as well as a characteristic evolution in waveform within the period comparable to the inharmonic tones in Fig. 3, illustrate the dephasing effect from the string inharmonicity in a real piano.

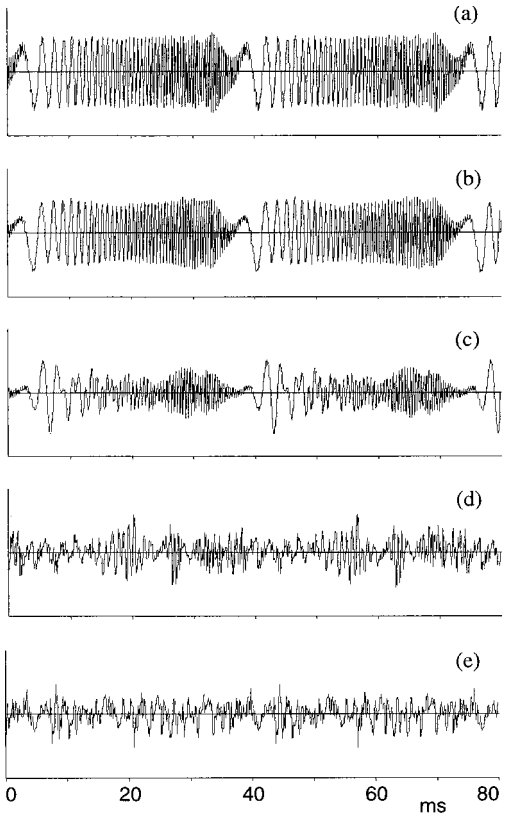


FIG. A1. A comparison between the waveforms of a computed stimulus tone (Schroeder, harmonic) and the sound pressure at a listener's ear in the three different presentation conditions in the listening tests; (a) computer output, (b) inside the headphones, (c) speaker-near condition, (d) speaker-far condition. An additional example showing the sound pressure waveform at about 4 m distance from a high-quality loudspeaker in a moderately reverberant lecture room is shown in (e). All waveforms have been normalized to approximately the same amplitude.

A tempting question for further research is the extent to which a phase-randomization process (''from order to disorder'') contributes to the character of real piano bass tones. First, it would be desirable to study phase effects in more realistic synthesized tones with piano-like spectral dynamics. Also, it would be desirable to conduct measurements on various pianos to determine the stages of phase transformations from string vibrations to radiated tones. Preliminary experiments indicate that some of the phase information in signals with a pronounced peak structure (like the sine-phase tone) can survive surprisingly well in reverberant conditions when radiated from loudspeakers. Thus, room reverberation alone may be insufficient to sweep out any ordered phase information in radiated piano tones.

The question then remains how large the dephasing effect is due to the radiation from the soundboard. An indication that the soundboard probably plays a major role in the dephasing process comes from experiences of preparing a

grand piano with a contact microphone for amplification or recording. The position of the pick-up relative to the bridge is then highly important. As the phase randomization due to the soundboard is missing, pitch glides indicating the dephasing due to string inharmonicity can be recognized in bass notes for certain positions of the contact microphone, in particular, when placed close to the bass bridge. A similar phenomenon may be of importance when listening to normally recorded piano music. If the microphone has been placed at a small distance from the soundboard, the signal may contain ordered phase relations among partials. This may explain in part the experience that recorded piano music may sound quite different when listening over loudspeakers (even of high quality) compared to headphones. A headphone presentation offers much better phase characteristics [Blauert and Laws (1978)] and avoids the reverberation of the listening room.

Whatever may be the relations between the contributions from the different parts of the piano to the phase randomization, the very attack portion of the tone would be the most significant part for communicating the ''ordered'' phase information. The listener needs information about the initial, more or less ordered, phases in order to be able to follow the continuous evolution toward a randomized state. The importance of the attack is further underlined by (1) the known fact that if focuses the attention on the new tone [Moore (1997)], and (2) the presence of a direct sound component not contaminated by multiple room reflections.

Regarding the presence of ''squeals'' that were attributed to temporary phase concentrations, it might be appropriate to revisit the ''wolf tones'' in the piano bass range observed by Knoblaugh (1929, 1944), and attributed by him to the longitudinal string modes. In view of later experimental evidence that the longitudinal modes decay relatively fast,[4] phase concentrations deserve to be investigated as an alternative explanation of the phenomenon.

In summary, our study indicates that the randomization of the relative phases among partials due to inharmonicity is an important factor for the evolution of the timbre in bass tones with wide spectrums. This effect probably needs to be taken into account in a detailed description of the sound generation in the piano bass range. An immediate application could be foreseen in high-quality piano synthesis.

ACKNOWLEDGMENTS

This project was supported by a NATO Postdoctoral Science Fellowship awarded to the first author and a grant from the Natural Sciences and Engineering Research Council of Canada to the third author. We are grateful to Sonya Dal Cin and Jung-Kyong Kim, Queen's University, for running participants and data preparation.

APPENDIX: PHASE RESPONSE

The amplitude and phase responses of the headphones and loudspeakers used in the listening tests will influence the sound presented to the listeners. While the amplitude responses are specified by the manufacturers, data on the phase responses are not readily available. A determination of the phase responses is possible by measuring the group delays [see, e.g., Blauert and Laws (1978)]. For headphones this needs some kind of (standardized) coupler that simulate the enclosed air volume in the ear. The results of such tests indicate that the differences in group delay characteristics (group delay versus frequency) among different manufacturers and models are relatively large [Blauert and Laws (1978); Hirara and Ueda (1990); Hirara (1997)]. Listening tests have shown, however, that in most cases the phase distortion will be below perceptibility [Blauert and Laws (1978)].

A determination of the group delays does not allow an easy interpretation of how severely the computed stimuli are distorted by imperfections in the phase responses. For this reason another approach was taken. The resulting sound pressure waveform for one of the stimuli tones (harmonic, Schroeder) was recorded with a high-quality electret miniature microphone (diam. 8 mm, height 6 mm) at the position of the listener's ear in each of the three presentation conditions in the listening tests; (inside the) headphones, speaker near and speaker far (see Fig. A1). In this way the combined effect of amplitude and phase distortions could be examined, as well as the influence of the room. An additional case was included, showing the sound pressure at a distance of about 4 m from a high-quality hi-fi loudspeaker (Audio Pro A4–14 Mk II) in a moderately reverberant lecture room with a reverberation radius of 1.1 m (reverberation time $T_{60}=0.76$ s, room volume $V=300\,\mathrm{m}^3$).

According to the waveforms in Figs. A1(a)–(d), it is evident that there is progressive deterioration in the presented sounds when going from presentation over headphones (b) to speaker near (c) and speaker far (d), as compared to the computed original signal in (a). In the headphone presentation the sound pressure is essentially a replica of the computed signal and all characteristic features of the stimulus are present, including the characteristic upward frequency sweep. In the speaker-near condition the sweep is still relatively well preserved, although modified by large-amplitude fluctuations. In the speaker-far condition the sweep is barely discernible, being severely deformed by high-frequency reverberation ripples. In the lecture room presentation the stimulus is still more deteriorated. Even at a moderate distance of 4 m from the hi-fi loudspeaker, the sound has no resemblance with the original signal [see Fig. A1(e)], the reason being that the reverberant sound dominates.

[1]SOUNDSWELL™ Signal Workstation, Nyvalla DSP AB, Åldermansvägen 19-21, SE-17148 Solna, Sweden.

[2]SPRUCE™ Listening Test 2.0, Nyvalla DSP AB, Åldermansvägen 19–21, SE-17148 Solna, Sweden.

[3]The phase spectrum was probably an overlooked covariant factor in the Fletcher *et al.* (1962) experiments on the timbral importance of inharmonicity in pianos [see Galembo and Cuddy (1997)].

[4]According to Podlesak and Lee (1987), the decay rate of the longitudinal string component in the lowest bass notes is about 100 dB/s.

Askenfelt, A., and Jansson, E. V. (**1993**). ''From touch to string vibrations. III: String motion and spectra,'' J. Acoust. Soc. Am. **93**, 2181–2196.

Attneave, F. (**1959**). *Applications of Information Theory to Psychology: A Summary of Basic Concepts, Methods and Results* (Holt-Dryden, New York).

Benade, A. (**1976**). *Fundamentals of Musical Acoustics* (Oxford University Press, New York).
Bismark, G., von (**1974**). ''Sharpness as an attribute of the timbre of steady sounds,'' Acustica **30**, 146–159.
Blauert, J., and Laws, P. (**1978**). ''Group delay distortions in electroacoustical systems,'' J. Acoust. Soc. Am. **63**, 1478–1483.
Conklin, H. A., Jr. (**1996**). ''Design and tone in the mechanoacoustic piano. Part 3. Piano strings and scale design,'' J. Acoust. Soc. Am. **100**, 1286–1298.
Crowder, R. G. (**1989**). ''Imagery for musical timbre,'' J. Exp. Psychol. **15**, 472–478.
Exley, K. (**1969**). ''Tonal properties of the pianoforte in relation to bass bridge mechanical impedance,'' J. Sound Vib. **9**, 420–437.
Fastl, H., and Stoll, G. (**1979**). ''Scaling of pitch strength,'' Hear. Res. **7**, 293–301.
Fletcher, H. (**1964**). ''Normal vibration frequencies of a stiff piano string,'' J. Acoust. Soc. Am. **36**, 203–209.
Fletcher, H., Blackham, E. D., and Stratton, R. (**1962**). ''Quality of piano tones,'' J. Acoust. Soc. Am. **34**, 749–761.
Galembo, A., and Askenfelt, A. (**1999**). ''Representation of signals and estimation of spectral parameters by inharmonic comb filters with application to the piano,'' IEEE Trans. Speech Audio Process. **7**, 197–203.
Galembo, A., Askenfelt, A., and Cuddy, L. L. (**1998**). ''Some characteristic features of piano bass tones,'' *Proceedings of the 4th International Acoustical Colloquium*, Zvolen, Slovakia, 15–16 October 1998.
Galembo, A., and Cuddy, L. L. (**1997**). ''String inharmonicity and the timbral quality of piano bass tones: Fletcher, Blackham, and Stratton (1962) revisited,'' Paper presented at the Meeting of the Society for Music Perception and Cognition, MIT, Cambridge.
Gelfand, S. A. (**1998**). Hearing: *An Introduction to Psychological and Physiological Acoustics*, 3rd ed. (Marcel Dekker, New York).
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.
Hansen, V., and Madsen, E. R. (**1974**). ''On aural phase detection,'' J. Audio Eng. Soc. **22**, 10–14.
Hartmann, W. (**1988**). ''Pitch perception and the segregation and integration of auditory entities,'' in *Auditory Function*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 623–645.
Hartmann, W. (**1998**). *Signals, Sound and Sensation* (Springer-Verlag, New York).
Helmholtz, H. L. F., von (**1877**). *Die Lehre von den Tonempfindungen als Physiologische Grundlage für die Theorie der Musik* (English translation by A. J. Ellis, 1885, reprinted by Dover, New York, 1954).
Hesse, H. P. (**1982**). ''The judgement of musical intervals,'' in *Music, Mind and Brain: The Neuropsychology of Music*, edited by M. Clynes (Plenum, New York), Chap. 11.
Hirara, T. (**1997**). ''Physical characteristics of headphones in psychoacoustical experiments,'' J. Acoust. Soc. Jpn. **53**, 798–806 (in Japanese).
Hirara, T., and Ueda, K. (**1990**). ''Investigation of headphones suitable for psychophysical experiments,'' J. Acoust. Soc. Am. Suppl. 1 **87**, S142.
Hundley, T. C., Benioff, H., and Martin, D. W. (**1978**). ''Factors contributing to the multiple rate of piano tone decay,'' J. Acoust. Soc. Am. **64**, 1303–1309.
Iverson, P., and Krumhansl, C. L. (**1989**). ''Pitch and timbre interaction in isolated tones and in sequences,'' J. Acoust. Soc. Am. Suppl. 1 **86**, S58.
Iverson, P., and Krumhansl, C. L. (**1993**). ''Isolating the dynamic attributes of musical timbre,'' J. Acoust. Soc. Am. **94**, 2595–2603.
Knoblaugh, A. F. (**1929**). ''The wolf tone of the pianoforte,'' Ph.D. thesis, University of Cincinnati.
Knoblaugh, A. F. (**1944**). ''The clang tone of the pianoforte,'' J. Acoust. Soc. Am. **16**, 102.
Lichte, W. H., (**1941**). ''Attributes of complex tones,'' J. Exp. Psychol. **28**, 455–480.
Licklider, J. C. R. (**1957**). ''Effects of changes in the phase pattern upon the sound of a 16-harmonic tone,'' J. Acoust. Soc. Am. **29**, 780(A).
Martin, D. W. (**1947**). ''Decay rates of piano tones,'' J. Acoust. Soc. Am. **19**, 535–541.
Mathes, R. C., and Miller, R. L. (**1947**). ''Phase effects in monaural perception,'' J. Acoust. Soc. Am. **19**, 780–797.
Moore, B. C. J. (**1977**). ''Effects of relative phase of the components on the pitch of three-component complex tones,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London).
Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing*, 4th ed. (Academic, Cambridge University, UK).
Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1985**). ''Relative dominance of individual partials in determining the pitch of complex tones,'' J. Acoust. Soc. Am. **77**, 1853–1860.
Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1985**). ''Thresholds for the detection of inharmonicity in complex tones,'' J. Acoust. Soc. Am. **77**, 1861–1867.
Nordmark, J. O. (**1978**). ''Frequency and periodicity analysis,'' in *Handbook of Perception, Volume 4: Hearing* (Academic, New York).
Patterson, R. D. (**1987**). ''A pulse ribbon model of monaural phase perception,'' J. Acoust. Soc. Am. **82**, 1560–1586.
Pierce, J. R. (**1992**). *The Science of Musical Sound* (Freeman & Co., New York).
Pincus, S. (**1991**). ''Approximate entropy as a measure of system complexity,'' Proc. Natl. Acad. Sci. U.S.A. **88**, 2297–2301.
Podlesak, M., and Lee, A. R. (**1987**). ''Longitudinal vibrations in piano strings,'' J. Acoust. Soc. Am. Suppl. 1 **81**, S61.
Podlesak, M., and Lee, A. R. (**1989**). ''Effect of inharmonicity on the aural perception of initial transients in low bass tones,'' Acustica **68**, 61–66.
Rayleigh (**1894**). *The Theory of Sound* (MacMillan & Co., Ltd., London), Vol. 1, pp. 298–301.
Reinholdt, A., Jansson, E., and Askenfelt, A. (**1987**). ''Analysis and synthesis of piano tone,'' J. Acoust. Soc. Am. Suppl. 1 **81**, S46.
Rimski-Korsakov, A., and Diakonov, N. (**1952**). *Musical Instruments: Methods of Research and Calculation* (in Russian) (Rosgizmestprom Press, Moscow).
Schouten, J. F. (**1940**). ''The residue and the mechanism of hearing,'' Proc. K. Ned. Akad. Wet. **43**, 991–999.
Schouten, J. F., Ritsma, R. J., and Cardozo, B. L. (**1962**). ''Pitch of the residue,'' J. Acoust. Soc. Am. **34**, 1418–1424.
Schroeder, M. R. (**1970**). ''Synthesis of low-peak-factor signals and binary sequences with low autocorrelation,'' IEEE Trans. Inf. Theory **IT-16**, 85–89.
Schroeder, M. R. (**1975**). ''Models of hearing,'' Proc. IEEE **63**, 1332–1350.
Schuck, O. H., and Young, R. W. (**1943**). ''Observations on the vibrations of piano strings,'' J. Acoust. Soc. Am. **15**, 1–11.
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.
Singh, P. G. (**1987**). ''Perceptual organization of complex tone sequences: A tradeoff between pitch and timbre,'' J. Acoust. Soc. Am. **82**, 886–899.
Singh, P. G., and Hirsh, I. J. (**1992**). ''Influence of spectral locus and *F*0 changes on the pitch and timbre of complex tones,'' J. Acoust. Soc. Am. **92**, 2650–2661.
Terhardt, E. (**1972**). ''Zur Tonhöhenwahrnehmung von Klängen,'' Acustica **26**, 133–199.
Wessel, D. L., Bristow, D., and Settel, Z. (**1987**). ''Control of phrasing and articulation in synthesis,'' *Proceedings of the International Computer Music Conference* (Computer Music Association, San Francisco), pp. 108–116.
Wightman, F. L. (**1973a**). ''Pitch and stimulus fine structure,'' J. Acoust. Soc. Am. **54**, 397–406.
Wightman, F. L. (**1973b**). ''The pattern-transformation model of pitch,'' J. Acoust. Soc. Am. **54**, 407–416.

Vocal intensity in falsetto phonation of a countertenor: An analysis by synthesis approach

Kenneth Tom
Department of Speech Communication, California State University Fullerton, 800 North State College Boulevard, Fullerton, California 92831

Ingo R. Titze
Department of Speech Pathology and Audiology, National Center for Voice and Speech, University of Iowa, Iowa City, Iowa 52242

(Received 22 August 2000; revised 19 June 2001; accepted 26 June 2001)

An analysis by synthesis paradigm was implemented to model glottal airflow and vocal tract acoustics for the falsetto phonation of a trained countertenor. Changes in vocal intensity were measured as a function of subglottal pressure, open quotient of the time-varying glottal airflow pulse, and formant tuning. The contributions of laryngeal adduction (open quotient of the glottal flow pulse) and of formant tuning to intensity change were derived from modeled data. The findings were: (1) Subglottal pressure accounted for almost 90% of the variation in SPL in falsetto phonation. (2) The open quotient of the glottal flow pulse was remarkably constant in these falsetto phonations, and thus did not affect vocal intensity significantly. (3) Formant tuning occurred in two out of nine possibilities for the vowel /ɑ/. These instances did not support the concept of systematic exploitation of formant tuning. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1396331]

PACS numbers: 43.75.Rs, 43.70.Aj, 43.70.Jt [RDA]

I. INTRODUCTION

Physiologically, the voice of an adult male is capable of producing falsetto register, with varying degrees of vocal intensity, in the upper part of the fundamental frequency range (Colton and Hollien, 1972; Hollien, 1974, 1977). Although there is a range of fundamental frequencies that can be produced in either falsetto or modal (chest) register, the highest fundamental frequencies are normally produced in falsetto register. Loud, high-pitched falsetto phonation is therefore considered one extreme of the functional range of the vocal mechanism (Titze, 1988, 1994; Welch *et al.*, 1988).

Culturally speaking, vocalization in falsetto register is not a rare phenomenon. In speech, falsetto voice can be used as a phonemic contrast, or as an element of prosody (Laver, 1980). Emotions such as surprise, fear, happiness, and anguish are expressed vocally in the form of sighs, squeals, giggles, and other outbursts of high-pitched falsetto vocalization (Scherer, 1995). Falsetto is also used by males to imitate women's or children's voices.

Singing in falsetto is part of the musical practice of many of the world's cultures, occurring in contexts as varied as the Beijing opera, the traditional chanting of Pygmy tribes in central Africa, Alpine yodeling, Hawaiian folk singing, and Western pop music (Malm, 1967; Alley, 1982; MacKerras, 1983). In the sacred music of England, it had become tradition by the 17th century to have the alto voice parts of a choir sung by countertenors, i.e., adult tenors or baritones using falsetto voice (Giles, 1982). In the last two decades, there has been a renaissance in the art of the countertenor vocal soloist in Europe and North America, especially in the performance of early music and Baroque opera (Gurewitsch, 1999). Countertenor soloists are adult males with tenor or baritone/bass speaking voices, who generally use a falsetto-based vocal technique for singing alto or soprano parts (Giles, 1982; Welch *et al.*, 1988).

Falsetto phonation has also been used by voice therapists and teachers of classical singing voice for rehabilitation and training (Brown, 1996). Voice therapists have used falsetto for treating functional voice disorders such as vocal nodules, ventricular phonation, and muscle tension dysphonia (Boone and McFarlane, 1994; Colton and Casper, 1990; Perlman, 1992). Singing teachers have used falsetto phonation with male students to develop a mixed register (''head voice'') and thereby decreased forceful productions of high-pitched vocalization (Frisell, 1964; Reid, 1978).

Although control of pitch in falsetto register has been discussed amply (van den Berg, 1963; Hirano *et al.*, 1969; Titze, 1994), there are few comprehensive studies regarding physiologic control of vocal intensity in falsetto. Most studies examining physiologic control of vocal intensity have avoided high-pitched or falsetto phonation (Holmberg *et al.*, 1988; Stathopoulos and Sapienza, 1993a; Holmberg *et al.*, 1995). This is to a large degree due to practical and theoretical limitations of measurement and analysis techniques currently available to separate source from vocal tract characteristics.

In a single-subject study, like ours to be described, Ishiki (1964) found mean airflow to be the most salient variable for vocal intensity change. At a given falsetto pitch, he found glottal resistance to be almost constant, and his data suggested that intensity was proportional to the 3.3 power of flow. Colton (1973) found the intensity range produced in falsetto register to be much smaller than the intensity range in modal register. The difference between most and least intense phonations in falsetto ranged from 13 dB SPL for the nonsingers to 18 dB SPL for the singer subjects, which was about half the range found in modal register.

The regulation of vocal intensity can be understood as coordinated contributions of possibly three major components: subglottal pressure, glottal adduction, and vocal tract resonance (Titze and Sundberg, 1992). The goal of this study was to describe changes in vocal intensity in falsetto phonation as a function of changes in these three variables. To circumvent the limitations of having to measure and remove formant structure directly in high-frequency phonations, an analysis by synthesis approach, patterned after Titze, Mapes, and Story (1994), was employed. This approach made explicit use of three-dimensional measurement of the vocal tract geometry (Tom *et al.*, 2001). Since this is very time-consuming and energy-draining for the subject, only a single singer was studied.

High-pitched phonations are problematic for determining formant frequencies and bandwidths because there are few harmonics with which to ''sample'' the vocal tract spectrum. In addition, the amplitudes of successive harmonics in the spectrum of falsetto phonation drop off rapidly, at a rate of 16 dB/octave or more, providing reduced levels of acoustic information for a linear inverse filter above 1000 Hz. This is the traditional methodological problem.

II. METHODS

This study implemented an analysis by synthesis approach as a practicable alternative to linear inverse filtering. This involved ''fitting'' a computer model to a particular subject by (1) incorporating cross-sectional vocal tract area functions derived from direct imaging of the subject's vocal tract; (2) adjusting glottal parameters of the model such that closure conditions were similar to those observed via videostroboscopic images and electroglottograms of the subject's phonations, and (3) using measured subglottal pressure and fundamental frequency as input values for simulation. The model simulated oral pressure, oral flow, and vocal fold contact area, all of which were compared to those of the subject.

A. Subject and phonatory conditions

The subject was a 45-year-old male with extensive training in Western classical singing, including 12 years of vocal study as a baritone and 6 years of study as a countertenor. An active performer and voice teacher, he had sung as a countertenor for the past 11 years, performing early music with a falsetto-based technique to vocalize in the alto range. The subject's medical and recent health history were unremarkable. He was a native speaker of general North American English. The subject's fundamental frequency range spanned from 69 Hz ($C\sharp_2$) to 392 Hz (G_4) in chest register and from 147 Hz (D_3) to 587 Hz (D_5) in falsetto register.

Data recordings were made under eight phonatory conditions (see Table I). There were three pitch levels, separated by a distance of a musically perfect fourth (low pitch, 262 Hz; medium pitch, 349 Hz; high pitch, 466 Hz). Each pitch was sung at two intensity levels, *mezzo piano* and *fortissimo*. In addition, two chest register conditions were produced for control, one at a comfortable and one at a loud intensity level. The pitches of these phonations were not prescribed, but were selected by the subject as B-flat$_2$ (117 Hz) and D_3 (149 Hz), respectively.

TABLE I. Phonatory conditions for data collection, vowel /ɑ/.

Pitch/Condition	Register	Pitch (F_0)	Loudness
B-flat$_4$ ff	Falsetto	B-flat$_4$ (466 Hz)	Very loud (ff)
Bflat$_4$ mp	Falsetto	B-flat$_4$ (466 Hz)	Moderately soft (mp)
F_4 ff	Falsetto	F_4 (349 Hz)	Very loud (ff)
F_4 mp	Falsetto	F_4 (349 Hz)	Moderately soft (mp)
C_4 ff	Falsetto	C_4 (262 Hz)	Very loud (ff)
C_4 mp	Falsetto	C_4 (262 Hz)	Moderately soft (mp)
Speech	Chest	D_3 (147 Hz)	Very loud
Speech	Chest	B-flat$_4$ (117 Hz)	Comfortable loudness

B. Acoustic recordings

Microphone, sound-level meter, and electroglottographic (EGG) signals were recorded simultaneously on a digital instrumentation tape recorder (Sony PC108-M). Six tokens of the vowel /ɑ/, prolonged for 2 to 3 s, were recorded at each of the eight conditions, a total of 48 utterances. Recording bandwidths were 20, 10, and 10 kHz, respectively. Audio signals were transduced with a prepolarized condenser microphone (AKG C451 EB with a CK 22 capsule). Preamplification and amplification were provided by an AKG N-62E phantom power source and a Digital Sound Corporation (DSC 240) amplifier, respectively. Intensity level was recorded from the output of a digital sound-level meter (Bruel & Kjaer 2230) fitted with a condenser microphone (Bruel & Kjaer 4155) set for slow response and linear weighting. Calibration tones were produced at four intensity levels, with a reference sound source at an approximate frequency of 570 Hz. The intensity levels of these tones were monitored with the digital sound-level meter. In addition, a Synchrovoice research electroglottograph (EGG) was used to transduce time-varying vocal fold contact area.

Recordings were made in a sound-treated booth (IAC) approximately $3\times3\times3.8$ m. In order to record data from phonations analogous to those produced during imaging (described below), the acoustic recordings were made with the subject in a supine position. The subject was reclined in a dental chair, parallel to the floor, with a built-in headrest adjusted for the subject's head stability and comfort. The sound-level meter and microphone were mounted in a fixed position on a microphone stand at a distance of 0.5 m, directly above the subject's lips. EGG electrodes were placed on the skin surface superficial to the laminae of the thyroid cartilage. Appropriate placement was verified by monitoring signal outputs on an oscilloscopic display. Pitch cues for the different pitch conditions were given from an electronic keyboard (Casio MT-35) before each set of six tokens of phonation.

C. Aerodynamic recordings

Transduction of wide-bandwidth oral airflow and estimates of subglottal pressure were accomplished using established techniques (Rothenberg, 1973; Smitheran and Hixon, 1981; Löfqvist, Carlborg, and Kitzing, 1982; Holmberg, Hill-

man, and Perkell, 1988; Kitajima and Fujita, 1990). Phonatory tasks were strings of seven smoothly and continuously uttered repetitions of the syllable /pɑ/ at a rate of approximately 1.5 syllables per second. Each syllable string was repeated five times for a total of 35 syllable repetitions at each of the phonatory conditions. In order to avoid beginning and/or end effects, only the three center syllables of each string were analyzed, i.e., three center syllables from each of the five strings resulted in 15 data points per phonatory condition. Subglottal pressure was inferred from the intraoral pressure during mouth occlusion in the /pɑ/ syllable. This intraoral air pressure was sensed with a small metal tube (approximately 20 mm in length, with an inner diameter of 1 mm) inserted into a small opening in a circumferentially vented face mask (Glottal Enterprises MA-1). When the mask was in place, the proximal end of the metal tube entered the corner of the mouth, the distal end being connected to a low-bandwidth differential pressure transducer (Glottal Enterprises PTL-1). The frequency response of this transducer was rated flat up to 30 Hz. Pressure signals were calibrated with a series of six pressure levels, ranging from 0.0–2.49 kPa as monitored with a Dwyer U-tube manometer.

Oral airflow was transduced with a high-resolution pneumotachograph attached to the circumferentially vented face mask (mentioned above). This transducer system (Glottal Enterprises MS 100 A2, Glottal Enterprises PTW-1) had a flat frequency response to approximately 1000 Hz. Flow signals were calibrated by using a series of nine flow levels over a range of 0–2198 ml/s as monitored by a rotameter (Gilmont #5 Flowmeter, E-5930).

Intensity level and EGG signals were obtained simultaneously with these aerodynamic signals, as described above. The aerodynamic signals were recorded in the same environment as the acoustic signals, with the subject in supine position. The subject held the flow mask firmly over his nose and mouth to ensure a good seal during the entire phonation of each pitch/loudness condition. The mask was removed from the subject's face between each condition and the flow and pressure voltages of the mask electronics readjusted to zero to minimize dc signal drift.

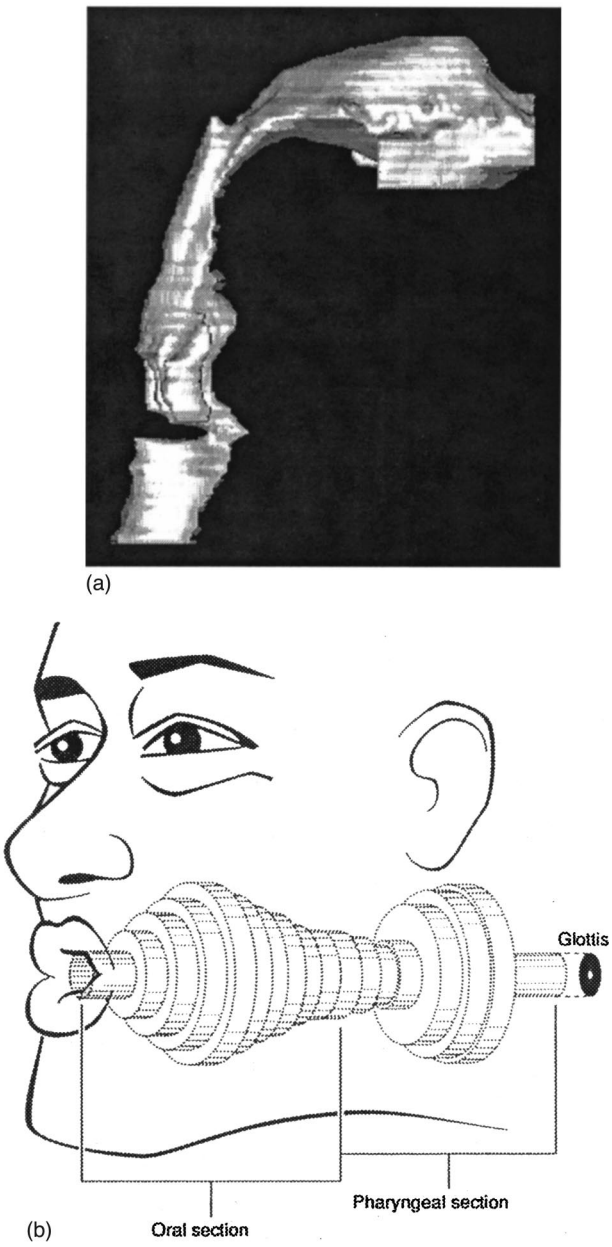


FIG. 1. Sketch of vocal tract. (a) Electron beam computerized tomography and (b) Computer-simulated cylindrical tube version.

D. Videostroboscopy

Laryngeal videostroboscopic images were made using a Machida LY-C30 rigid laryngoscope with a Bruel & Kjaer Rhino-Larynx Stroboscopic light source (type 4914). Images were recorded on S-VHS videotape with simultaneous sound for reference purposes. With the subject comfortably seated upright in a dental chair, the rigid scope was introduced orally without the use of anesthetic spray. The eight phonatory conditions were recorded starting with low-pitched comfortable speech and loud speech in chest register, and then the low, medium, and high-pitch conditions in falsetto. Phonations were made on an /i/ vowel in order to accommodate the laryngoscope in the oral cavity. (It is difficult to see the vocal folds during an /ɑ/ vowel, because the position of the tongue and epiglottis in the pharynx obscures their view.)

E. Vocal-tract imaging

High-resolution images of the vocal tract during phonation of an /ɑ/ vowel on the eight pitch and loudness conditions were acquired using electron beam-computed tomography (EBCT). Each series of contiguous axial scans was reconstructed as a volumetric (3D) image set of the vocal tract during a particular phonatory condition. Cross-sectional area functions were derived from measurements of the 3D reconstructed vocal tracts. The specific techniques used for image acquisition and volume reconstruction and measurement of vocal tract cross-sectional areas have been described by Tom (1996) and Tom *et al.* (2001). An example of a volumetric airway EBCT scan is shown in Fig. 1(a) for the /ɑ/ vowel.

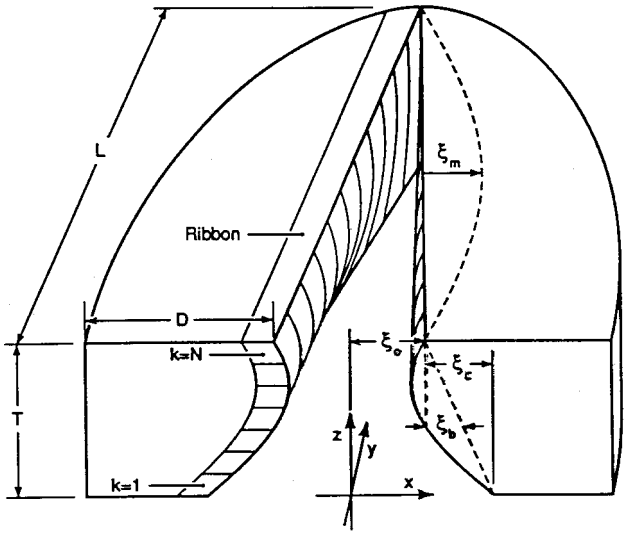


FIG. 2. Sketch of vocal fold model with geometric parameters.

F. Analysis by synthesis

A kinematic model of vocal fold vibration (Titze, 1984) was combined with a one-dimensional, soft-wall, lossy wave reflection algorithm for sound propagation in the vocal tract (Liljencrants, 1985; Story, 1995). This analysis by synthesis paradigm has been used in previous experiments (Titze, 1984; Titze, Mapes, and Story, 1994). The kinematic model represents a first level of interaction between the simulated voice source and vocal tract. It calculates pressure and flow behaviors through a dynamic three-dimensional glottis (Fig. 2), coupled to cylindrical tube approximations of the subglottal and supraglottal vocal tract. Subglottal and supraglottal acoustic pressures affect glottal airflow through the dynamic three-dimensional glottis, which is modeled as

$$g(y,z,t)=\mathrm{Max}\{0,2[\bar{\xi}(y,z)+\xi(y,z,t)]\}, \tag{1}$$

where $\bar{\xi}$ is the postural (prephonatory) component and ξ is the dynamic (oscillatory) component. The postural component is

$$\bar{\xi}=(1-y/L)[\xi_0+(\xi_c-4\xi_b z/T)(1-z/T)], \tag{2}$$

where L is vocal fold length, T is the vocal fold thickness, ξ_0 is the half width of the glottis at the vocal process (Fig. 2), ξ_c is the net convergence of the medial surface of the vocal folds, and ξ_b is the bulging of the medial surface. The time-varying component is written as

$$\xi=\xi_m \sin(\pi y/L)\left[\sin\omega t-\frac{\omega z}{c}\cos\omega t\right], \tag{3}$$

where ξ_m is the displacement amplitude of the center ($y=L/2$) of the vocal folds, ω is the angular frequency, and c is a vertical surface wave velocity in the vocal fold tissues (about 1 m/s). With this model, oscillation is driven (not self-sustained) and Eq. (3) represents two principal modes, a compressional mode (first term in brackets) and a shear mode (second term in brackets). The airflow is calculated on the basis of incident and reflected pressures above and below the glottis in an interactive fashion (Titze, 1984). Parameters for variation are subglottal pressure P_s, fundamental frequency $F_0=\omega/2\pi$, and the glottal shaping parameters ξ_0, ξ_c, and ξ_b. The amplitude of vibration ξ_m was computed by an empirical rule, based on unpublished data on excised human larynges in our laboratory

$$\xi_m=0.1L\left(\frac{P_s-P_{th}}{P_{th}}\right)^{1/2}. \tag{4}$$

In this expression, P_{th} is the oscillation threshold pressure, given by Titze (1994) as

$$P_{th}=0.14+0.06(F_0/120)^2. \tag{5}$$

The vocal tract was simulated with an even number of contiguous cylinders, the length of which depended on the sampling frequency and overall vocal tract length. At the sampling frequency used in this study, 44 100 Hz, each vocal tract section was 0.396 825 cm long. With the vocal tract normalized to 44 sections, this resulted in an overall vocal tract length of 17.46 cm. Figure 1(b) shows how the cylindrical sections are used in the model (fewer than 44 are shown for simplicity).

To match the model to measured data, approximately 25 cycles of each EGG file, taken from the center portion of the /ɑ/ token, were first used to estimate the open quotient for the particular phonation. This became a seed value for the synthesis program. Computer simulations were completed for each of the six tokens of each of the six falsetto pitch/loudness conditions, as well as the two chest register conditions, for a total of 48 cases. Vocal tract area functions, fundamental frequency, and lung pressure, as measured from the subject, remained constant inputs to the model. Because we know it to be a ''register'' parameter, ξ_b, the bulging of the medial surface, was kept constant for each register condition. Register was held constant by the subject for each phonation, as specified for each phonatory condition. Other glottal shaping parameters, including ξ_0 (the vocal process glottal half width), ξ_c (net glottal convergence), and ξ_p (posterior glottal gap size) were optimized (MINPAC, free software from Argonne National Laboratory, Chicago) by fitting the FFT magnitude spectrum of the microphone signal to the model output. Optimization consisted of minimizing the least-squared error as follows:

$$\mathrm{Error}=\frac{1}{n}\sum (A_n-\bar{A_n})^2=\mathrm{minimum}, \tag{6}$$

where A_n is the nth harmonic amplitude of the model and $\bar{A_n}$ is the corresponding harmonic amplitude of the measured data. Harmonics up to 4000 Hz were included in the sum.

Figure 3 illustrates a comparison between a modeled spectrum and the spectrum of a particular phonation from the subject. (The spectrum is a perfect line spectrum because a Fourier series of a single cycle of the model was calculated and matched to the data.) The vertical lines represent the subject's phonations; the asterisks represent values as simulated by the model.

To optimize the relevant source parameters (adduction, convergence, and posterior gap), videostroboscopic images were first examined to get a range of seed values. For the vocal process glottal half width ξ_0, values between 0.0 and 0.02 cm were chosen as trials. Larger values gave basically

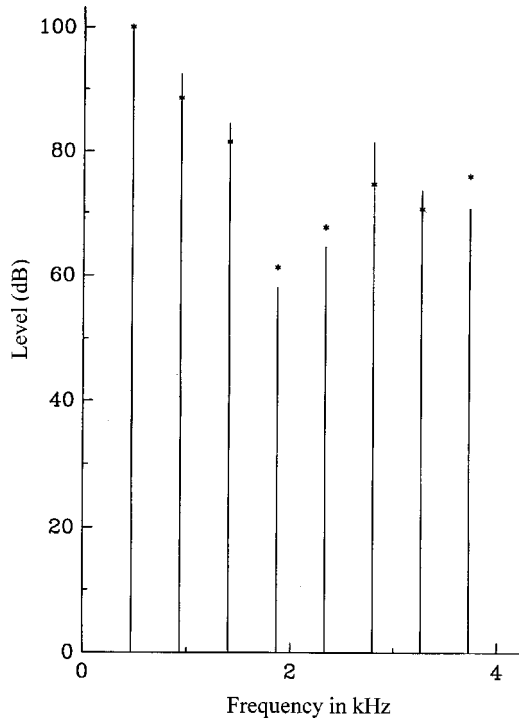


FIG. 3. Multitude spectrum for B-flat$_4$, very loud, token 5. Lines are measured data and asterisks are the model.

sinusoidal wave shapes. For the net glottal convergence ξ_c, values between 0.4 and 0.9 were chosen. Lower values, e.g. 0.1 or 0.2, created a relatively square vocal fold shape, known to be associated with the greater thyroarytenoid muscle contraction, as in chest voice. A value of 0.2 is most often used in simulations of chest register speech. The higher values yielded more rounded and triangular vocal fold shapes (from a coronal perspective). The glottal parameter, ξ_p, posterior glottal gap size, was quantified in terms of an average diameter from the vocal process to the posterior wall. The seed values chosen ranged from 0.000 to 0.075 cm. Each optimization run for a given token of phonation produced 192 trials, using the entire matrix of combinations of glottal parameter values.

When there were several candidates with similarly low least-squared error, secondary matching criteria were used to rule out poorer candidates. Since time waveforms of microphone, EGG, and oral flow were available, visual comparisons were made (Fig. 4). The best fit was the one that provided both a good spectral match and a good visual match (which included the correct phase information in the signals). Table II shows the measured values and optimized parameter input values used to simulate the best-fit falsetto phonations.

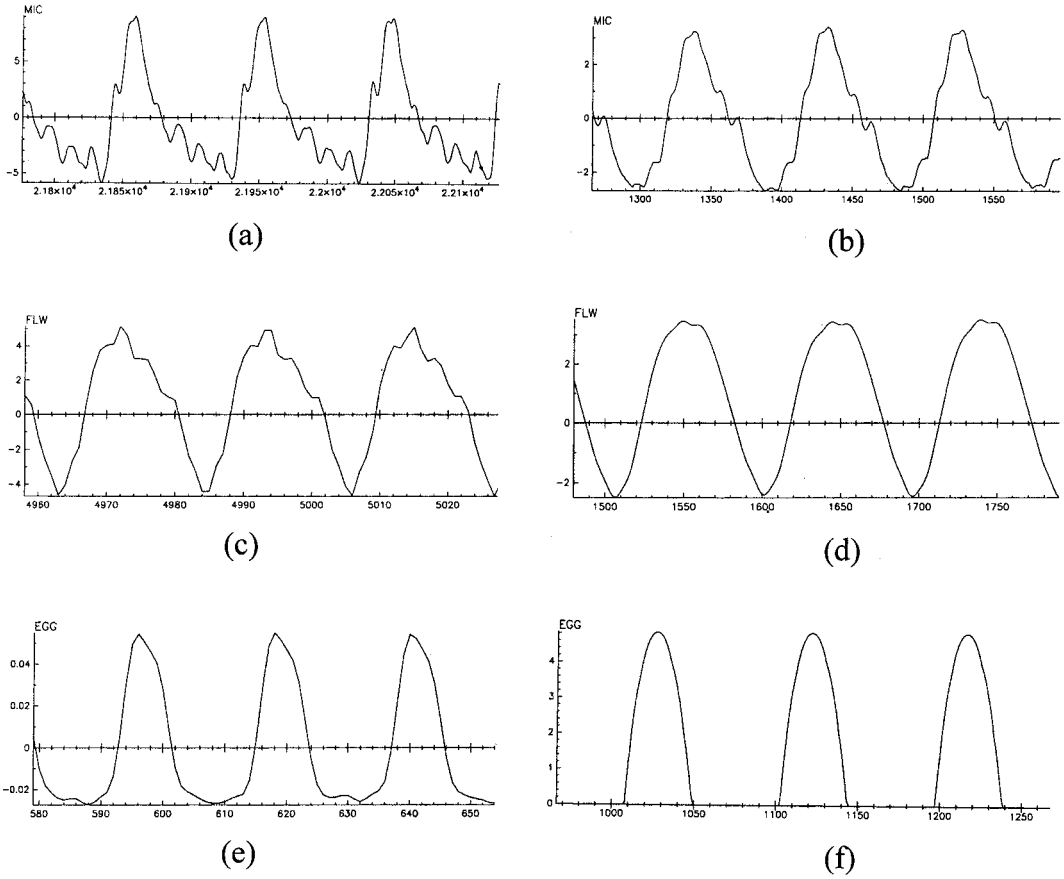


FIG. 4. Comparison of measured (left panels) and simulated (right panels) waveforms for B-flat$_4$, very loud, token 5: (a) and (b) are microphone; (c) and (d) are oral flow; (e) and (f) are EGG.

TABLE II. Measured and optimized input values for parameters used for best-fit simulations of falsetto phonations.

Pitch & loudness condition and token number	Glottal half-width (cm)	Net glottal convergence	Posterior glottal gap (diameter in cm)	Total squared error (number of optimized harmonics)
Bff_4 (Fundamental frequency=466 Hz; subglottal pressure=2.36 kPa)				
1	0.0175	0.4	0.075	410 (8)
2	0.0175	0.4	0.075	249 (8)
3	0.0050	0.9	0.075	175 (8)
4	0.0100	0.7	0.075	148 (8)
5	0.0100	0.7	0.075	127 (8)
6	0.0100	0.9	0.050	128 (8)
Bmp_4 (Fundamental frequency=466 Hz; subglottal pressure=1.76 kPa)				
1	0.0125	0.4	0.075	533 (8)
2	0.0125	0.9	0.050	642 (8)
3	0.0075	0.6	0.075	402 (8)
4	0.0175	0.7	0.050	441 (8)
5	0.0175	0.7	0.075	574 (8)
6	0.0075	0.6	0.075	379 (8)
Fff_4 (Fundamental frequency=349 Hz; subglottal pressure=1.72 kPa)				
1	0.0125	0.9	0.050	1460 (11)
2	0.0125	0.9	0.075	916 (11)
3	0.0125	0.9	0.075	975 (11)
4	0.0125	0.9	0.075	1134 (11)
5	0.0125	0.9	0.075	1215 (11)
6	0.0125	0.9	0.075	1078 (11)
Fmp_4 (Fundamental frequency=349 Hz; subglottal pressure=1.07 kPa)				
1	0.0200	0.6	0.050	564 (11)
2	0.0200	0.7	0.075	634 (11)
3	0.0200	0.7	0.050	784 (11)
4	0.0200	0.7	0.075	808 (11)
5	0.0200	0.6	0.050	1008 (11)
6	0.0125	0.7	0.075	411 (11)
Cff_4 (Fundamental frequency=262 Hz; subglottal pressure=1.25 kPa)				
1	0.0200	0.7	0.075	1029 (15)
2	0.0200	0.9	0.075	658 (15)
3	0.0200	0.5	0.050	1459 (15)
4	0.0200	0.9	0.075	1543 (15)
5	0.0200	0.9	0.075	2291 (15)
6	0.0200	0.7	0.075	1401 (15)
Cmp_4 (Fundamental frequency=262 Hz; subglottal pressure=0.79 kPa)				
1	0.0175	0.7	0.075	633 (15)
2	0.0175	0.7	0.075	409 (15)
3	0.0200	0.9	0.050	759 (15)
4	0.0200	0.7	0.075	1640 (15)
5	0.0200	0.7	0.075	1182 (15)
6	0.0200	0.9	0.050	438 (15)

Exact matches between measured and simulated data for acoustic (oral pressure), aerodynamic (mean oral flow, dynamic glottal airflow), and EGG signals were not expected from the computer simulations. The isomorphism of the model is compromised to a degree by simplifications in the parametrization of variables made in order to make the model practical with regard to computation time. Additionally, the simplified nature of its glottis cannot completely capture the complex dynamic geometry of a given subject's glottis during phonation, and hence, related airflow simulations.

The simulation for vocal fold contact area (EGG) in SPEAK was calculated by determining a differential contact length, and then integrating the contact area of the medial surface over the differential thickness of each vocal fold layer. The parameter of vocal fold thickness is not specified and thus only approximated with the rule that vocal fold thickness is inversely related to length, using data from Hollien (1960) for nominal thickness. Another possible source of error lies in the difficulties encountered transducing the EGG signal in falsetto phonations. Relatively small changes in position of the thyroid cartilage can reduce signal strength/accuracy significantly. The minimal vertical vocal fold contact of small-amplitude falsetto phonations can also reduce

signal amplitude, and thus lower signal/noise ratios, further reducing the accuracy of the measured signal.

III. RESULTS AND DISCUSSION

A. Laryngeal imaging

Videotaped images of laryngeal videostroboscopy of the subject's phonations at all eight phonatory conditions were visually inspected with regard to glottal configuration and vibratory behavior. It was concluded that all phonations were produced with bilateral symmetry. The subject's falsetto phonations were produced with an elongated glottis without any noticeable glottal gap at the vocal processes of the arytenoid cartilages. Vocal fold contact during vibration was relatively brief, but complete along the length of the membranous folds. The entire arytenoid cartilage appeared to be adducted.

During chest register phonations, the vocal folds appeared bulkier in mass and shorter in length. The amplitude of vibration (lateral excursion) was much greater than in falsetto phonations. In both speech phonation conditions (comfortable and loud intensities), glottic closure was less firm at the vocal processes. In some of these samples, a small posterior glottal gap was evident. Closure along the glottal midline during vibration was mostly complete and closed phases were longer in chest voice than in falsetto.

B. Subglottal pressure

The subject's subglottal pressures at speech conditions (0.57 and 0.95 kPa) were within normal limits (Holmberg *et al.*, 1988). His falsetto productions were associated with subglottal pressures ranging from 0.79 to 2.36 kPa, which are also within the range of subglottal pressures produced by other tenors (Titze and Sundberg, 1992) and baritones (Sundberg *et al.*, 1999). The linear correlation between SPL and subglottal pressure, i.e., *excess pressure over threshold* $(P_s - P_{th})$ plotted in logarithmic units, is shown in Fig. 5(a). Data points are for all falsetto conditions (15 data points per each of six conditions). The correlation coefficient for the regression line, R^2, was 0.88. Thus, it appears that almost 90% of the variation in SPL in falsetto can be explained by changes in subglottal pressure. The slope of the regression line corresponded to an increase of 8.13 dB per doubling of *excess pressure over threshold*, in agreement with the 8–9-dB increase in SPL predicted by Titze and Sundberg (1992) for their nonfalsetto singers and speakers.

Titze and Sundberg (1992) have found that the effect of F_0 on threshold pressure becomes clear when *fractional excess pressure over threshold* $(P_s - P_{th})/P_{th}$ [the term raised to the $\frac{1}{2}$ power in Eq. (4)], is plotted against SPL. Data for their tenor subjects separated into a vertical stack of three nearly parallel lines, the increases in value of the *y* intercepts corresponding to higher phonation threshold pressures required by increasing fundamental frequency. Sundberg *et al.* (1999) found that data for two of their five baritone subjects followed this pattern. When plotted as described above, the data for all of their subjects resulted in linear regressions with slopes ranging from 6.3 to 13.66. This corresponded to increases in SPL from 4.37 to 9.47 dB per doubling of *fractional excess pressure over threshold*.

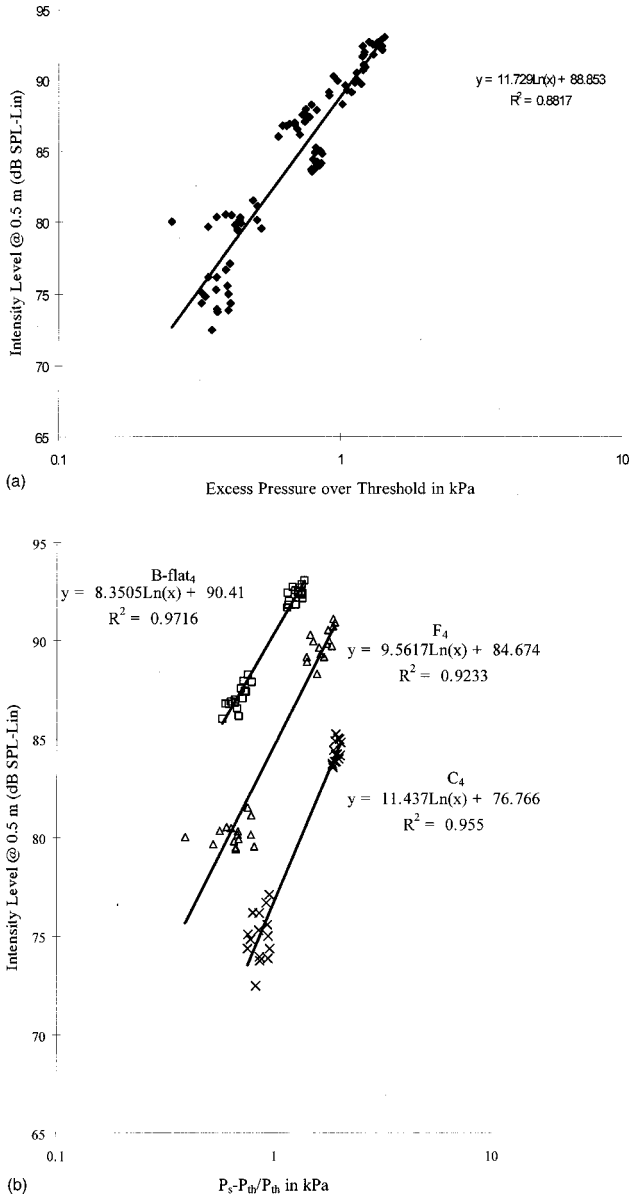


FIG. 5. (a) Sound-pressure level versus excess pressure over threshold, falsetto pitches. (b) Sound-pressure level versus fractional excess pressure over threshold, falsetto pitches.

Figure 5(b) shows the linear correlation between *fractional excess pressure over threshold* $(P_s - P_{th})/P_{th}$, plotted in logarithmic units, and SPL for our countertenor subject. The data follow the pattern predicted by Titze and Sundberg (1992). The slopes of the three lines were 8.35, 9.56, and 11.44 for the 466, 359, and 262-Hz falsetto conditions, respectively. This corresponded to increases of 5.79, 6.63, and 7.93 dB per doubling of $(P_s - P_{th})/P_{th}$, well within the range of increases found by Sundberg *et al.* (1999). Since $(P_s - P_{th})/P_{th}$ determines the amplitude of vibration of the vocal folds (Titze and Sundberg, 1992), it is expected that it also determines the SPL and the peak flow.

C. Glottal adduction and airflow

Open quotients associated with falsetto phonation (obtained by model matching) for the subject varied only between 0.848 to 0.865 (Fig. 6). This is a surprisingly low

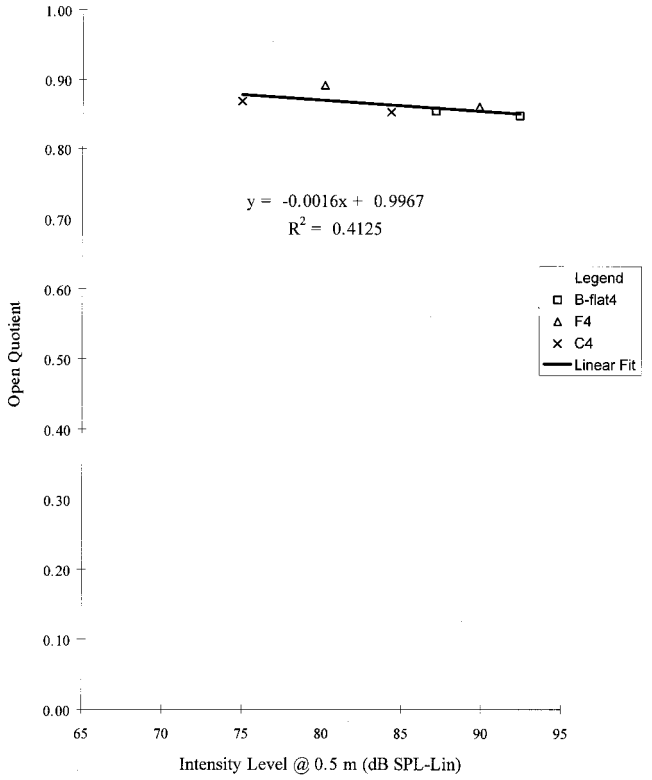


FIG. 6. Open quotient versus sound-pressure level, falsetto pitches.

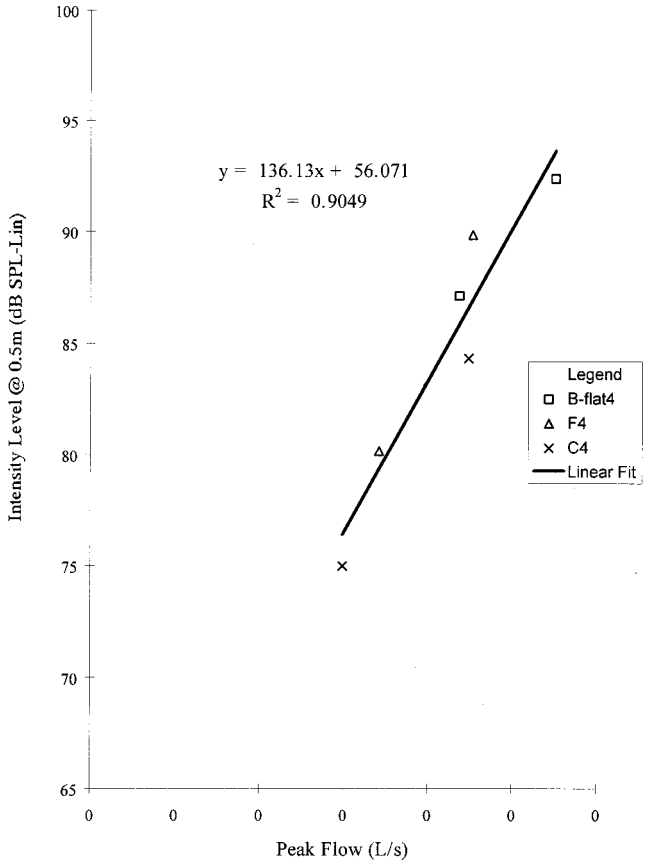


FIG. 7. Sound-pressure level versus peak flow, falsetto pitches.

variation, given that, in general, this measure can range from 0.2 in pulse phonation to 1.0 in breathy phonation (Timcke, von Leden, and Moore, 1959). It may be that the subject intentionally avoided using changes in glottal adduction and thereby stabilized the open quotient, in order to keep the timbre constant. Another possible interpretation is that he was not willing to risk an involuntary register transition (''cracking'') into the heavier chest register. A register transition was certainly possible at two of the three falsetto pitch levels (low and medium), which he could also produce in chest register. A primary goal of some singers' technique is to keep the vocal timbre as stable as possible (for some musical styles). Most likely, the stability of the glottal adduction is related to that technical goal. The obvious outcome in this singer was that adduction (glottal valving) did not contribute to changes in SPL.

Both peak flow and maximum flow declination rate appear to be strongly related to vocal intensity (Figs. 7 and 8). This has been repeatedly observed (Holmberg *et al.*, 1988; Stathopoulos and Sapienza, 1993a,b; Sundberg, Titze, and Scherer, 1993; Sundberg *et al.*, 1999). Three possible strategies for increasing the maximum flow declination rate have been pointed out by Sundberg *et al.* (1993). The first is increasing peak flow, i.e., raising the amplitude of the glottal pulse by increasing subglottal pressure. This was the primary strategy used in the falsetto phonations by the singer subject in this study. The second strategy for change in maximum flow declination rate is to decrease the open quotient, which raises the energy of the higher-frequency partials. This mechanism was not used by the countertenor subject, as discussed above. The third strategy is to change the acoustic load of the vocal tract. When F_0 is below F_1, the vocal tract offers an inertive load, which causes the glottal waveform to skew significantly to the right, or towards the moment of closure (Rothenberg, 1981). A deliberate narrowing of the epilaryngeal tube in the relation to the pharynx can raise the inertance and create greater skewing (Rothenberg, 1981; Titze and Story, 1997). The flow waveforms of this subject in this study tended to have relatively small amounts of waveform skewing (Fig. 4), suggesting that this mechanism was also not used.

D. Formant tuning

The relationship between formant frequencies of the vocal tract and harmonic frequencies of the source was examined in order to assess the use of formant tuning by the subject, i.e., the intentional adjustment of formant frequencies to approach or coincide with a harmonic. Each of the six falsetto pitch/loudness conditions was examined for the presence of any instances where the first or second formant was very close to (or coincided with) a source harmonic. When the frequency of the first or second formant fell within 40 Hz of a spectral harmonic, this was considered as a possible instance of formant tuning (Sundberg, 1993). Figure 9 illustrates the changes in formant frequencies (F_1 and F_2) from speech phonation (produced in chest register) at comfortable loudness to phonation at B-flat$_4$ (466 Hz), on the vowel /ɑ/. The frequencies of voice source harmonics are represented by vertical and horizontal grid lines. Each point on the figure represents, simultaneously, the location of the frequency of the first formant (in the horizontal dimension) and of the

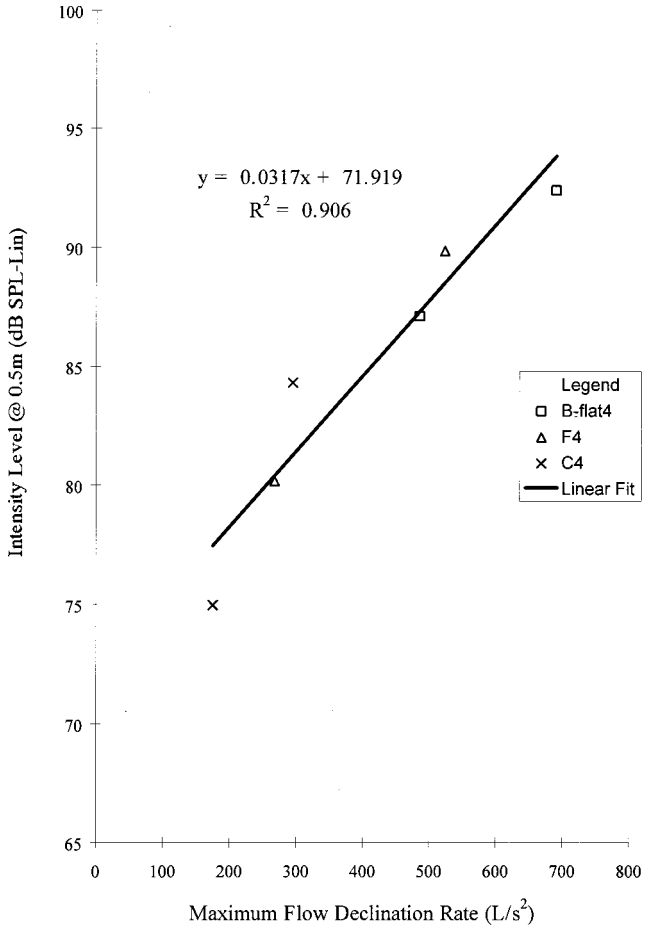


FIG. 8. Sound-pressure level versus negative peak flow derivative (maximum flow declination rate) for falsetto pitches.

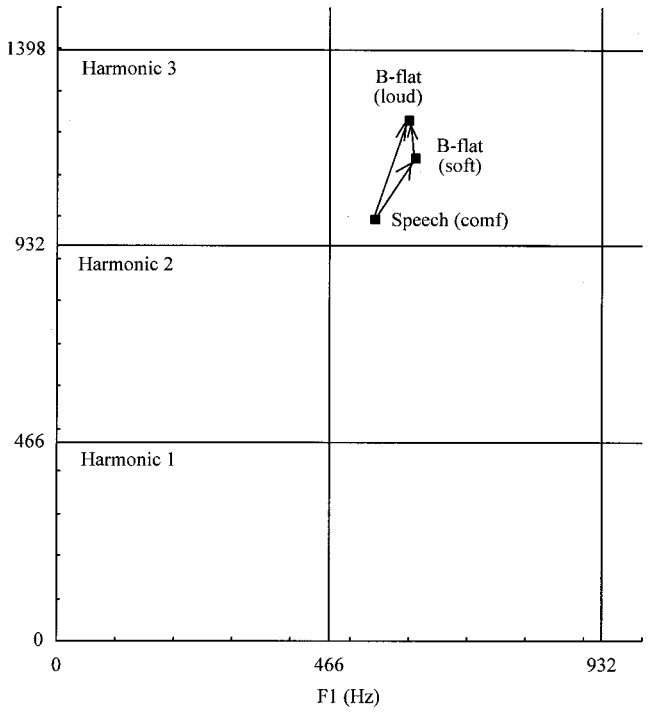


FIG. 9. Formant trajectories in F_1–F_2 space, for the vowel /ɑ/ as it migrates from speech to falsetto singing.

second formant (in the vertical dimension). Changes in the position of a data point towards a grid line would represent a tuning effect. A change in position towards the intersection of two harmonic lines represents a double tuning effect. It is seen that no specific tuning was sought by the subject. This was somewhat surprising to us because it would seem that the countertenor would benefit from some formant tuning, given that the production is oriented toward boosting the fundamental rather than higher harmonies.

Only two of the six falsetto conditions met the criterion of a harmonic being within 40 Hz of a formant. These two cases do not strongly support the notion of any systematic and intentional use of formant tuning, at least for the vowel /ɑ/. Perhaps for this vowel, intentional formant tuning would entail significant downward migration of F_1 and F_2, which would result in perceivable vowel distortion and perhaps vocal timbre changes.

It was not possible to test the effect of formant tuning comprehensively within the limitations of the current study. To do so would have required formant data for several vowels spanning the pitch and loudness range of the falsetto register. Risks associated with radiation exposure during EBCT imaging did not allow for the collection of additional vowel data. If formant tuning were to be used as an intentional strategy to increase vocal intensity in high-pitched falsetto phonation, one might expect that it would be similar to formant tuning in the soprano voice, as described by Sundberg (1975), whereby F_1 would be adjusted to be slightly greater or equal to F_0. Both the soprano and the countertenor produce phonations with a relatively steep spectral slope (few harmonics), so that reinforcement of the fundamental with the first formant may not risk destabilizing the spectral balance. A systematic evaluation of formant tuning would require the acquisition of formant information (via imaging) for vowels in which F_1 is normally lower than F_0, e.g., /i/ and /u/, which would provide a plausible acoustic context in which F_1 could migrate towards F_0. This must be left for a subsequent study.

IV. CONCLUSIONS

Computer models of phonation have grown in both flexibility and sophistication, making increasingly detailed analysis by synthesis feasible in singing research. This is particularly useful when there is significant nonlinear interaction between the source and the resonator, such that linear inverse filtering is suspect. The present study has shown that analysis by synthesis was fruitful for the study of intensity control in falsetto phonation of a trained countertenor. Our findings were as follows: (1) Subglottal pressure accounted for almost 90% of the variation in SPL in falsetto phonation. (2) Glottal adduction, measured in terms of the open quotient of the modeled glottal flow pulse, did not change significantly in these falsetto phonations, and thus, did not affect vocal intensity. (3) Formant tuning, or coincidences between harmonics and formants, occurred in only two out of the six possibilities for the vowel /ɑ/. These cases did not support systematic use of formant tuning in countertenor singing, but our results were not comprehensive enough to include the critical vowels /u/ and /i/.

Although the results of this study are based on data from a single subject, they may be generalizable to the use of falsetto in voice therapy and in the training of the singing voice. In particular, the strategies used for intensity control do not differ much from those used in other phonation types, yet the mechanism is not vocally abusive. Further studies are needed to examine intraindividual and interindividual differences in vocal intensity control strategies, not only to acquire a better understanding of the relevant mechanisms, but also to guide clinical practitioners and voice teachers in the use of falsetto voice in their work.

ACKNOWLEDGMENTS

This research was supported in part by Research Grant Number 5 R01 DC 00387-08 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

Alley, R. (**1982**). *Peking Opera* (New World, Beijing).
Boone, D., and McFarlane, S. (**1994**). *The Voice and Voice Therapy*, 5th ed. (Prentice-Hall, Englewood Cliffs, NJ).
Brown, O. (**1996**). *Discover Your Voice* (Singular, San Diego).
Colton, R. H. (**1973**). ''Vocal Intensity in the Modal and Falsetto Registers,'' Folia Phoniatr. **25**, 62–70.
Colton, R. H., and Casper, J. K. (**1990**). *Understanding Voice Problems, A Physiologic Perspective for Diagnosis and Treatment* (Williams & Wilkins, Baltimore).
Colton, R. H., and Hollien, H. (**1972**). ''Phonational Range in the Modal and Falsetto Registers,'' J. Speech Hear. Res. **15**, 708–713.
Frisell, A. (**1964**). *The Tenor Voice* (Bruce Humphries, Somerville, MA).
Giles, P. (**1982**). *The Countertenor* (Frederick Muller, London).
Gurewitsch, M. (**1999**). ''For the countertenors, recognition and roles,'' in *The New York Times*, pp. AR 29, AR 31, 4 April 1999.
Hirano, M., Ohala, J., and Vennard, W. (**1969**). ''The Function of Laryngeal Muscles in Regulating Fundamental Frequency and Intensity of Phonation,'' J. Speech Hear. Res. **12**, 616–628.
Hollien, H. (**1960**). ''Vocal Pitch Variation Related to Changes in Vocal Fold Length,'' J. Speech Hear. Res. **3**, 150–156.
Hollien, H. (**1974**). ''On Vocal Registers,'' J. Phonetics **2**, 125–143.
Hollien, H. (**1977**). ''The Registers and Ranges of the Voice,'' in *Approaches to Vocal Rehabilitation*, edited by M. Cooper and H. C. Cooper (Charles C. Thomas, Springfield, IL), pp. 76–121.
Holmberg, E., Hillman, R., and Perkell, J. (**1988**). ''Glottal Airflow and Transglottal Air Pressure Measurements for Male and Female Speakers in Soft, Normal, and Loud Voice,'' J. Acoust. Soc. Am. **84**, 511–529.
Holmberg, E., Hillman, R., Perkell, J., Guiod, P. C., and Goldman, S. L. (**1995**). ''Comparisons Among Aerodynamic, Electroglottographic, and Acoustical Spectral Measures of Female Voice,'' J. Speech Hear. Res. **38**, 1212–1223.
Isshiki, N. (**1964**). ''Regulatory Mechanism of Voice Intensity Variation,'' J. Speech Hear. Res. **7**, 17–29.
Kitajima, K., and Fujita, F. (**1990**). ''Estimation of Subglottal Pressure With Intraoral Pressure,'' Acta Oto-Laryngol. **109**, 473–478.
Laver, J. (**1980**). *The Phonetic Description of Voice Quality* (Cambridge University Press, New York).
Liljencrants, J. (**1985**). ''Speech Synthesis with a Reflection-Type Line Analog,'' DS dissertation, Dept. of Speech Comm. and Music Acoust., Royal Inst. of Tech., Stockholm, Sweden.
Löfqvist, A., Carlborg, B., and Kitzing, P. (**1982**). ''Initial Validation of an Indirect Measure of Subglottal Pressure During Vowels,'' J. Acoust. Soc. Am. **72**, 633–635.
MacKerras, C. (**1983**). *Chinese Theater, From its Origins to the Present Day* (University of Hawaii Press, Honolulu).
Malm, W. P. (**1967**). *Music Cultures of the Pacific, the Near East and Asia* (Prentice-Hall, Englewood Cliffs, NJ).
Perlman, A. (**1992**). Personal communication.
Reid, C. L. (**1978**). *Bel Canto: Principles and Practices* (Joseph Patelson, New York).
Rothenberg, M. (**1973**). ''A New Inverse Filtering Technique for Deriving the Glottal Airflow Waveform During Voicing,'' J. Acoust. Soc. Am. **53**, 1632–1645.
Rothenberg, M. (**1981**). ''Acoustic Interaction Between the Glottal Source and the Vocal Tract,'' in *Vocal Fold Physiology*, edited by K. Stevens and M. Hirano (University of Tokyo Press, Tokyo), 305–328.
Scherer, K. R. (**1995**). ''Expression of Emotion in Voice and Music,'' J. Voice **9**, 235–248.
Smitheran, J. R., and Hixon, T. J. (**1981**). ''A Clinical Method for Estimating Laryngeal Airway Resistance During Vowel Production,'' J. Speech Hear Disord. **46**, 138–146.
Stathopoulos, E. T., and Sapienza, C. M. (**1993a**). ''Respiratory and Laryngeal Function of Women and Men During Vocal Intensity Variation,'' J. Speech Hear. Res. **36**, 64–75.
Stathopoulos, E. T., and Sapienza, C. (**1993b**). ''Respiratory and Laryngeal Function of Children During Vocal Intensity Variation,'' J. Acoust. Soc. Am. **94**, 2531–2543.
Story, B. H. (**1995**). ''Physiologically Based Speech Simulation Using an Enhanced Wave-Reflection Model of the Vocal Tract,'' Ph.D. dissertation, University of Iowa.
Sundberg, J. (**1975**). ''Formant Technique in a Professional Female Singer,'' Acustica **32**, 89–96.
Sundberg, J., Titze, I. R., and Scherer, R. C. (**1993**). ''Phonatory Control in Male Singing: A Study of the Effects of Subglottal Pressure, Fundamental Frequency, and Mode of Phonation on the Voice Source,'' J. Voice **7**, 15–29.
Sundberg, J., Anderson, M., and Hultqvist, C. (**1999**). ''Effects of Subglottal Pressure Variation on Professional Baritone Singers' Voice Sources,'' J. Acoust. Soc. Am. **105**, 1965–1971.
Timcke, R., von Leden, H., and Moore, P. (**1959**). ''Laryngeal Vibrations: Measurement of the Glottic Wave,'' Arch. Otolaryngol. **69**, 438–444.
Titze, I. R. (**1984**). ''Parametrization of the Glottal Area, Glottal Flow and Vocal Fold Contact Area,'' J. Acoust. Soc. Am. **75**, 570–580.
Titze, I. R. (**1988**). ''A Framework for the Study of Vocal Registers,'' J. Voice **2**, 183–194.
Titze, I. R. (**1994**). *Principles of Voice Production* (Englewood Cliffs, NJ).
Titze, I. R., Mapes, S., and Story, B. (**1994**). ''Acoustics of the Tenor High Voice,'' J. Acoust. Soc. Am. **95**, 1133–1142.
Titze, I. R., and Story, B. (**1997**). ''Acoustic Interactions of the Voice Source with the Lower Vocal Tract,'' J. Acoust. Soc. Am. **101**, 2234–2243.
Titze, I. R., and Sundberg, J. (**1992**). ''Vocal Intensity in Speakers and Singers,'' J. Acoust. Soc. Am. **91**, 2936–2946.
Tom, K. (**1996**). ''Intensity Control in Male Falsetto Phonation: An Analysis by Synthesis Approach,'' Ph.D. dissertation, University of Iowa.
Tom, K., Titze, I. R., Hoffman, E. A., and Story, B. (**2001**). ''Three-dimensional Vocal Tract Imaging and Formant Structure: Varying Vocal Register, Pitch and Loudness,'' J. Acoust. Soc. Am. **109**, 742–747.
van den Berg, J. (**1963**). ''Vocal Ligaments versus Vocal Registers,'' The NATS Bulletin, Dec. 1963, 16–21, 31.
Welch, G. F., Sergeant, D. C., and MacCurtain, F. (**1988**). ''Some Physical Characteristics of the Male Falsetto Voice,'' J. Voice **2**, 151–163.

Temperature dependence of ultrasonic enhancement with a site-targeted contrast agent

Christopher S. Hall, Jon N. Marsh, and Michael J. Scott
Washington University School of Medicine, St. Louis, Missouri 63110 and Barnes–Jewish Hospital, St. Louis, Missouri 63110

Patrick J. Gaffney
St. Thomas' Hospital, London SE1 7EH, United Kingdom

Samuel A. Wickline and Gregory M. Lanza[a)]
Washington University School of Medicine, St. Louis, Missouri 63110 and Barnes–Jewish Hospital, St. Louis, Missouri 63110

(Received 6 February 2001; accepted for publication 23 June 2001)

Molecular imaging contrast agents specifically detect the biochemical ''signatures'' of disease before anatomical manifestations are apparent. Sensitive and specific localization of fibrin both *in vivo* and *in vitro* has been demonstrated with the use of a ligand-directed liquid perfluorocarbon nanoparticle. Since the acoustic properties of perfluorocarbons are known to vary with temperature, it was hypothesized that temperature could be used to augment the magnitude of enhancement imparted by targeted nanoparticles. Accordingly, the acoustic backscatter of two different substrates, nitrocellulose membrane and human plasma clot, targeted by the nanoparticles was measured at temperatures ranging from 27° to 47 °C in 5 °C increments. Classic avidin–biotin interactions were utilized to couple biotinylated nanoparticles to avidin-conjugated nitrocellulose membranes. Ultrasonic contrast enhancement of the nitrocellulose membrane at 25 MHz, measured by acoustic microscopy, increased from 2.0 ± 0.3 dB at 27 °C to 3.7 ± 0.4 at 47 °C. In a similar experiment, antifibrin nanoparticles bound to human plasma clots also exhibited temperature-dependent ultrasonic signal enhancement ranging from 13.9 ± 1.5 dB at 27 °C to 18.1 ± 1.5 dB at 47 °C. The increase in ultrasonic contrast enhancement measured was well described by a simple, acoustic transmission line model with temperature-dependent impedance. These results suggest that temperature-dependent changes in acoustic backscatter may be used to further differentiate tissues targeted with site-specific nanoparticles from surrounding normal soft tissues. [DOI: 10.1121/1.1395584]

PACS numbers: 43.80.Cs, 43.80.Ev, 43.80.Qf, 43.80.Vj [FD]

I. INTRODUCTION

Molecular imaging is a promising new area of medical diagnostic and therapeutic imaging that utilizes specific ligand-targeted contrast agents to detect molecular markers of disease.[1] The overall goal of molecular imaging is complementary to traditional clinical imaging techniques, which detect pathologies that are clinically manifested by adversely affecting either form or function of a specific organ or system. Molecular imaging attempts to define the presence of a disease by detection of molecular epitopes that cause future pathological conditions. Because molecular markers cannot be resolved with traditional clinical imaging techniques, the use of new technologies such as targeted contrast agents is necessary to aid and amplify detection.

Site-targeted ultrasonic contrast agents represent one such construct that promises to increase the sensitivity of molecular imaging for both *in vivo* and *in vitro* applications. Efforts to develop site-targeted ultrasonic contrast agents have been employed in the past with varying degrees of success. Most acoustic contrast agents that have been under development can be divided into three categories: liposomes,[2–4] microbubbles,[5–8] and nano-emulsions.[9–11] Liposomes are spherical bileaflet membrane vesicles produced spontaneously by phospholipids in water. Multilamellar lipid bilayers produced through a dehydration–rehydration process can form internal vesicles within a liposome and lead to increased acoustic reflectance.[2–4] In the second approach, microbubbles have been proposed for site-targeted modalities in addition to their perfusion applications. Microbubbles have been targeted towards thrombi,[6,12] avidin-coated petri disk,[8] and activated endothelial cells.[7] Other investigators have examined the interaction of thrombus with site-targeted agents. In particular, Unger *et al.* have observed successful binding of MRX-408, a bubble-based contrast agent, both *in vitro* and *in vivo*.[6]

The third approach for site-targeted ultrasonic contrast agents involves nano-emulsions.[9] We have developed a site-targeted ultrasonic contrast agent which is a liquid perfluorocarbon, lipid-encapsulated nanoparticulate emulsion (~250 nm in diameter). Previous studies by our laboratory have shown that perfluorocarbon nanoparticulate emulsions

a)Author to whom correspondence should be addressed. Address for correspondence: Cardiology Division, Campus Box 8086, Washington University School of Medicine, 660 S. Euclid Ave., St. Louis, MO 63110-1093; electronic mail: greg@cvu.wustl.edu

provide substantial enhancement when targeted towards *in vitro* and *in vivo* thrombi preparations.[10,12,13] In addition, the binding characteristics have been further elucidated in a paper by Hall *et al.* that describes the time evolution of ultrasonic enhancement due to the progressive binding of the contrast agent to a human plasma clot.[14] The site-targeted contrast system described in this paper utilizes a three-step process which relies on (1) the binding of a biotinylated antifibrin (NIB 1H10) antibody[15,16] to the substrate; (2) circulating avidin as an intermediate step to bind to the biotin molecules on the antibody; and (3) binding biotinylated contrast agent to the affixed avidin. This three-step approach was used for its ease in *in vitro* experiments, although the authors have recently shown the success of a single-step system that binds an antibody directly to the contrast particle.[11]

The subject of this work is the potential enhancement of the contrast based on the physical behavior of the nanoparticles. In particular, we examine the temperature dependence of enhancement from substrates targeted with a nanoparticulate emulsion. Previous work has described a physical model for the enhancement observed from targeted nanoparticulate emulsion.[10] A simple, acoustic transmission line model is dependent on the acoustic impedance of the contrast agent that leads to varying acoustic enhancement from specific substrates[14] and with selection of perfluorocarbon.[13,17] The impetus for this study was the observation that the speed of sound of perfluorocarbons is highly sensitive to temperature,[18] and therefore a small change in temperature may significantly affect the acoustic impedance mismatch between contrast agent and targeted substrate.

II. THEORETICAL MODELING

To explain the observed enhancement of ultrasonic scattering from a surface targeted with the liquid nanoparticle perfluorocarbon emulsion, a simple acoustic transmission line model has been developed.[10] This model attempts to account for several experimental features of the contrast agent: the poor intrinsic scatterer cross section for emulsions in suspension and the observed marked increase in backscatter when the contrast agent attaches to a substrate.

The incoming ultrasonic wave is modeled as a plane wave approaching the interface provided by a substrate covered with a targeted contrast agent. The incoming wave passes through a nonattenuating medium and then encounters the contrast agent, which is modeled as a continuous thin layer completely covering the substrate. The substrate is modeled as having a thickness great enough to preclude any internal reflections from interfering with the observed front-wall echo. In this study, the experimental substrates consist of plasma clots, as well as a nonbiological medium, a nitrocellulose membrane.

The reflection coefficient from the combined system can be expressed as

$$r(k)=r_{12}+\frac{t_{12}t_{21}r_{23}e^{2ikd}}{1-r_{21}r_{23}e^{2ikd}}, \tag{1}$$

where r and t are the complex amplitude reflection and transmission coefficients, respectively, between water (1), contrast (2), and substrate (3), k is the wave number of the ultrasonic wave in the contrast layer, and d is the thickness of the contrast layer. The typical thickness of a contrast layer consisting of one covering of the substrate by the emulsion particles is about 250 nm. The reflection and transmission coefficients are functions of the complex impedance, and therefore are primarily controlled by the density and speed of sound in each material.

For the purpose of this paper, we examine the temperature dependence of reflection enhancement that is related to the temperature dependence of the transmission and reflection coefficients. The observation that the speed of sound of perfluorocarbons is highly dependent on the ambient temperature implies that the acoustic impedance can be expressed as

$$Z(T)=\rho(T)\,\nu(T), \tag{2}$$

so

$$\frac{\partial Z(T)}{\partial T}=\frac{\partial \rho(T)}{\partial T}\,\nu(T)+\frac{\partial \nu(T)}{\partial T}\,\rho(T), \tag{3}$$

where $Z(T)$ is the complex acoustic impedance, $\rho(T)$ is the mass density, $\nu(T)$ is the speed of sound, and T is the temperature. The temperature variation for the density of the perfluorocarbon changes less rapidly than that of the speed of sound. Previously published reports,[18] for a selection of perfluorocarbons, describe a change of as much as 4% in the density and 42% change in speed of sound as temperature extends from room temperature to 25 °C above.

The reflection and transmission coefficients are expressed as

$$r_{ab}=\frac{Z_b-Z_a}{Z_b+Z_a} \quad \text{and} \quad t_{ab}=\frac{2Z_b}{Z_b+Z_a}, \tag{4}$$

where

$$\frac{\partial R_{ab}(T)}{\partial (T)}=\frac{\partial T_{ab}(T)}{\partial T}=\frac{-2}{(Z_b+Z_a)^2}\left(Z_b\frac{\partial Z_a}{\partial T}-Z_a\frac{\partial Z_b}{\partial T}\right). \tag{5}$$

The expression above can be further simplified, if one of the materials has a very weak dependence on temperature in comparison to the other, by ignoring the second term. Combining Eqs. (3) and (5) shows the direct dependence of the reflection and transmission coefficients on the temperature dependence of speed of sound.

The enhancement provided by the layer of contrast agent can be expressed in decibels as

$$\text{Enhancement}=20\log(|R(k)|/|R_0|), \tag{6}$$

where R_0 is the reflection coefficient of the substrate surface without the contrast agent.

For the calculations shown in Figs. 3 and 6, we modeled the contrast agent as a single layer (250 nm) bound to a substrate with velocities $v_{\text{clot}}=v_{\text{water}}$ and

$v_{\text{nitrocellulose}} = 147$ mm/μs;

densities $\rho_{\text{clot}} = 1.01$ g/cm^3

and

$\rho_{\text{nitrocellulose}} = 1.141$ g/cm^3.

The velocity of the perfluorocarbon was experimentally determined at each temperature and the reported density of 1.73 g/cm^3 was used. Figures 3 and 6 are further discussed in the Results section.

III. METHODS

A. Sample preparation

1. Nitrocellulose

Flat nitrocellulose membranes were prepared for contrast binding with a diaminohexane spacer and activated with glutaraldehyde for protein conjugation. Nitrocellulose discs (2.5-cm diameter) were immersed in 1,6 diaminohexane (2.5% w/v, *p*H 11.9) for 60 min under constant rotary agitation. The membranes were next washed under constant agitation for approximately 12 h in 1-M acetic acid followed by 12 h in ultrapure water with several changes of each medium. The diaminoalkane-modified nitrocellulose membranes were then exposed to 1% glutaraldehyde in 0.5 M $NaHCO_3/Na_2CO_3$, *p*H 10, for 15 min followed by a 3-h wash in several changes of ultrapure water. The diaminohexane-modified, glutaraldehyde-activated nitrocellulose membranes were stored dry at 4 °C until use.

Avidin (50 μg) dissolved in 0.1-M phosphate buffered saline (PBS) (*p*H 7.2–7.4) was placed onto the center of each membrane with a microliter syringe and allowed to dry. Next, each membrane was washed with three, 5-min changes of PBS-0.1% Tween 20. Dehydrated milk powder suspended in PBS-01% Tween 20 was used to block glutaraldehyde activated protein binding sites left unoccupied after the application of avidin for 20 min. Each protein was removed with three 5-min isotonic, PBS washes.

Five nitrocellulose discs were utilized for exposure to biotinylated perfluorooctane (PFO) particles. Control disks were prepared by utilizing nitrocellulose membranes completely blocked as above.

2. Human plasma clots

Plasma clots were produced on nitrocellulose membranes by combining human citrated plasma (375 μl) and 100-mM calcium chloride (25 μl) with three units of thrombin in a plastic mold placed on the membranes. The plasma was allowed to coagulate slowly at ambient temperature and then transferred to PBS until exposure to the control or targeted contrast system.

Clots were then incubated with the biotinylated 1h10 antibody overnight in PBS in a 4 °C cold room under gentle agitation. The clots were then rinsed and exposed to 100 μg of avidin under gentle agitation for 1 h at room temperature. The clots were then rerinsed and ready for exposure to contrast agent in the manner described below in Sec. III B. Control clots ($n=7$), *not* pretargeted with biotinylated antibody or avidin, were exposed in identical manner to the biotinylated contrast agent. Treated clots ($n=9$) were exposed to the biotinylated contrast agent which contained perfluorooctane.

B. Ultrasonic acquisition

A 25-MHz, spherically focused transducer (0.63-cm diameter, 2.54-mm focal length, Panametrics V324) was mounted on a gantry consisting of three orthogonal sleds. The transducer was translated in a raster scan format by a computer-controlled motion controller (Aerotech Unidex 12) with 100-μm resolution. The pulses sent to the motor from the motion controller were counted in a digital counter (National Instruments PCI-1200) and then a trigger was generated for a digital delay generator (Stanford Research Systems DG535). The delay generator then sent a trigger for the pulser (Panametrics 5900) and for the digitizing oscilloscope (Hewlett-Packard 54510B), as well as a delayed trigger for the real-time digitizer (Tektronix RTD720A). Traces representing the backscattered ultrasonic wave were captured on the fly as the transducer was scanned over the surface of the clot in a 68×68 (6.8 mm×6.8 mm) grid at 100-μm resolution. The traces were then transferred from the real-time digitizer to the controlling computer (Apple Power Macintosh 7300) over GPIB for image reconstruction and data storage. Acquisition typically took about 4 min per scan.

The sample chamber consisted of a fully enclosed well with an acoustic aperture to allow insonification of the sample. The chamber was attached through two ports to silicone tubing (Masterflex Platinum, i.d.=1/8 in.) that allowed perfusion of the contrast agent over the sample. A roller pump (Masterflex, Cole-Parmer, Inc.) was used to drive the flow at a rate of 20 mL/min. The flow system was filled with 20 mL of 50-mM phosphate buffer. The sample chamber and enclosed sample were positioned vertically so that no passive settling of the contrast agent could occur. After initial location of the sample, a bolus of 100 μL of the contrast agent was delivered through an injection port and ultrasonic monitoring was performed initially and after 60 min of exposure. The chamber and tubing were then flushed with phosphate buffer.

After confirmation of successful targeting of the contrast agent by imaging, the temperature was varied in 5 °C increments from 27 °C to 47 °C using an immersion heater controlled by a temperature regulator (DigiSense, Cole-Parmer, Inc.). The entire water bath was placed on top of a magnetic stirrer plate to allow for adequate mixing and homogeneous temperature distribution throughout the bath. At each temperature point, the focus of the transducer was determined by observing the reflection from a steel plate. The front wall of the sample was then placed at the focus of the transducer.

C. Data analysis

The reflected ultrasonic signals were full-wave rectified and used to render a peak-detected c-scan image so that a user-defined region of interest could be drawn around the clot or nitrocellulose sample. The signals representing the reflection of the interrogating wave of ultrasound from the surface of the sample were isolated with a rectangular windowing function. The placement of the window was care-

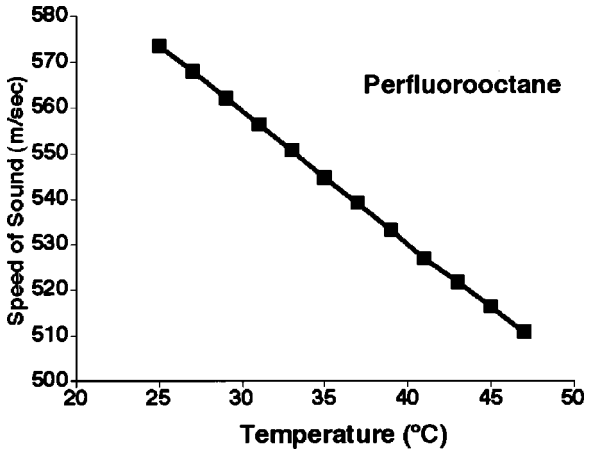


FIG. 1. Temperature dependence of the speed of sound for perfluorooctane.

fully controlled in the case of the thin nitrocellulose samples by an automatic algorithm that placed the end of the window midway between the front- and back-wall echo of the nitrocellulose paper. The isolated signal was then fast-Fourier transformed and the average power over the usable bandwidth (17 to 35 MHz, as determined by 10-dB down points) was calculated in the logarithmic domain. This ''integrated power'' was then sorted for all of the points in the region of interest and the 100 brightest points were retained for analysis. The integrated power determined at every point in the scan was also used to render images of the change in ultrasonic enhancement of the clot. The frequency-dependent reflection enhancement was averaged for the 100 brightest points and then normalized by subtracting the reflection enhancement for the control scan. This process was performed for each sample.

D. Perfluorocarbon speed of sound

Speed of sound measurements were made for pure perfluorooctane liquid (bp 105 °C) at discrete temperatures by placing 8 mL of solution in a sealed, vertically mounted sample chamber. The back of the chamber consisted of a stainless-steel reflector which extended past the fully enclosed well to allow for water-path and sample-path measurements. The chamber was mounted so that the stainless-steel reflector was perpendicular to the insonifying beam. The times of flight from the transducer to front wall of the chamber and from the transducer to the stainless-steel plate were determined for nine independent locations over the sample. The speeds of sound were then averaged together for each temperature. The temperature was changed by 2-deg increments from 25° to 47 °C by heating the surrounding water bath and allowing time for the sample to reach equilibrium (typically 20 to 25 min). Because of the mounting of the chamber, any gas was trapped at the top of the chamber and outside of the insonifying path. Speed of sound was calculated using a previously published algorithm.[19]

IV. RESULTS

Speed of sound measurements are summarized in Fig. 1. As has been previously reported for liquid perfluorocarbons,[18] the speed of sound decreases with increasing temperature. A simple linear fit (i.e., $V_{\mathrm{pfo}} = -2.872T + 645.4$, $R^2 = 0.99$) to the data is shown through the data points. Using this fit, the speed of sound for perfluorooctane is 574 m/s at 24.7 °C, which is in good agreement with previously published results,[17] 579 m/s. The change in speed of sound over the measured temperature range is 63 m/s or 10.9%.

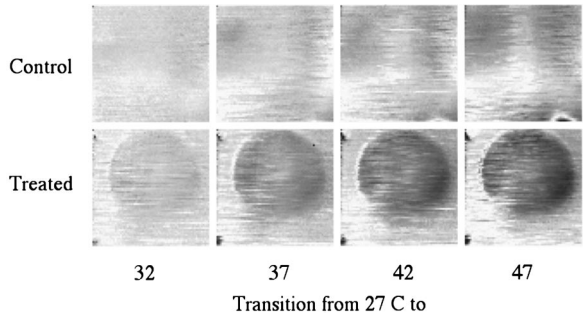


FIG. 2. Difference in enhancement from 27 °C at specific temperatures for specially prepared nitrocellulose membrane. Darker grays represent greater changes in enhancement.

Figure 2 demonstrates the change of ultrasonic reflected power with temperature for targeted and control nitrocellulose membranes. The membranes have been spotted with a single drop of avidin resulting in the almost perfectly circular feature. The picture shows the change in reflected enhancement for the transition from the baseline temperature of 27 °C to each individual temperature point. Darker areas represent areas of larger changes in enhancement and can be seen to increase substantially for targeted nitrocellulose in comparison to the control nitrocellulose.

The 100 brightest sites for each nitrocellulose membrane sample were selected to determine the average frequency-dependent reflection enhancement. Each curve was then normalized to the respective reflection enhancement from the control nitrocellulose membrane at the same temperature. The progressive increase in ultrasonic backscatter enhancement with temperature is further quantified in the top panel of Fig. 3 with the bandwidth (17 to 34 MHz) limited average. The theoretical model described in Sec. II was used in conjunction with the experimentally determined speeds of sound to predict the reflection enhancement. These predictions are presented in the bottom panel of Fig. 3. The predicted trends are in good agreement with the experimental results, demonstrating increasing enhancement with increasing temperature.

The inherently poor echogenicity of the native human plasma clots is shown in the left portion of Fig. 4. The figure illustrates two images of the integrated reflected power of a plasma clot before and after delivery of the contrast agent, with the gray scale representing a logarithmic depiction of the reflection enhancement. Darker areas on the right picture represent areas where the enhancement of the clot is increased by the presence of site-targeted ultrasonic contrast agent. The uneven nature of the acoustic enhancement is discussed in the following section.

The same plasma clot depicted in Fig. 4 is also shown in Fig. 5, where the gray scale now represents the change in enhancement brought about by increases in temperature in

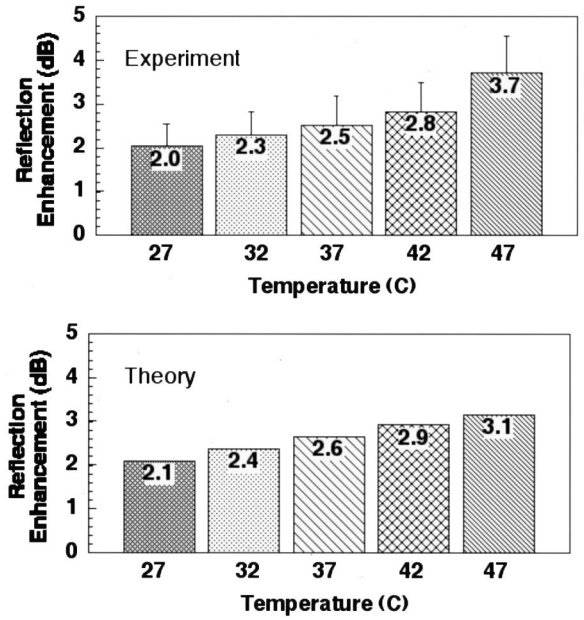


FIG. 3. Top panel contains experimental results for the reflection enhancement of nitrocellulose membrane (n=5) over control (n=5) at specific temperatures. Bottom panel contains predictions from theoretical model.

5 °C increments from the 27 °C baseline. As in the case of targeted nitrocellulose, targeted human plasma clots show substantial increase in echogenicity with temperature while the control clots (unexposed to contrast agent) show little change over this temperature range. Figure 6 shows the results of the reflection enhancement for the 100 brightest sites on the clots. The top panel shows the results of averaging over the usable bandwidth of the transducer (17 to 34 MHz). A linear fit to this plot of reflection enhancement versus temperature yields a correlation coefficient of R=0.99 and a slope of 0.21 dB/C. The results for the theoretical predictions are shown in the bottom panel of the figure.

To test for the reversibility of the temperature effect, two samples were heated and then cooled and the reflection enhancement recorded for each temperature point. The ''hysteresis'' curves are plotted in Fig. 7. Within the error bars of the measurements, the magnitude of enhancement was identical at each temperature, independent of the direction of temperature change. This result suggests that the

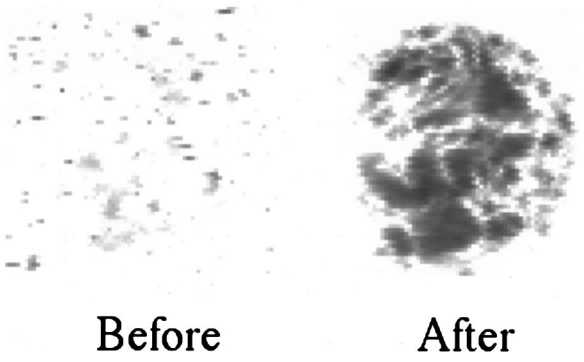


FIG. 4. The gray scale represents the ultrasonic reflection from a human plasma clot before (left panel) and after (right panel) targeting with contrast agent. Darker gray sale represents larger reflection.

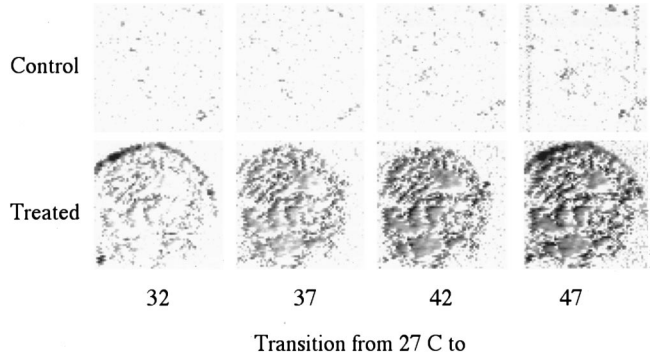


FIG. 5. Difference in enhancement from 27 °C at specific temperatures for site-targeted human plasma clot. Darker grays represent greater changes in enhancement.

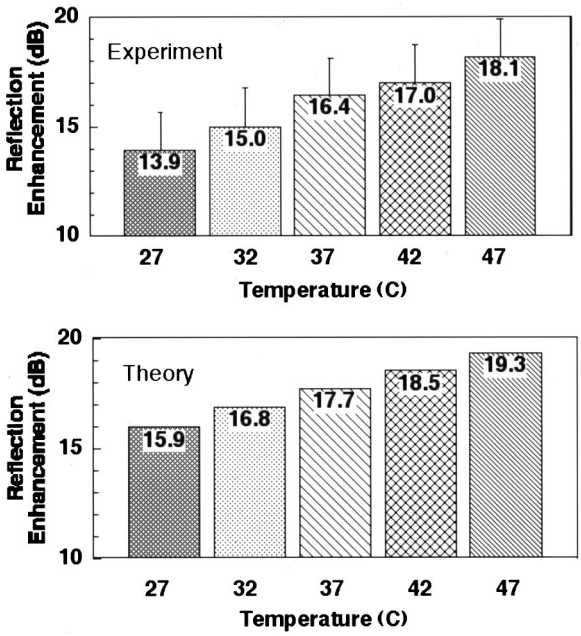


FIG. 6. Top panel contains experimental results for the reflection enhancement of human plasma clot (n=9) over control (n=7) at specific temperatures. Bottom panel contains predictions from theoretical model.

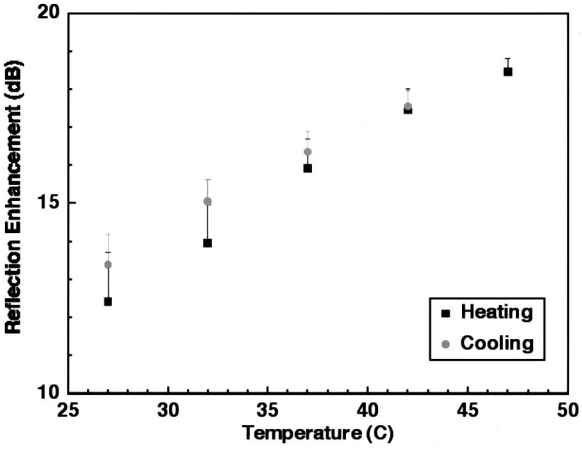


FIG. 7. A hysteresis plot of the reflection enhancement from human plasma clot as a function of temperature when heated (circles) and then cooled (squared).

temperature-dependent enhancement is not due to an irreversible change to the contrast agent with temperature.

V. DISCUSSION

This report illustrates a novel approach to detection and enhancement of molecular markers with the use of a site-targeted ultrasonic contrast agent comprising a perfluorocarbon nanoparticle emulsion. We present evidence that the contrast enhancement from such an agent is sensitive to the ambient temperature and may provide a useful means of differentiating the contrast agent from the surrounding tissue. Moreover, a mathematical model of backscatter behavior can be applied to predict this effect and in fact was instrumental in both the discovery and interpretation of the temperature dependence of contrast enhancement with perfluorocarbon nanoparticles.

One of the challenges confronting the use of site-targeted contrast agents is the sensitive detection and differentiation of the particles from the surrounding soft tissue. Detection of pathological changes on or near vascular surfaces may be compromised because the targeted substrate itself is echogenic or the signal from the surface may be somewhat view- or angle dependent. Several techniques have been developed to solve this issue. Second harmonic or harmonic and power harmonic Doppler imaging attempt to address this problem and allow differentiation of microbubbles in circulation from tissue.[20–24] However, soft tissue may also exhibit a second harmonic backscattered signal. Furthermore, the contrast agent may manifest velocities too slow for the sensitivity of Doppler techniques. Unlike microbubbles, which have inherently high acoustic impedance and resonant phenomenon responsible for enhanced backscatter cross section, the mechanism for increased reflection enhancement from the site-targeted nanoparticle emulsion has been postulated to be due to an increased acoustic impedance mismatch at the target surface where particles bind and accumulate.[10]

The intrinsic, acoustic impedance of the contrast agent is dominated by the perfluorocarbon liquid encapsulated within the nanoparticle. Previous reports from this laboratory have shown that the magnitude of enhancement is dependent on the type of perfluorocarbon chosen for the contrast agent and is proportional to its acoustic impedance.[13] The surface impedance mismatch between surrounding soft tissue and the contrast agent can be altered by careful selection of perfluorocarbon[13] or by adjusting the perfluorocarbon impedance through external means.

One such approach is the subject of this study, namely, changes in the temperature of the agent, which decreases the speed of sound within the perfluorocarbon. Figure 1 illustrates speed of sound (and therefore acoustic impedance) decreases at a rate of roughly 2.9 m/s-C for perfluorooctane. A proportional change in acoustic impedance occurs as a result of this temperature dependence. Future examination of other biologically safe perfluorocarbons may reveal even greater temperature sensitivity.

The increased impedance mismatch between the bound nanoparticles and the substrate that occurs with increased temperature results in improved reflection enhancement from targeted substrates (Figs. 3 and 6). In comparison, the control nitrocellulose and control human plasma clots both show only a small change in reflected power as the temperature increases. The targeted or enhanced substrates both exhibit detectable changes in enhancement with as little as a 5 °C change in temperature. The correlation between contrast enhancement and temperature for nitrocellulose ($R=0.95$) and plasma clot ($R=0.99$) is excellent. Figures 2 and 5 quantify the changes in reflection enhancement as a function of temperature. Overall, the increase in enhancement is 0.08 dB/C for nitrocellulose and 0.21 dB/C for plasma clot.

An experiment was added to the study to test for the hysteresis of the reflection enhancement effect with temperature. In this experiment, we sought to determine if heating the site-targeted contrast agent caused an irreversible change in its composition. The results of Fig. 7 suggest that this is not the case, as the associated enhancement is reproducible whether the sample is being heated or cooled. Any phase change from a lipid-encapsulated liquid perfluorocarbon nanoparticle to a gaseous microbubble would be associated with a substantial increase in volume that could lead to leakage or disruption of the lipid shell. Such a change in the shell would be irreversible and might lead to an inability of the microbubble to remain bound in flow conditions. The fact that the same reflection enhancement is obtainable for the sample as it is either heated or cooled argues that such an irreversible change is unlikely. Thus, we propose that the nanoparticles should be sufficiently stable *in vivo* for purposes of contrast enhancement that employs local heating effects.

The model outlined in Sec. II of this paper describes an approximation to the experimental setup. The contrast agent is modeled as a confluent layer of perfluorocarbon covering a flat substrate, which may not exist perfectly under physiologic conditions. Nevertheless, the simple model successfully predicts the overall trend for the enhanced reflectivity after binding to a specular surface. The acoustic transmission line model predicts an increase in enhancement by 3.7 dB for nitrocellulose and 19.3 dB for plasma clot at 47 °C. These values compare with the experimentally determined values of 3.7 and 18.1 dB for nitrocellulose and clot, respectively.

While the change in enhancement with temperature correlates well with theoretical predictions, the overall magnitude is slightly different for the predicted enhancement than that measured by experiments. Several possible factors may play a role in this small discordance. It is possible that the emulsion particles do not form a single confluent layer, but may instead form a partial layer that possesses acoustic parameters combining the surrounding medium and perfluorocarbon. This combination may have further increased acoustic impedance and therefore result in enhancement lower than the model might predict. A second consideration is that the physical properties used for the initial acoustic impedance of clot and nitrocellulose are not completely accurate. For the purposes of this paper and previously published results, we have used an estimation of the speed of sound and density of human plasma clots that is consistent with the reflection measured from unenhanced clot. Third, the first term in Eq. (3) is ignored, because we did not measure the

density variation of perfluorooctane as a function of temperature. A fourth consideration is partial phase conversion of the perfluorocarbon core with elevated temperature. Since the temperatures applied are less than half the boiling point of perfluorooctane (bp 105 °C), it is unlikely that any acoustically significant conversion of fluorocarbon liquid to gas occurred during this study. Recent data from our laboratory demonstrate that the acoustic attenuation profile of nanoparticles increases linearly with frequency as expected and is unaffected by changes in ambient pressure, mechanical index, duration of the acoustic pulse train, etc. as opposed to microbubbles, which are very sensitive to all of the above conditions (Hughes, unpublished).[25]

Whereas it is clear that the scattering cross section of a single microbubble exceeds that of a single emulsion nanoparticle, the use of microbubbles may be associated with potential drawbacks for some molecular imaging applications. Unlike nanoparticles, which by virtue of their size and stability have a circulatory half-life in excess of 1 h, microbubbles have a short circulatory half-life that diminishes the time available to localize and adequately bind to molecular epitopes. In contradistinction to nanoparticles, conventional microbubble contrast agents are much more susceptible to destruction at typical clinical imaging frequencies and energies, which may limit the opportunity for thorough characterization targeted pathologies. Implementation of stiffer bubble shells may improve circulatory longevity and resistance to insonification destruction but also could alter particle scattering characteristics. However, increasing microbubble shell stability may increase the persistence of the contrast agent within the blood pool and concomitantly decrease the relative contrast between the blood pool and site-targeted pathology. Nanoparticles are inherently nonechogenic in circulation until bound, overcoming the high scattering limitation of microbubbles, diminishing the potential for false-positive results, and markedly increasing their contrast-to-noise ratio.

A potential limitation for clinical utilization of this method is that human tissue is reported to undergo nonreversible adverse changes after prolonged exposures to temperatures above 43 °C. Studies on the effects of hyperthermia report significant tissue damage after exposure for minutes at temperatures exceeding 43 °C.[26,27] The authors are not currently aware of any reported effects of short time exposures to temperatures above normal body temperature. In a future application of this technology we foresee the bound nanoparticles as ''acoustic beacons'' that can be identified and segmented in the displayed image by their rapidly changing ultrasonic response to brief (i.e., microsecond) focused heating pulses. Further research will be required to determine the rates of temperature change versus the ultrasonic pulse applied as well as experiments to resolve the challenge of adjusting for shifting focus of an ultrasonic transducer due to the altered speed of sound in tissue with increased temperature.[28,29]

The results presented in this paper ultimately might be useful for differentiating acoustic contrast agents bound to molecular markers from the surrounding tissue. Such an imaging system might use the backscattered information of two different temperatures to construct a ''difference'' image similar to those in Figs. 2 and 5. The specific approach to heating tissue locally remains to be developed, but ultrasound itself is clearly one of the best candidates to accomplish this task. In our opinion, this approach has the potential to facilitate differentiation of materials according to the temperature dependence of their intrinsic acoustic impedance. The contrast agent used in this study, targeted perfluorocarbon nanoparticles, exhibits a dramatic change in speed of sound with temperature and acoustic impedance and therefore is an excellent candidate for improving the sensitivity of molecular imaging with ultrasound imaging methods.

ACKNOWLEDGMENTS

The authors acknowledge support from NIH (No. R01-59865), Wolff Charitable Trust, Bracco Diagnostics, Inc. SCA&I Fellowship Award, AHA Heartland Affiliate No. 9960282Z, Barnes-Jewish Hospital Research Foundation Grant (No. 5408), ACC/Searle Career Development Award in Cardiovascular Disease, NCI Unconventional Innovations Program (N01-CO-07013-32).

[1] R. Weissleder, ''Molecular Imaging: Exploring the Next Frontier,'' Radiology **212**, 609–614 (1999).
[2] H. Alkan-Onyuksel, S. M. Demos, G. M. Lanza, M. J. Vonesh, M. E. Klegerman, B. J. Kane, J. Kuszak, and D. D. McPherson, ''Development of Inherently Echogenic Liposomes as an Ultrasonic Contrast Agent,'' J. Pharm. Sci. **85**, 486–490 (1996).
[3] S. M. Demos, H. Onyuksel, J. Gilbert, S. I. Roth, B. Kane, P. Jungblut, J. V. Pinto, D. D. McPherson, and M. E. Klegerman, ''*In Vitro* Targeting of Antibody–conjugated Echogenic Liposomes for Site-Specific Ultrasonic Image Enhancement,'' J. Pharm. Sci. **86**, 167–171 (1997).
[4] S. M. Demos, H. Alkan-Onyuksel, B. J. Kane, K. Ramani, A. Nagaraj, R. Greene, M. Klegerman, and D. D. McPherson, ''*In Vivo* Targeting of Acoustically Reflective Liposomes for Intravascular and Transvascular Ultrasonic Enhancement,'' J. Am. Coll. Cardiol. **33**, 867–875 (1999).
[5] R. F. Mattrey and M. P. Andre, ''Ultrasonic Enhancement of Myocardial Infarction with Perfluorocarbon Compounds in Dogs,'' Am. J. Cardiol. **54**, 206–210 (1984).
[6] E. C. Unger, T. P. McCreery, R. H. Sweitzer, D. Shen, and G. Wu, ''*In Vitro* Studies of a New Thrombus-specific Ultrasonic Contrast Agent,'' Am. J. Cardiol. **81**, 58G–61G (1998).
[7] F. S. Villaneuva, R. J. Jankowski, S. Klibanov, M. L. Pina, S. M. Alber, S. C. Watkins, G. H. Brandenburger, and W. R. Wagner, '' Microbubbles Targeted to Intercellular Adhesion Molecule-1 Bind to Activated Coronary Artery Endothelial Cells,'' Circulation **98**, 1–5 (1998).
[8] A. L. Klibanov, M. L. Hughes, J. K. Wojdyla, J. N. Marsh, C. S. Hall, J. G. Miller, J. H. Wible, and G. H. Brandenburger, ''Targeting of Ultrasound Contrast Material: Selective Imaging of Microbubbles *in Vitro*,'' Acad. Radiol. **5**, S243–S246 (1998).
[9] G. Lanza, K. Wallace, M. Scott, W. Cacheris, D. Abendschein, D. Christy, A. Sharkey, J. Miller, P. Gaffney, and S. Wickline, ''A Novel Site-targeted Ultrasonic Contrast Agent with Broad Biomedical Application,'' Circulation **94**, 3334–3340 (1996).
[10] G. Lanza, R. Trousil, K. Wallace, J. Rose, C. Hall, M. Scott, J. Miller, P. Eisenberg, P. Gaffney, and S. Wickline, ''*In Vitro* Characterization of a Novel, Tissue-targeted Ultrasonic Contrast System with Acoustic Microscopy,'' J. Acoust. Soc. Am. **104**, 3665–3672 (1998).
[11] G. M. Lanza, D. R. Abendschein, C. S. Hall, J. H. Marsh, M. J. Scott, D. E. Scherrer, and S. A. Wickline, ''Molecular Imaging of Stretch-induced Tissue Factor Expression in Carotid Arteries with Intravascular Ultrasound,'' Invest. Radiol. **35**, 227–234 (2000).
[12] G. Lanza, K. Wallace, S. Fischer, D. Christy, M. Scott, R. Trousil, W. Cacheris, J. Miller, P. Gaffney, and S. Wickline, ''High Frequency Ultrasonic Detection of Thrombi with a Targeted Contrast System,'' Ultrasound Med. Biol. **23**, 863–870 (1997).
[13] J. N. Marsh, C. S. Hall, M. J. Scott, R. J. Fuhrhop, P. J. Gaffney, S. A. Wickline, and G. M. Lanza, ''Enhancement of Reflectivity by Specific Perfluorocarbon Emulsions used in Site-targeted Ultrasound Contrast

Agent,'' SPIE International Symposium on Medical Imaging, San Diego, CA (2000).
[14] C. S. Hall, J. N. Marsh, M. J. Scott, P. J. Gaffney, S. A. Wickline, and G. M. Lanza, ''Time Evolution of Enhanced Ultrasonic Reflection Using a Fibrin-targeted Nanoparticulate Contrast Agent,'' J. Acoust. Soc. Am. (in press).
[15] T. Edgell, F. McEvoy, P. Webbon, and P. J. Gaffney, ''Monoclonal Antibodies to Human Fibrin: Interaction with other Animal Fibrins,'' Thromb. Haemostasis **75**, 595–599 (1996).
[16] S. Raut and P. J. Gaffney, ''Evaluation of the Fibrin Binding Profile of Two Anti-fibrin Monoclonal Antibodies,'' Thromb. Haemostasis **76**, 56–64 (1996).
[17] C. S. Hall, G. M. Lanza, J. H. Rose, R. J. Kaufmann, R. W. Fuhrhop, S. H. Handley, K. R. Waters, J. G. Miller, and S. A. Wickline, ''Experimental Determination of Phase Velocity of Perfluorocarbons: Applications to Targeted Contrast Agents,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 75–84 (2000).
[18] K. L. Narayana and K. M. Swamy, ''Temperature Variation of Ultrasonic Velocity in Some Low Velocity Fluorocarbon Liquids,'' Acoust. Lett. **9**, 137–143 (1986).
[19] I. Y. Kuo, B. Hete, and K. K. Shung, ''A Novel Method for the Measurement of Acoustic Speed,'' J. Acoust. Soc. Am. **88**, 1679–1682 (1990).
[20] P. N. Burns, ''Harmonic Imaging with Ultrasound Contrast Agents,'' Clin. Radiol. **51**, 50–55 (1996).
[21] J. D. Kasprzak, B. Paelinck, F. J. Ten Cate, W. B. Vletter, N. deJong, D. Poldermans, A. Elhendy, A. Bouakaz, and J. R. T. C. Roelandt, ''Comparison of Native and Contrast Enhanced Harmonic Echocardiography for Visualization of Left Ventricular Endocardial Border,'' Am. J. Cardiol. **83**, 211–217 (1999).
[22] J. M. Rubin, R. O. Bude, P. L. Carson, R. L. Bree, and R. S. Adler, ''Power Doppler US: A Potentially Useful Alternative to Mean-frequency-based Color Doppler US,'' Radiology **190**, 853–856 (1994).
[23] R. Senior, S. Kaul, P. Soman and A. Lahiri, ''Power Doppler Harmonic Imaging: A Feasibility Study of a New Technique for the Assessment of Myocardial Perfusion,'' Am. Heart J. **139**, 245–251 (2000).
[24] K. T. Spencer, J. Bednarz, V. Mor-Avi, L. Weinhert , J. Tan, I. Godoy, and R. M. Lang, ''The Role of Echocardiographic Harmonic Imaging and Contrast Enhancement for Improvement of Endocardial Border Delineation,'' J. Am. Soc. Echo. **13**, 131–138 (2000).
[25] M. S. Hughes (unpublished).
[26] S. B. Barnett, G. R. ter Harr, M. C. Ziskin, W. L. Nyborg, K. Maeda, and J. Bang, ''Current Status of Research on Biophysical Effects of Ultrasound,'' Ultrasound Med. Biol. **20**, 205–218 (1994).
[27] M. W. Miller and M. C. Ziskin, ''Biological Consequences of Hyperthermia,'' Ultrasound Med. Biol. **15**, 707–722 (1989).
[28] Z. Sun and H. Ying, ''A Multigate Time of Flight Technique for Estimation of Temperature Distribution in Heated Tissue: Theory and Computer Simulation,'' Ultrasonics **37**, 107–122 (1999).
[29] R. Seip and E. S. Ebbini, ''Noninvasive Estimation of Tissue Temperature Response to Heating Fields using Diagnostic Ultrasound,'' IEEE Trans. Biomed. Eng. **42**, 828–839 (1995).

Use of a dual-pulse lithotripter to generate a localized and intensified cavitation field

Dahlia L. Sokolov[a)]
Department of Bioengineering, University of Washington, Seattle, Washington 98105 and
Applied Physics Laboratory, Center for Industrial and Medical Ultrasound, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105

Michael R. Bailey
Applied Physics Laboratory, Center for Industrial and Medical Ultrasound, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105

Lawrence A. Crum
Department of Bioengineering, University of Washington, Seattle, Washington 98105 and
Applied Physics Laboratory, Center for Industrial and Medical Ultrasound, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105

(Received 16 January 2001; accepted for publication 20 June 2001)

Localizing cavitation to the kidney stone in extracorporeal shock wave lithotripsy may be desirable since cavitation appears to play a major role in both stone comminution and renal tissue damage. A method has been developed to localize and intensify cavitation damage *in vitro*. Cavitation fields in water were filmed with a high-speed digital video camera. In a conventional lithotripter (CL), the shock wave produced by a single source creates a 2×10 cm cylindrical cloud of bubbles in water. Bubbles in the CL field collapse simultaneously along the focal axis to produce a nearly uniform 1-mm×8-cm line of pits in 25-μm-thick aluminum foil. Our dual-pulse lithotripter (DPL) uses two shock wave sources, facing each other, confocal, and triggered simultaneously to create a 4×5 cm cylindrical cloud of bubbles that collapse over a range of times and strengths such that the greatest pit damage on foils is contained within a few square millimeters of the focus. The time for bubbles to grow and collapse was measured with a focused hydrophone and compared with calculations based on the Gilmore equation. Pressure doubling due to synchronous arrival of the two pulses at the focus created increased bubble growth and increased foil pit depth. Asynchronous timing between the two pulses elsewhere in the DPL field resulted in disruption of radial dynamics and negligible pitting to foils. Translation of bubbles was also investigated, both numerically and experimentally. While net translation was calculated to be <0.3 mm in all cases, the rapid acceleration of bubbles in a small region may contribute to their premature destruction in that region. Overall, radial dynamics were found to be largely responsible for the observed pattern of cavitation in the dual-pulse lithotripsy field.
[DOI: 10.1121/1.1394221]

PACS numbers: 43.80.Sh, 43.80.Gx, 43.25.Yw [FD]

I. INTRODUCTION

Extracorporeal shock wave lithotripsy (ESWL) has been in clinical use worldwide since the mid-1980's. While still a favored and apparently low-risk method for treating renal calculi, ESWL may actually result in considerable renal tissue damage.[1,2] Many patients, including those who have recurring stones, pediatric patients, and those who already have compromised renal function, may be particularly at risk for permanent damage from lithotripsy treatment.[3,4] Acoustic cavitation—the violent growth and collapse of bubbles induced by sound waves—appears to play a major role in both stone comminution and renal tissue damage. We have developed a method for localizing and intensifying cavitation damage with the goal of increasing the rate of stone fragmentation while decreasing damage to surrounding tissue. This article reports measurement of localized and intensified cavitation activity and includes calculations indicating that changes in radial bubble dynamics largely account for both localization and intensification.

When cavitation bubbles collapse near a rigid surface, they may generate high-speed liquid jets directed toward the surface.[5–9] Using various real and artificial stone targets, several investigators[10–13] have demonstrated that such cavitation microjets play a significant role in renal stone fragmentation during shock wave treatment. Destructive interactions also occur between shock-wave-induced cavitation bubbles and cells or tissue. A collapsing bubble may radiate shock waves[14,15] and even generate free radicals,[16] and a rapidly expanding bubble may generate shear fluid flow,[17] all of which may contribute to tissue damage. Using *in vitro* cell suspensions and cell cultures, several researchers have reported a strong correlation between cavitation activity and cell damage.[13,18–20] Kodama and Tomita[21] used a gelatin surface as an *in vivo* model to study bubble dynamics. While cautioning that specific mechanisms are tissue dependent,

a)Electronic mail: dsokolov@apl.washington.edu

Kodama and Tomita reported that oscillatory bubble motion and liquid jets both contribute to tissue damage.

Several investigators have developed experimental methods to modify the shock-wave-induced cavitation field. Lewin *et al.*[22] and Tavakkoli *et al.*[23] built a piezoelectric shock wave generator that made it possible for them to manipulate the shape of the waveform and measure corresponding changes in cavitation-induced bioeffects. Bailey *et al.*[24,25] built a pressure-release reflector to reverse the order of the positive and negative components of the waveform while maintaining the same peak pressures, thereby producing significant changes in cavitation dynamics. Delius *et al.*,[26] Bailey,[27] Prieto and Loske,[28] Huber *et al.*,[29] and Xi and Zhong[30] all varied shock wave pulsing sequences to modify cavitation events. Indeed, some of these studies report improved stone fragmentation,[28,30] or, conversely, substantially decreased cavitation damage.[25] However, all of the methods mentioned here appear to affect the cavitation field uniformly; while cavitation damage may be intensified or mitigated, there has been no report of localization by these methods.

We demonstrate that two confocal shock wave sources, facing each other and fired simultaneously, both localize and intensify the cavitation damage *in vitro*. The Gilmore equation[31] was used to examine the cavitation dynamics for the dual pulses. Maximum bubble radius, bubble collapse time, and the acoustic pressure of shock waves radiated upon bubble collapse were calculated for positions along the focal axis by inputing a dual-pulse sequence with the appropriate interpulse delay for each position. Similarly, translational displacement and velocity were calculated using the equation for conservation of momentum. Experiments were carried out with electrohydraulic shock wave sources. High-speed photography, acoustic detection, and pitting on thin aluminum foils were used to measure maximum bubble radius, bubble collapse time, and pit depth, respectively. Relative changes in numerically and experimentally determined values for maximum bubble radius and bubble collapse time are compared.

Only radial motion of the cavitation bubbles and primary radiation (Bjerknes) forces are addressed in these studies. Results indicate a high correlation between radial motion, both calculated and observed, and the pattern of cavitation and cavitation damage observed in the dual-pulse field. Bubble–bubble interactions, fluid momentum associated with acoustic streaming, and shock wave reflections might also contribute, but were not studied here.

II. NUMERICAL METHODS

The Gilmore equation[31] was used to calculate radial bubble dynamics and the equation for conservation of momentum was used to calculate bubble translation for multiple locations in the dual-pulse cavitation field. Both models contain the following assumptions:

(i) A single bubble exists in an infinite medium (water).
(ii) The bubble remains spherical at all times.
(iii) Conditions are spatially uniform inside the bubble.
(iv) The initial bubble radius (3–30 μm) is much less than the pulse length of the acoustic excitation.
(v) Bulk viscous effects can be ignored.
(vi) The bubble is initially at equilibrium.
(vii) The air within the bubble behaves as an ideal gas.
(viii) Radial bubble motion is unaffected by bubble translation.
(ix) Dual pulses obey the principle of linear superposition.
(x) Bubble translation occurs only parallel to the focal axis.

The single bubble assumption can be challenged at the center of the dual-pulse field, where high-speed photographs reveal densely packed cavitation activity and some bubble agglomeration. Furthermore, while the first shock wave normally encounters a bubble in the micron-size range, the second shock wave may encounter a bubble that has already grown to tens or even hundred of microns. While this new ''initial'' bubble radius is still much smaller than the pulse length of the acoustic excitation, it may not be smaller than the shock thickness; hence the bubble may be subject to nonuniform compression or expansion.[32] However, in most cases, the assumption of bubble sphericity is based on experimental results, as discussed later. Linear superposition of dual pulses was also confirmed experimentally, as discussed later. Although much simpler than the clinical case, these models have been used extensively in lithotripsy and medical acoustics and have held up extremely well.[15,27,32,33]

A. Radial bubble dynamics

The radial response of a bubble in water, driven by a single lithotripter pulse or two pulses with increasing interpulse delay t_d, was numerically calculated from the Gilmore equation for bubble dynamics,[31] as described by Church:[15]

$$R\left(1-\frac{\dot{R}}{C}\right)\ddot{R}+\frac{3}{2}\left(1-\frac{\dot{R}}{3C}\right)\dot{R}^2 = \left(1+\frac{\dot{R}}{C}\right)H+\frac{\dot{R}}{C}\left(1-\frac{\dot{R}}{C}\right)R\frac{dH}{dR}, \tag{1}$$

where R is the bubble radius, $\dot{R}=dR/dt$ is the velocity of the bubble wall, $\ddot{R}=d^2R/dt^2$ is the acceleration of the bubble wall, C is the speed of sound in the water at the bubble wall, and H is the difference between the enthalpy in water at the bubble wall relative to the enthalpy in the water far from the bubble. The terms on the left side of Eq. (1) are inertial terms and those on the right side represent work done on the bubble by acoustic pressure. Pressure p and density ρ of the water are related by a modified Tait equation,[34,35]

$$p=A(\rho/\rho_0)^m-B, \tag{2}$$

where $A=c_0^2\rho_0/P_0m$, $B=A-1$, and $m=7$ are empirical constants, and ρ_0 is the equilibrium density of the liquid. Included in the approximation is the assumption of isentropic liquid flow; however, in lithotripsy, the approximation is presumed inaccurate for the short time periods when pressure in the liquid exceeds 10^4 MPa.[15] The modified Tait equation is used to calculate the enthalpy,

$$H=\int_{p_\infty(t)}^{p(R,1)}\frac{dp}{\rho}=\frac{mA^{1/m}}{(m-1)\rho_0}\{[p(R,t)+B]^{(m-1)/m}-[p_\infty(t)+B]^{(m-1)/m}\}, \quad (3)$$

and sound speed,

$$C^2=\frac{dp}{d\rho}=[c_0^2+(m-1)H]^{1/2}, \quad (4)$$

where $c_0=Am\sqrt{\rho_0}$ is the linear sound speed of the liquid, p_∞ is the pressure of the liquid far from the bubble, and $p(R,t)$ is the pressure of the liquid at the bubble interface.

Diffusion of gas across the bubble wall was calculated from the zeroth-order solution to the diffusion equation,[36] which gives the instantaneous number of moles:

$$n=n_0-4(\pi D)^{1/2}\int_0^\tau F(\tau')(\tau-\tau')^{-1/2}\,d\tau', \quad (5)$$

where D is the diffusion constant and n_0, whose value is dependent on other initial conditions, is the number of moles of gas initially present in the bubble. The terms in the integral and the solution technique are fully described by Church.[15]

The sound radiated into the water by the pulsating bubble P_r can be calculated using an approximation made by Akulichev[37] and described by Church:[15]

$$P_r=A\left\{\frac{2}{(m+1)}+\frac{(m-1)}{(m+1)}\left[1+\frac{G(m+1)}{rc^2}\right]^{1/2}\right\}^{2m/(m-1)}-B, \quad (6)$$

where A, B, and m are constants of the Tait equation and $G=R(H+U^2/2)$. Maximum pressure radiated by the bubble P_{rm}, calculated for a distance of 1 cm from the center of the bubble, was used as an indicator of the strength of a spherical collapse. Finite-amplitude propagation losses were not included in this calculation. While neglecting such losses may have affected the absolute values calculated for P_{rm}, it is unlikely to have affected relative values, as reported here. Maximum radius $R_{\max}$ and bubble collapse time t_c were also used as quantitative descriptors of cavitation activity.

The numerical code was implemented in Fortran, using a fourth-order Runga–Kutta routine. Either measured or analytical waveforms were used in the analysis. For the analytical waveforms, the equation was the same as used by Church:[15]

$$P=2KP_a e^{-\alpha r}(1-e^{\beta t})\cos\left(2\pi ft+\frac{\pi}{3}\right), \quad (7)$$

with $K=1.1$, decay constant $\alpha=9.1\times10^{-5}\,\mathrm{s}^{-1}$, frequency $f=83.3\,\mathrm{kHz}$, and $P_A=40\,\mathrm{MPa}$ unless otherwise noted. As described later, measured waveforms and delays were determined experimentally by hydrophone measurements along the focal axis of the lithotripters. Numerical results with two measured or two analytical shock waveforms were similar when normalized to a single pulse of the same type. Unless otherwise noted, numerical results presented are for the measured pulses seen at the bottom of Fig. 1. To generate each dual-pulse waveform used in calculations, the single measured pulse was added to itself with the appropriate time delay.

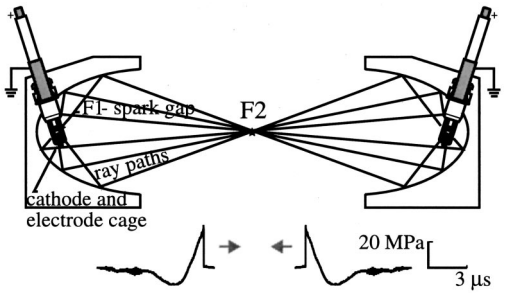


FIG. 1. Schematic diagram of dual-pulse lithotripter. Electrodes at F1 of confocal ellipsoidal reflectors are simultaneously triggered to produce opposing shock wave pulses propagating toward F2. The waveforms illustrated here were measured at the focus of a single reflector.

The values used for the physical constants are as follows: $c_0=1485\,\mathrm{m/s}$, $\rho_0=1000\,\mathrm{kg/m^3}$, $P_0=1.013\times10^5\,\mathrm{N/m^2}$, $D=2.42\times10^{-5}\,\mathrm{cm^2/s}$, $\sigma=70\,\mathrm{dyn/cm}$, $\mu=0.001\,\mathrm{Pa\cdot s}$, and $\eta=\gamma=1.4$.

B. Bubble translation

The radiation force on a bubble subjected to a single shock wave or two oppositely traveling shock waves was calculated for different positions in the dual-pulse lithotripsy field. The waveforms and bubble sizes were the same as used to calculate radial dynamics. The translation of a bubble in the direction of shock wave propagation is estimated using momentum conservation in one dimension, as described by Watanabe and Kukita:[38]

$$m_b\frac{du_b}{dt}=-V\frac{dP}{dz}-\frac{1}{2}\rho_0\frac{d}{dt}(Vu_r)-\frac{1}{2}\rho_0|u_r|u_rAC_d, \quad (8)$$

where m_b and V are the mass and volume of the gas inside the bubble, z is the distance along the reflector's focal axis, $u_b=dz/dt$ is the translational velocity of the bubble, $u_r=u_b-u_l$ is the relative velocity between the bubble and the liquid, A is the projected area of the bubble, and C_d is the drag coefficient. The velocity of the liquid, u_I is calculated by

$$u_I=\frac{\Delta p}{\rho_0 c_0}=\frac{p_{sw1}}{\rho_0 c_0}-\frac{P_{sw2}}{\rho_0 c_0}, \quad (9)$$

where p_{sw1} and p_{sw2} are the time dependent shock wave pressure profiles for pulses 1 and 2, respectively. The drag coefficient is expressed in terms of the Reynolds number:

$$C_d=27.0\,\mathrm{Re}^{-0.78}, \quad (10)$$

where the Reynolds number is defined by

$$\mathrm{Re}=\frac{2\rho_0 u_r R}{\mu}. \quad (11)$$

The force on the bubble, given by Eq. (8), was numerically integrated to determine bubble velocity u_b and position δ as a function of time. The drag coefficient given in Eq. (10) was proposed by Crum[39] as a best fit to measurements on oscillating air bubbles translating through liquids. Crum's measurements were made up to a Reynolds number of 200.

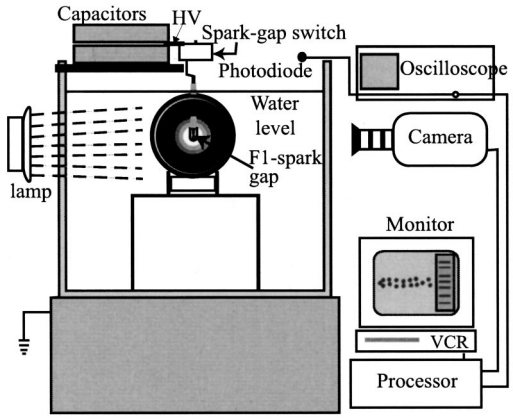


FIG. 2. Transverse view of dual-pulse lithotripter and instruments used in imaging of cavitation fields with a high-speed video camera. Only one reflector is pictured here. The cavitation field is illuminated from the left, and filmed from the right, through the acrylic walls of the tank. Photodiodes, recording the sparks at F1, were used to trigger the camera as well as the oscilloscope.

Eller and Flynn[40] calculated that rapidly translating bubbles remain stable, approximately in the shape of an oblate spheroid, for a wide range of conditions. In fact, they predicted stability for a relative bubble velocity of up to 50 m/s for 0.1 μm bubbles, which is equivalent to a Reynolds number of 10 in water, well within the range of Crum's analysis. The maximum Reynolds number in this analysis exceeds 10 by several orders of magnitude, sometimes predicting bubbles velocities of up to 10^{10} m/s. Given Eller and Flynn's prediction for shape stability of translating bubbles, and extrapolating their analysis to the minimum bubble radius in our own analysis, we assumed bubble destruction if relative bubble velocity exceeded approximately 100 m/s at any time before the inertial collapse.

Numerical computation of bubble translation was implemented in Matlab (Mathworks, Natick, MA) using a fourth-order Runga–Kutta routine. First, as previously described, values for R and $\dot{R}$ were calculated using Eq. (1). These values were subsequently used in the radiation force calculations, maintaining the same acoustic pressure distributions, initial bubble sizes, and physical constants previously defined.

III. EXPERIMENTAL METHODS

A. Overview

All experiments were conducted with electrohydraulic shock wave sources modeled after that of the Dornier HM3 clinical lithotripter (Dornier Medical Systems, GmbH). Figures 1 and 2 illustrate the main components of the dual-reflector lithotripter, including a water bath, spark sources consisting of high-voltage capacitors and underwater electrodes, and brass reflectors. Each spark, generated with refurbished Dornier electrodes (Service Trends, Kennesaw, GA), produces a spherically diverging pressure wave at the internal focus, F1, of the ellipsoidal reflector. This wave reflects from the ellipsoidal surface and converges at the external focus, F2.

Three techniques were used to measure and record cavitation activity: high-speed digital photography, passive cavitation detection, and pitting in thin aluminum foil. Analyses were carried out primarily for bubbles at various positions along the z axis. To differentiate measured from calculated variables, measured variables are labeled with an asterisk.

B. Water bath

The water bath is contained in an acrylic tank with a width of 59 cm and length of 95 cm. The bottom of the tank has a slant of 14 degrees, similar to that of the Dornier HM3. To simplify positioning of the two reflectors, an acrylic shelf was built to create a horizontal surface on the bottom of the tank. The reflectors sit on acrylic bases approximately 7 cm in height and the water level for experiments was 0.5 cm above the top surface of the reflectors.

A water-degassing system is incorporated into the research lithotripter, as described by Cleveland.[41] Gas content was measured with a model 51B oxygen meter (YSI, Incorporated, Yellow Springs, OH). Before each experiment, the water was partially degassed to 3–4 parts per million of O_2. Due to the design of the filtering and degassing system, as well as the large volume and exposed surface area of the water tank, it is not possible to completely degas the water or to keep out all particulate matter that may trap cavitation nuclei. Conductivity of the water was maintained at 600 μS/cm and tested with a Model EP water conductivity meter (Myron L. Company, Carlsbad, CA).

C. Spark sources and reflectors

Figures 1 and 2 illustrate some of the main components and configuration of the dual-pulse lithotripter. Two ellipsoidal brass reflectors were positioned on the horizontal shelf such that their major axes were collinear, with a common external focus F2. Each reflector was part of a system consisting of the reflector, an electrode, a pair of 40-nF low-inductance capacitors and a nitrogen-enclosed spark-gap switch (EG&G, Princeton, NJ). Both systems were connected, through high-voltage diodes, to a high-voltage source set at 18 kV for most experiments. Sparks from the two electrodes were simultaneously triggered by a 10-V square pulse, output by a function generator (model 33120A, Hewlett Packard, Palo Alto, CA), and sent in parallel to the two spark-gap switches. As suggested by Coleman *et al.*[42] each electrode was ''burned in'' with 150 sparks before use and was replaced after 2000 sparks. Coleman *et al.* showed that after electrodes were burned in, the pulse amplitude increased 0.01%/3 sparks and there was no noticeable change in wave shape.

One reflector is a Dornier HM3 reflector with dimensions: major half axis $a=13.8$ cm, minor half axis $b=7.75$ cm, and depth $=12.5$ cm. The other reflector has the same ellipsoidal dimensions and a depth of 13.5 cm, producing no measurable difference in waveform or amplitude.

D. Measurement of pressure wave forms

Axial waveforms from our conventional lithotripter were previously measured by Bailey *et al.*[24] They positioned

a reference shock wave hydrophone (Sonic Technologies, Hatboro, PA), consisting of a 5×5 cm PVDF membrane window held in a plastic cassette, along the z axis of a single reflector. This hydrophone was calibrated, by the manufacturer, by substitution calibration with a NIST-traceable hydrophone. The geometric diameter of the sensitive element on the hydrophone was less than 0.5 mm. Peak pressures were averaged for ten measurements at each of several locations along the z axis. A peak positive pressure value of approximately 40 MPa was found for a charging voltage of 18 kV. The duration of the positive spike is approximately 1 μs and the total pulse duration is approximately 4–5 μs. This waveform is in good agreement with those measured in the HM3 by Coleman[42] and Cleveland.[41]

A similar hydrophone was placed along the z axis of the dual-pulse lithotripter to confirm that dual pulses, in terms of waveforms and peak pressures, obey the principle of linear superposition. That is, the fields from two coaxial, opposing reflectors superimpose to produce an acoustic field that is essentially twice the magnitude of that of a single reflector, at the focus. Figure 3 demonstrates three measured waveforms: a single pulse at $z=0$, dual pulses at $z=0$, and dual pulses at $z=2$ cm. These particular waveforms, while plotted here on the same voltage scale, were not converted to a pressure scale and not used in any calculations. The waveform in Fig. 3(b1) is clipped because the vertical scale on the oscilloscope was set too small to capture the full positive peak. The positive peak pressure is less than twice the single pulse peak pressure, and broader, because the positive peaks are ~1 μs in duration and the two pulses did not perfectly overlap in this measurement. Negative pressure nearly doubles in value from the single pulse to the dual pulses at $z=0$.

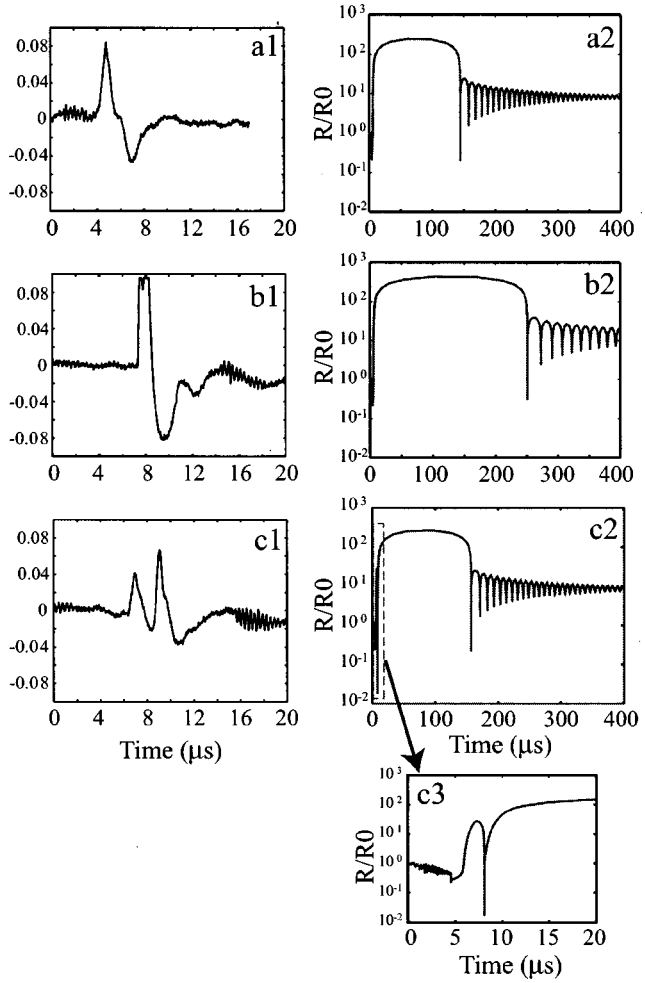


FIG. 3. Numerical radius-time curves (right column) for corresponding representative measured waveforms (left column) in regions A and B of the dual-pulse cavitation field and for a single pulse at $z=0$. R_0 is 3 μm. In (a2), a single shock wave pulse (a1) causes a bubble to grow to a maximum radius of 246 μm and collapse at 144 μs. In (b2), two such pulses arrive simultaneously (b1), resulting in a maximum bubble radius of 440 μm and a bubble duration of 250 μs. This corresponds to the case at $z=0$. In (c2), a bubble situated in region B experiences two pulses in quick succession (c1). The first 20 μs of the trace in (c2) are expanded in (c3) to show that the positive peak of the second pulse acts on the rapidly expanding bubble to cause a premature forced collapse.

E. Cavitation imaging by high-speed digital photography

Cavitation at and around F2 was initially recorded with a Kodak Ektapro 4540 high-speed digital camera (Eastman Kodak Company, Rochester, NY) with frame rates as high as 18 000 frames per second. The main components of the Kodak imaging system are included in Fig. 2. The camera was triggered by the signal from a photodiode that recorded the spark at F1. The cavitation field was illuminated with a 1000-W light bulb, either from behind or in front of the cavitation field, relative to the camera. At the frame rate used in these studies, the sensor displayed 64×128 pixel images, presented in 256 grayscale levels, with a field of view as small as $2\times6.7\,\mathrm{cm}^2$. Pixel size was >0.16 mm^2. Images were stored on VHS videocassette and digitized using a video capture card on a Macintosh computer.

Cavitation fields were later recorded with an Imacon 200 high-speed digital camera (DRS Technologies, Cupertino, CA) with interframe and exposure times both as low as 5 ns. The camera and a flash lamp were triggered by a TTL signal achieved by an analog to digital conversion of the current discharge signal generated by a coil around the high-voltage capacitor grounding cable. The field of view (FOV) used in these experiments was as small as $0.44\times0.54\,\mathrm{cm}^2$. The Imacon camera had eight double-gated channels, allowing for capture and storage of 16, 1280×1024 CCD pixel frames at a time. Pixel size was >18 μm^2. Images were stored on CD-ROM for later analysis on a Macintosh computer.

Field of view was determined for each camera and lens setting by imaging a ruler in the focal plane of the camera. Given the known FOV for each image, maximum bubble radii $R^*_{\max}$ were measured directly from the digitized images viewed on a Macintosh Computer with Adobe Photoshop (Adobe Systems, Inc., San Jose, CA). Bubble density was also determined from these images, using a volume defined by the FOV and measured depth of the cavitation field.

F. Passive acoustic detection of cavitation

Cavitation collapse was also detected acoustically using a passive cavitation detection (PCD) system. In this method, cavitation was detected by measuring the shock waves emitted during bubble collapse.[32,33,43] The measurement transducer contains a concave piezoceramic sensing element with

a 200-mm radius of curvature, a 100-mm aperture, a 1.08-MHz resonance frequency and a 3.9-mm full width half-maximum focal diameter. The acoustic axis of the detector was perpendicular to the focal axis of the reflector(s). A high-pass filter (model 3202, Krohn-Hite, Avon, MA) with a 300-kHz cutoff frequency was used to remove the low-frequency signal of the radial mode of the transducer. An oscilloscope (Model TDS 744A, Tektronix, Inc., Beaverton, OR) with a 2 GS/s digitizing rate and a 500-MHz analog bandwidth was used to acquire and digitize signals from the transducer. The oscilloscope was triggered by the same TTL signal used to trigger the Imacon camera, and a 180-μs delay was added between receipt of the trigger and the start of data acquisition.

Collapse time t_c^* was defined as the time between receipt of the signal corresponding to the arrival of the lithotripsy pulse at the focus, and receipt of the radiated pressure signal from the collapsing bubble. Church and others[15,25] have used t_c^* as an indicator of cavitation intensity, with increasing t_c^* corresponding to increasing collapse strength. The PCD measurements were used to determine collapse times for bubbles at different locations in the cavitation fields, including $z=0$ and $z=2$ cm. For two pulses, the position, z, is related, by the speed of sound, to the delay in time between the arrival of the first and second pulse at that point, such that $z=0$ corresponds to an interpulse delay, $t_d^*=0$. The second pulse is delayed relative to the first for any point $|z|>0$ because the distance traveled to this point by a shock wave originating from the far reflector is greater than that of a shock wave originating from the near reflector. Because of small jitter in timing (<3 μs) between firing of the two electrodes, associated with uncertainty in triggering, the $z=0$ position shifts along the axis by as much as 3 to 4 mm from shot to shot. Variation in t_c^* along the axis is expected because of the changes in negative pressure amplitude as the shock wave propagates. Variation at each location is expected because of small spark-to-spark jitter in strength, shape, and location of the spark. Even a 1-mm shift in the location of the spark can lead to a noticeable shift in the location and size of the focal volume.[44] Cleveland *et al.*[41] measured the variation in positive and negative peak pressures and found that positive peak pressure varied by as much as 15% while negative peak pressure varied by less than 3%.

G. Cavitation detection by pitting of aluminum foil

When a bubble collapses near a rigid surface, such as a stone, it may collapse asymmetrically, generating a high-speed water jet directed toward the surface.[5–9] The depth and distribution of pits created on thin foils has been used to quantify damage from lithotripsy cavitation fields *in vitro*.[18,25,45,46] The respective sizes and intensities of the single-pulse and dual-pulse cavitation fields at collapse were studies using 25-μm-thick sheets of aluminum foil placed at F2 ($z=0$), in a plane containing the z axis, as illustrated in Fig. 4. Foils were $9\times12\,\mathrm{cm}^2$ and were stretched taut and secured over the face of a plastic (PVC) semi-cylindrical frame of inner diameter 9.6 cm, such that the frame itself

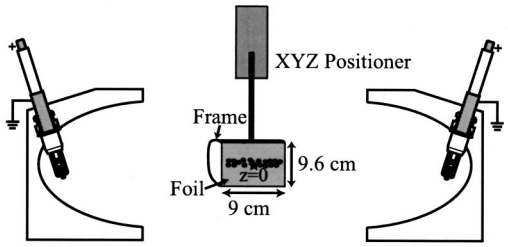


FIG. 4. Diagram of position and size of 25-μm-thick aluminum foils used in experiments. The approximate location and size of the conventional lithotripter cavitation field are indicated by the black dots on the foil surface.

was >2.8 cm outside of the focal volume at all times. Foil and frame were positioned and supported from above using an XYZ positioning system (Velmex Industries, Bloomfield, NY).

Following exposure to either a conventional lithotripter (CL) or dual-pulse lithotripter (DPL) cavitation field, each foil was carefully removed from its frame and mounted on flat, smooth sheets of Lexan using tape around the edges only. Care was taken to minimize any disturbance to the cavitation pits on the foils. Foils were digitally scanned onto a Macintosh computer for viewing with Adobe Photoshop and quantitatively analyzed using a noncontact profilometer (New View200, Zygo Corporation, Middlefield, CT). Peak-to-valley (*PV*) distances of disturbances in the foil surface were measured for 2.1×2.8 mm square regions along the entire focal length of the CL foils and at $z=0$ and $z=2$ cm on the DPL foils. *PV* was chosen as the value of interest because the foils, having bisected the cavitation field, contained pits generated from both sides.

IV. RESULTS

First, results for the conventional lithotripter field are presented. Subsequently, experimental and numerical results for the dual-pulse lithotripter are presented by location in the cavitation field and compared to those for the focal region of the conventional lithotripter. Finally, results of bubble translation calculations are presented for the single- and dual-pulse fields.

A. The CL field

Figure 5 contains a chronological series of photographs for both the conventional lithotripter (CL) and dual-pulse lithotripter (DPL) fields. The CL cavitation field is approximately 1.5 cm in diameter and 10 cm long, at its maximum. The estimated density of the CL field in Fig. 5(a), at maximum radius $R_{\max}^*$, is 70 bubbles per cubic centimeter, corresponding to an average center to center bubble spacing of ~3 mm. $R_{\max}^*$ for individual bubbles is 0.5 mm. High-speed videos show that bubbles in the CL field grow and collapse in 300 ± 20 μs. PCD measurements yielded collapse times of 340 ± 31 μs at F2, 343 ± 31 μs at F2$-$2 cm (prefocus), and 307 ± 18 μs at F2$+$2 cm (postfocus).

Figure 6(a) contains an image of a foil subjected to a single CL pulse that originated from the left. The collapse of CL bubbles results in a line of pits, approximately 1 mm in width and 8 cm in length. The pit depths along this line are

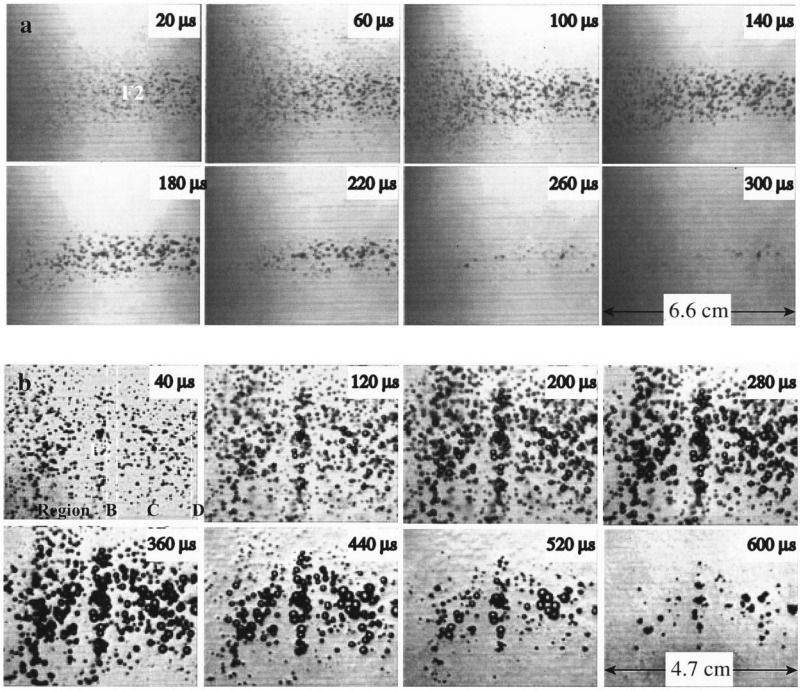


FIG. 5. Time series of front-lit (a) conventional lithotripter (CL) and (b) dual-pulse lithotripter (DPL) cavitation fields in water. In the CL field, a single shock wave pulse arrives from the right. In the DPL, one pulse arrives from each side to meet simultaneously at F2 ($z=0$). The DPL field of view (FOV) is 30% smaller than CL FOV. The time on each frame indicates time after the pulse(s) arrive(s) at F2. For this particular experiment, the water was not degassed as in all other experiments. The white texture in the last four frames of the DPL field is an imaging artifact resulting from double gating of the framing channels.

distributed with a mean $PV=56.73\pm18.07\,\mu$m and maximum PV, $PV_{\max}=95.59\,\mu$m. Qualitative inspection of the CL foils did not lead to any conclusions about distribution of pit depths along the 8-cm line.

B. The DPL field: Localization and intensification

Figure 5(b) contains a series of images of the dual-pulse cavitation field. The dual-pulse lithotripter (DPL) field is axially shorter (<5 cm) than the 10-cm long conventional lithotripter (CL) cavitation field; therefore, the field of view of the DPL images is smaller than that of the CL images. In the transverse dimension, the DPL field is 4 cm across, as compared to the 1.5-cm-diam CL field. However, while the CL field is nearly uniform in bubble size, time to collapse, and intensity of collapse along the length of thin foils, the DPL field concentrates its most destructive cavitation activity at F2. The region of damage is very similar in appearance to the central band in Fig. 5(b), which contains the largest and longest-lasting bubbles. The foil in Fig. 6(b) especially illustrates the localized nature of the DPL damage. Instead of the long line of damage along the reflector's axis seen in Fig. 6(a), the most intense damage from the DPL, while longer transverse to the axis, is isolated to within just 2 mm of F2 in the axial direction.

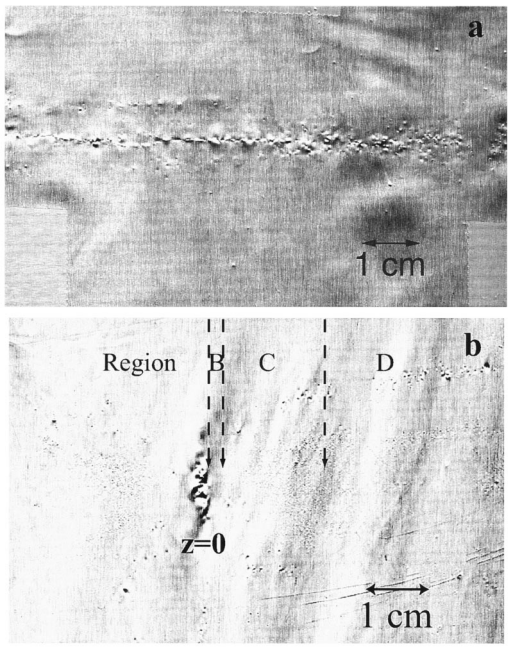


FIG. 6. Aluminum foils, 25 μm thick, that were placed in a cavitation field to record the collapse of the bubbles as pits in the surface. In (a) a single shock wave pulse, originating from the left, generates an 8-cm-long, thin line of pits. In (b), two pulses, one from each side of the foil, meet simultaneously at $z=0$ to generate extensive damage at that location only, and counteract each other in regions B, C, and D to minimize cavitation and its impact on the foil.

In addition to the localization, the dual-pulse lithotripter cavitation field has a banded structure, symmetric about the center line, $z=0$, where the pulses arrive simultaneously. As shown in Fig. 5(b), the DPL cavitation field may be divided into four regions of analysis symmetric about $z=0$: (A) $0<z<1$ mm, (B) $1<z<3$ mm, (C) $3<z<20$ mm, and (D) $z>20$ mm. As previously described, interpulse delay is related to position in the field by $t_d=2z/c_0$. Figure 7 demonstrates calculated values of $R_{\max}$, t_c, and P_{rm} normalized to those for a single pulse for regions A, B, and C.

In region A, calculations of $R_{\max}$, t_c, and P_{rm} yield values nearly double the corresponding values for a single pulse. Figure 3 demonstrates radius-time (R-t) curves for representative measured waveforms in regions A and B of the dual-pulse cavitation field and for a single pulse at $z=0$. The column on the left shows the oscilloscope trace of the waveforms, with the y axis having values of volts. The column on the right shows the corresponding numerically calculated R-t traces. Figure 3(b2) demonstrates a radius-time (R-t) curve for a 3-μm bubble subject to two superim-

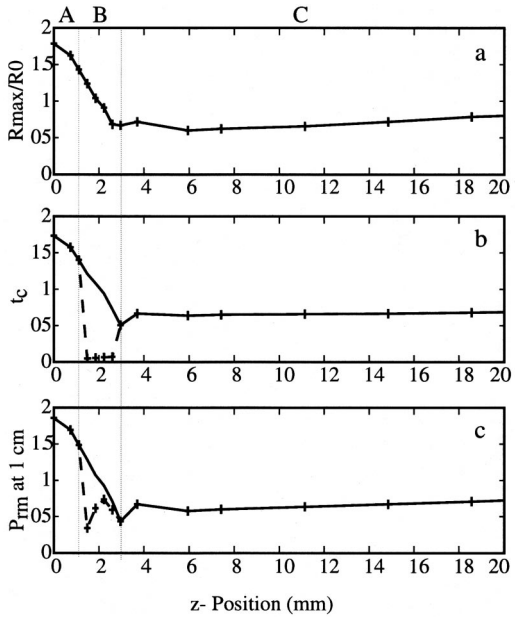


FIG. 7. These curves demonstrate numerically calculated (a) maximum radius R_{max}/R_0 (b) time to collapse t_c, and (c) maximum radiated pressure P_{rm}, for $R_0=3$ μm bubbles in regions A, B, and C of the dual-pulse cavitation field. All values are normalized to values calculated for a single pulse acting on the same bubble at $z=0$. The dashed lines in curves (b) and (c), within region B, indicate t_c and P_{rm} assuming that the bubbles that experience a forced collapse from the second pulse, as in Fig. 3(c3), do not regrow.

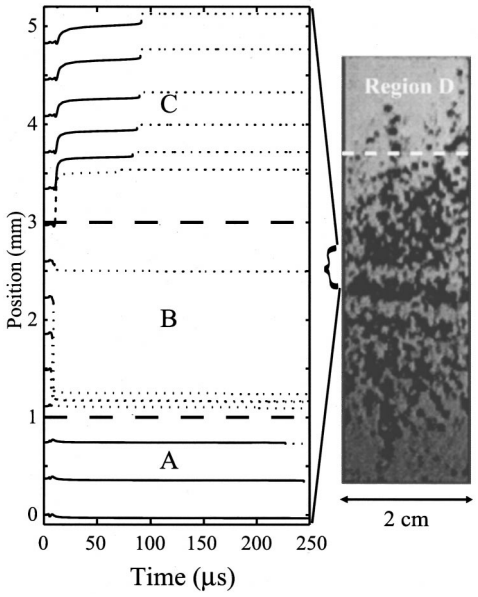


FIG. 8. These curves trace, in time, the calculated translation of bubbles within 5 mm of focus. The *y* axis of the plot designates position along the *z* axis of the reflector, with the first point of each trace corresponding to the initial position of a stable 3-μm bubble. When translation velocity exceeds ~100 m/s for a given bubble, the corresponding solid translation curve becomes a dotted curve. The photograph to the right of the figure is a snap shot in time of the bubbles near their maximum size, corresponding to a time of ~50 μs on the translation plot. Regions A, B, and part of C are designated by the heavy dashed lines at 1 and 3 mm on the translation plot and the border of region D is marked by the dashed white line on the photograph.

posed pulses at $z=0$. This is compared to the R-t curve in Fig. 3(a2), in which the same 3-μm bubble is subject to a single pulse at $z=0$. In Fig. 3(a2), the absolute values for t_c (144 μs) and R_{max} (246 μm) are less than experimental values, as others have reported.[15,25,33] This discrepancy might be explained by underrepresentation of the negative pressure in the shock wave by the membrane hydrophone. As elaborated by Bailey,[25] Staudenraus and Eisenmenger[47] proposed that under strong negative pressure, water tore away from the metalized sensitive element of the PVDF membrane hydrophone and thus cut short the negative tail of the measured waveform. However, the relative doubling of values from the single reflector to the dual reflectors is consistent in both experiment and calculations, and corresponds to the doubling of acoustic pressure at this position.

Experimental results in Fig. 5 demonstrate that R^*_{max} and t^*_c, for the dual pulses are also approximately double those for a single pulse. R^*_{max} changes by a factor of 2.2, from 0.5 to 1.1 mm, while t^*_c increases by a factor of 1.8 from 340 ± 31 μs to 608 ± 33 μs. Figure 6(b) contains an image of a foil subjected to the DPL cavitation field. The DPL field creates a $\sim 2\times 5$ mm^2 "crater" at $z=0$, with $PV=78.05$ ± 39.66 μm, approximately 1.4 times the value for CL ($p<0.03$), and $PV_{max}=123.4$ μm.

In region C, calculations yield values of R_{max}, t_c, and P_{rm} all in the range 0.6–0.7 times the corresponding values for a single pulse. High-speed videos reveal that the density of the bubble field in region C is lower than at $z=0$, and R^*_{max} and t^*_c values are measured to be 0.5–1.1 mm and 370–570 μs, respectively, equal to or greater than the corresponding CL values. The discrepancy between calculated and measured values in this region may be due to the fact that the numerical analysis does not include changes in the waveform as it propagates.

As demonstrated in Fig. 6(b), foil pits in region C are very shallow or even nonexistent. At the outer edge of region C ($z=2$ cm), $PV=18.46\pm 7.34$ μm and $PV_{max}=35.93$ μm, approximately one-fourth of the values recorded at $z=0$ in the DPL field, and approximately one-third of the values recorded for the CL field ($p<0.01$, student t-test).

In region D, bubbles that begin to grow when the first pulse passes disappear after the second pulse arrives. We model the second pulse as larger than the first because shock waves focused by the ellipsoidal reflectors used for these studies reach maximum positive amplitude at approximately 20 mm past the geometric focus F2, as verified by Bailey *et al.*[24] Numerical calculations with analytical waveforms indicate that, for positions $z>20$ mm, the larger positive peak of the second pulse can disrupt the growth of the bubble, resulting in a much reduced t_c.

Experimental measurements show that t^*_c and PV are both reduced in region D. The foil in Fig. 6(b) as well as the image in Fig. 8 most clearly indicate the absence of bubbles in this region. PCD measurements in this range show only weak signals in the time range of 370–530 μs, possibly corresponding to a few stray bubbles still within the focus of the detection transducer or strong collapses well outside the focus of the detector and closer to $z=0$. Because of the disrupted bubble growth in regions C and D, cavitation damage is localized to region A.

Figures 5(b) and 6(b) demonstrate the absence of bubbles and damage to foil, respectively, in region B. Calculations indicate that for the small interpulse delays that exist in this narrow region, there is a collapse immediately following the arrival of the second pulse, as shown in Figs. 3 (c2) and (c3). This forced collapse yields values for minimum bubble radius $R_{\min}$ as low as $R_0/100$, often much smaller than $R_{\min}$ calculated for an inertial collapse, most likely because too little time has passed for significant diffusion of gas into the bubble. P_{rm} values for the forced collapse vary from 0.3 to 0.7 times the value for the inertial collapse due to a single pulse at the focus. The dashed curves in Figs. 7(b) and (c) demonstrate the full range of interpulse delays for which the premature collapse is calculated to occur.

Even with a time resolution of 0.5 μs, only a few bubbles were ever detected in this region using the high-speed camera. The PCD, with a focal diameter of 3.9 mm, did not have adequate resolution to limit measurement of radiated sound to these narrow regions in the dual-pulse lithotripsy field. The role of cavitation dynamics and bubble translation in these gaps will be discussed later.

C. Bubble translation

The radius-time curves presented previously are in good agreement with the observed localization of the dual-pulse cavitation field. However, the role of bubble translation in localization was also investigated. Calculations predict that a single pulse at the focus causes a 3-μm bubble to translate <20 μm, with a maximum velocity <100 m/s. No axial translation is observed on high-speed video of the conventional lithotripter field.

Figure 8 presents calculated translation curves for the dual-pulse field. The figure includes translation curves for a 3-μm bubble at several initial positions within 5 mm of the focus as well as a photograph of the cavitation field created by two opposing pulses that originated from above and below the image. Time is on the *x* axis and *z* position is on the *y* axis of the plot. Each curve traces, in time, the movement of a bubble toward or away from $z=0$. Each solid curve, representing a single, spherical bubble, becomes a dotted curve as soon as the translation velocity for that bubble exceeds 100 m/s. The heavy dashed lines at 1 and 3 mm designate the borders for regions A, B, and C. In region A, net bubble translation is <0.1 mm and in region C, net bubble translation is <0.3 mm. In both A and C, bubble velocities do not exceed approximately 100 m/s until the final, inertial collapse, around 100 μs for region C. However, for the bubbles beginning at positions between approximately 1 and 3 mm—that is, for the bubbles in region B—calculated translation velocities exceed 100 m/s by several orders of magnitude as the second, delayed pulse acts on the growing bubble. The solid curves become dotted at this time of <10 μs, resulting in the approximately 2-mm ''gap'' evident in the figure.

Calculations were also carried out for bubbles up to 30 μm in radius. Net translation was slightly greater for larger bubbles in regions A and C, and approximately the same size and location of the region-B-like ''gap'' occurred close to the focus. No bubble translation, in any position in the dual-pulse lithotripter field, was observed in experiment.

V. DISCUSSION AND CONCLUSIONS

When two pulses are generated simultaneously from identical, opposing, and confocal shock wave sources, the resulting cavitation field demonstrates a localization and intensification of cavitation damage, as compared with that of a single shock wave source. Both numerical and experimental data indicate that radial bubble dynamics assume a principal role in the formation of the dual-pulse cavitation field. Bubbles at $z=0$, the focus of the lithotripter with *z* as the reflectors' common major axis, are subject to a doubling of acoustic pressure. Consequently, bubbles grow nearly twice as large, last nearly twice as long, and, upon collapse, radiate shock waves with nearly twice the acoustic pressure amplitude. Some bubble agglomeration was also observed at the focus of the dual-pulse cavitation field. In the 2-cm regions pre- and postfocal, spherically collapsing bubbles are calculated to radiate waves with only 40% of the pressure amplitude of waves radiated by bubbles at the focus. The results of the experiments with thin foils demonstrate that asymmetric collapses on foil are also considerably weaker, with pit depths decreasing by a factor of 4, and, on some foils, no visible pits in this area. Variation between foils may be explained by spatial jitter in the generation of the shock waves. Such jitter may cause slight movements of the focal axis away from the foil surface, as well as amplitude variations between the two pulses. If cavitation bubbles in 2-cm regions on either side of the focus are near the threshold for pitting in the Al foil, these small changes in the acoustic field, hence the cavitation field, may explain the absence of pits on many foils. No significant cavitation activity is observed or calculated in the regions beyond $|z|=2$ cm.

Bubble-free gaps were observed in high-speed videos and on foils in the region $1<|z|<3$ mm. Calculations showed that an instability is generated in this region because of the timing between the two shock wave pulses acting on the bubbles. Calculations of radial dynamics predict a rapid, yet weakly radiating forced collapse from the positive peak of the second pulse, yielding much smaller minimum radii than predicted for an inertial collapse. Because our model contains the assumption of bubble sphericity, calculations also indicated that bubbles would survive this forced collapse and rebound to a new maximum. It is possible that no such rebound occurs, and bubbles are destroyed at some point during this collapse. Calculations of translation, which is dependent on bubble wall velocity, predict high axial velocities and accelerations in this same region. Eller and Flynn[40] calculated that bubbles exceeding velocities of approximately 100 m/s no longer have shape stability. Therefore, the high radial and translation velocities in our analysis both point toward bubble destruction in this narrow gap. The underlying assumption of bubble destruction leads to the dashed collapse time versus interpulse delay curve presented in Fig. 7(b). Experiments support the conclusion that the bubbles are destroyed rather than just pushed away. On high-speed videos, there are only a few bubbles in this region where high radial and translation velocities occur immedi-

ately following the arrival of the second pulse. In contrast, regions of high bubble density maintain low translation and radial velocities until the final, inertial collapse. Finally, no bubble translation is observed experimentally.

It has been suggested that the dual-pulse cavitation field *in vivo* might result in a volume of damage confined primarily to the kidney stone and not extending into the surrounding renal tissue. In addition, the rate of stone fragmentation may increase significantly because of the increase, with each shot, in cavitation activity at the site of the stone. However, *in vivo* cavitation dynamics may vary greatly from those observed *in vitro*. Furthermore, if a large kidney stone of 10 mm is present at the beam focus, it is reasonable to expect that even *in vitro* cavitation activity would be different from that presented in this article. Claus-Dieter Ohl *et al.*[48] studied the effect of a rigid boundary on the explosive growth and translation of bubbles near the boundary. Measurements of stone fragmentation in the dual-pulse lithotripter have been made and presented previously.[49,50] In short, *in vitro* stone fragmentation, with artificial stones, was more than doubled at the focus of the dual-pulse lithotripter as compared to the conventional lithotripter. A more detailed analysis of *in vitro* stone fragmentation and tissue damage is currently underway.

At present, only a single shock wave source is used to treat kidney stones with ESWL. There are anatomic constraints on the positioning of the reflector relative to the human kidney. For example, the propagation path should not contain bone or air and ideally would not contain organs other than the kidney itself. Perhaps two reflectors can be oriented at an angle less than 180 degrees with respect to one another to produce the same localized effect. As an alternative approach, a localized and intensified field might be generated by the appropriate phasing of shock waves from a single source. In either case, the investigation reported here shows that two pulses can be used to intensify or mitigate cavitation activity, and that by spatially distributing the temporal delay between pulses, cavitation can be localized.

ACKNOWLEDGMENTS

This work was supported by NIH DK43881 and a NSF Graduate Research Fellowship. A portion of the research described in this article was performed in the Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the Department of Energy's Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory (Richland, WA). The authors would like to thank Dr. Glen Dunham of EMSL for his assistance with the profilometer measurements. The authors would also like to thank Dr. Tom Matula and Dr. Steve Kargl (APL) and Dr. Oleg Sapozhnikov (Moscow State University) for their helpful discussions and assistance with experimental and numerical methods to describe cavitation activity and Dr. Skip Kolve (APL) for his technical assistance with instrumentation. Finally, the authors would like to thank Dr. Claus-Dieter Ohl (TU Twente, The Netherlands) for his early correspondence and input regarding calculations of bubbles translation.

[1] A. P. Evan, L. R. Willis, B. Connors, G. Reed, J. A. McAteer, and J. E. Lingeman, "Shock wave lithotripsy-induced renal injury," Am. J. Kidney Dis. **17**(4), 445–450 (1991).
[2] A. P. Evan and J. A. McAteer, "Q-Effects of shock wave lithotripsy," in *Kidney Stones: Medical and Surgical Management*, edited by F. Coe, C. Pak, and G. M. Preminger (Raven, New York, 1996), Vol. 1, pp. 549–570.
[3] A. P. Evan, B. A. Connors, D. J. Pennington, P. M. Blomgren, J. E. Lingeman, N. S. Fineberg, and L. R. Willis, "Renal disease potentiates the injury caused by SWL," J. Endourol. **13**(9), 619–628 (1999).
[4] L. R. Willis, A. P. Evan, B. A. Connors, P. Blomgren, N. S. Fineberg, and J. E. Lingeman, "Relationship between kidney size, renal injury, and renal impairmen t induced by shock wave lithotripsy," J. Am. Soc. Nephrol. **10**(8), 1753–1762 (1999).
[5] C. F. Naude and A. T. Ellis, "On the mechanism of cavitation damage by nonspherical cavities collapsing in contact with a solid boundary," ASME J. Basic Eng. **83**, 648–656 (1961).
[6] J. K. Walters and J. F. Davidson, "The initial motion of a gas bubble formed in an invsicid liquid. Part 1. The two-dimensional bubble," J. Fluid Mech. **12**, 408–416 (1962).
[7] J. K. Walters and J. F. Davidson, "The initial motion of a gas bubble formed in an inviscid liquid. Part 2. The three-dimensional bubble and the toroidal bubble," J. Fluid Mech. **17**, 321–336 (1963).
[8] T. B. Benjamin and A. T. Ellis, "The collapse of cavitation bubbles and the pressures thereby produced against solid boundaries," Philos. Trans. R. Soc. London, Ser. A **260**, 221–240 (1966).
[9] L. A. Crum, "Acoustic cavitation," presented at the Ultrasonics Symposium, 1982.
[10] L. A. Crum, "Cavitation microjets as a contributory mechanism for renal calculi disintegration in ESWL," J. Urol. (Baltimore) **140**(6), 1587–1590 (1988).
[11] W. Sass, M. Braunlich, H. P. Dreyer, E. Matura, W. Folberth, H. G. Preismeyer, and J. Seifert, "The mechanisms of stone disintegration by shock waves," Ultrasound Med. Biol. **17**(3), 239–243 (1991).
[12] W. Sass, H. P. Dreyer, S. Kettermann, and J. Seifert, "The role of cavitational activity in fragmentation processes by lithotripters," J. Stone Dis. **4**(3), 193–207 (1992).
[13] A. J. Coleman and J. E. Saunders, "A review of the physical properties and biological effects of the high amplitude acoustic field used in extracorporeal lithotripsy," Ultrasonics **31**(2), 75–89 (1993).
[14] A. Vogel and W. Lauterborn, "Acoustic transient generation by laser-produced cavitation bubbles near solid boundaries," J. Acoust. Soc. Am. **84**, 719–731 (1988).
[15] C. C. Church, "A theoretical study of cavitation generated by an extracorporeal shock wave lithotripter," J. Acoust. Soc. Am. **86**, 215–227 (1989).
[16] T. R. Morgan, V. P. Laudone, W. D. Heston, L. Zeitz, and W. R. Fair, "Free radical production by high energy shock waves: Comparison with ionizing radiation," J. Urol. (Baltimore) **139**, 186–189 (1988).
[17] M. Lokhandwalla, J. A. McAteer, J. C. Williams, Jr., and B. Sturtevant, "Mechanical hemolysis in shock wave lithotripsy (SWL): II. Cell lysis due to shear," Phys. Med. Biol. **46**, 1245–1264 (2001).
[18] D. A. Lifshitz, J. C. Williams, Jr., B. Sturtevant, B. A. Connors, A. P. Evan, and J. A. McAteer, "Quantitation of shock wave cavitation damage in vitro," Ultrasound Med. Biol. **23(3)**, 461–471 (1997).
[19] J. C. Williams, Jr., J. F. Woodward, M. A. Stonehill, A. P. Evan, and J. A. McAteer, "Cell damage by lithotripter shock waves at high pressure to preclude cavitation," Ultrasound Med. Biol. **25**(9), 1445–1449 (1999).
[20] A. Sonden, B. Svensson, N. Roman, H. Ostmark, B. Brismar, J. Palmblad, and B. T. Kjellstrom, "Laser-induced shock wave endothelial cell injury," Lasers Surg. Med. **26**(4), 364–375 (2000).
[21] T. Kodama and Y. Tomita, "Cavitation bubble behavior and bubble-shock wave interaction near a gelatin surface as a study of *in vivo* bubble dynamics," Appl. Phys. B: Lasers Opt. **B70**, 139–149 (2000).
[22] P. A. Lewin, J. Y. Chapelon, J. L. Mestas, A. Birer, and D. Cathignol, "A novel method to control P+/P− ratio of the shock wave pulses used in the extracorporeal piezoelectric lithotripsy (EPL)," Ultrasound Med. Biol. **16**(5), 473–488 (1990).
[23] J. Tavakkoli, A. Birer, A. Arefiev, F. Prat, J. Y. Chapelon, and D. Cathignol, "A piezocomposite shock wave generator with electronic focusing capability: Application for producing cavitation-induced lesions in rabbit liver," Ultrasound Med. Biol. **23**(1), 107–115 (1997).
[24] M. R. Bailey, D. T. Blackstock, R. O. Cleveland, and L. A. Crum, "Comparison of electrohydraulic lithotripters with rigid and pressure-release ellipsoidal reflectors. I. Acoustic fields," J. Acoust. Soc. Am. **104**, 2517–2524 (1998).

[25] M. R. Bailey, D. T. Blackstock, R. O. Cleveland, and L. A. Crum, ''Comparison of electrohydraulic lithotripters with rigid and pressure-release ellipsoidal reflectors. II. Cavitation fields,'' J. Acoust. Soc. Am. **106**, 1149–1160 (1999).
[26] M. Delius, W. Mueller, A. Goetz, H-G. Liebich, and W. Brendel, ''Biological effects of shock waves: Kidney hemorrhage in dogs at a fast shock wave administration rate of fifteen hertz,'' J. Lithotripsy Stone Dis. **2**(2), 103–110 (1990).
[27] M. R. Bailey, ''Control of acoustic cavitation with application in lithotripsy,'' Ph.D. thesis, Technical Report ARL-TR-97-1, Applied Research Laboratories, The University of Texas at Austin, 1997.
[28] F. E. Prieto and A. M. Loske, ''Bifocal reflector for electrohydraulic lithotripters,'' J. Endourol **13**(2), 65–75 (1999).
[29] P. Huber, J. Debus, K. Jochle, I. Simiantonakis, J. Jenne, R. Rastert, J. Spoo, W. J. Lorenz, and M. Wannenmache, ''Control of cavitation activity by different shockwave pulsing regimes,'' Phys. Med. Biol. **44**(6), 1427–1437 (1999).
[30] X. Xi and P. Zhong, ''Improvement of stone fragmentation during shock-wave lithotripsy using a combined EH/PEAA shock-wave generator-*in vitro* experiments,'' Ultrasound Med. Biol. **26**(3), 457–467 (2000).
[31] F. R. Gilmore, ''The growth or collapse of a spherical bubble in a viscous compressible liquid,'' Rep. No. 26-4, California Institute of Technology, 1952, pp. 1–40.
[32] R. O. Cleveland, O. A. Sapozhnikov, M. R. Bailey, and L. A. Crum, ''A dual passive cavitation detector for localized detection of lithotripsy-induced cavitation *in vitro*,'' J. Acoust. Soc. Am. **107**, 1745–1758 (2000).
[33] A. J. Coleman, M. J. Choi, J. E. Saunders, and T. G. Leighton, ''Acoustic emission and sonoluminescence due to cavitation at the beam focus of an electrohydraulic shock wave lithotripter,'' Ultrasound Med. Biol. **18**(3), 267–281 (1992).
[34] P. A. Thompson, *Compressible-Fluid Dynamics* (Rensselaer Polytechnic Institute, Troy, NY, 1988), p. 102.
[35] D. A. Sullivan, ''Historical review of real-fluid isentropic flow models,'' J. Fluids Eng. **103**, 258–267 (1981).
[36] A. Eller and H. G. Flynn, ''Rectified diffusion during nonlinear pulsations of cavitation bubbles,'' J. Acoust. Soc. Am. **37**, 493–503 (1965).
[37] V. A. Akulichev, in *High Intensity Ultrasonics Fields*, edited by L. D. Rozenberg (Plenum, New York, 1971), pp. 239–259.
[38] T. Watanabe and Y. Kukita, ''Translational and radial motions of a bubble in an acoustic standing wave field,'' Phys. Fluids A **5**(11), 2682–2688 (1993).
[39] L. A. Crum, ''Bjerknes forces on bubbles in a stationary sound field,'' J. Acoust. Soc. Am. **57**, 1363–1370 (1975).
[40] A. Eller and H. G. Flynn, ''The equilibrium and stability of a translating cavity in liquid,'' J. Fluid Mech. **40(4)**, 785–803 (1967).
[41] R. O. Cleveland, M. R. Bailey, N. Fineberg, B. Hartenbaum, M. Lokhandwalla, J. A. McAteer, and B. Sturtevant, ''Design and characterization of a research electrohydraulic lithotripter patterned after the Dornier HM3,'' Rev. Sci. Instrum. **71**(6), 2514–2525 (2000).
[42] A. J. Coleman, J. E. Saunders, R. C. Preston, and D. R. Bacon, ''Pressure waveforms generated by a Dornier extra-corporeal shock-wave lithotripter,'' Ultrasound Med. Biol. **13**(10), 651–657 (1987).
[43] P. Zhong, I. Cioanta, F. H. Cocks, and G. M. Preminger, ''Inertial cavitation and associated acoustic emission produced during electrohydraulic shock wave lithotripsy,'' J. Acoust. Soc. Am. **101**, 2940–2950 (1997).
[44] B. Sturtevant, ''Shock wave physics of lithotripters,'' in *Smith's Textbook of Endourology*, edited by A. Smith, G. H. Badlani, D. H. Bagley, R. V. Clayman, J. H. Jordan, L. R. Kavoussi, J. E. Lingeman, G. M. Preminger, and J. W. Segura (Quality Medical, St. Louis, 1996), pp. 529–552.
[45] A. J. Coleman, J. E. Saunders, L. A. Crum, and M. Dyson, ''Acoustic cavitation generated by an extracorporeal shockwave lithotripter,'' Ultrasound Med. Biol. **13(2)**, 69–76 (1987).
[46] A. Philipp, M. Delius, C. Scheffczyk, A. Vogel, and W. Lauterborn, ''Interaction of lithotripter-generated shock waves with air bubbles,'' J. Acoust. Soc. Am. **93**, 2496–2509 (1993).
[47] J. Staudenraus and W. Eisenmenger, ''Fiber-optic hydrophone for ultrasonic and shock-wave measurements in water,'' Ultrasonics **31**(4), 267–273 (1993).
[48] C-D. Ohl, R. Geisler, and W. Lauterborn, ''Effect of a rigid boundary on bubble dynamics in ESWL,'' presented at the 141st Meeting of the Acoustical Society of America, Chicago, IL (2001).
[49] D. L. Sokolov, M. R. Bailey, F. Pulvermaker, and L. A. Crum, ''Increased damage to stones without increased damage to cells with a dual-reflector lithotripter,'' presented at the 2000 IEEE International Ultrasonics Symposium, San Juan, Puerto Rico, 2000.
[50] D. L. Sokolov, M. R. Bailey, and L. A. Crum, ''Effect of dual-reflector lithotripter on stone fragmentation and cell damage,'' J. Acoust. Soc. Am. **108**, 2518(A) (2000).

Erratum: "The matched-lag filter: Detecting broadband multipath signals with auto- and cross-correlation functions" [J. Acoust. Soc. Am. 109, 1997–2007 (2001)]

John L. Spiesberger
Department of Earth and Environmental Science, 158 Hayden Hall, University of Pennsylvania, 240 South 33rd Street, Philadelphia, Pennsylvania 19104-6316

(Received 12 June 2001; accepted for publication 25 June 2001)

[DOI: 10.1121/1.1396327]

PACS numbers: 43.60.Cg, 43.60.Gk, 43.60.Dh, 43.10.Vx

The following corrections are necessary on p. 2004 concerning the proof that auto-correlation functions do not prohibit forbidden arrangements of signals. First, the three paths should have amplitudes of $\sqrt{8}$ each, instead of 8. Second, the sentence starting with "Next, we note that the expected..." should be replaced with the following two sentences. Next, we note that the expected value of two lag-functions, {7,6,3,0} and {5,0,1,2}, is $E[R_{11}(p)]=\{6,3,2,1\}$. However, {7,6,3,0} is a forbidden arrangement of auto-correlation function amplitudes for paths arriving at either set of times {1,2,3} or {2,3,4}, and {5,0,1,2} is forbidden because it is impossible to have signals simultaneously at lags 0, 2, and 3.

Therefore, since $E[R_{11}(p)]=\{6,3,2,1\}$ is a physically possible auto-correlation function, a specified auto-correlation function does not exclude forbidden arrangements of signals.

Erratum: "Measured capacitance of a condenser microphone as a function of diaphragm displacement" [J. Acoust. Soc. Am. 108, 2134–2144 (2000)]

Xinche Yan[a] and Malcolm J. Crocker
Department of Mechanical Engineering, Auburn University, Auburn, Alabama 36849

Li Jun Zeng
TRW Automotive, 12025 Tech Center Drive, Livonia, Michigan 48150

(Received 20 February 2001; accepted for publication 13 April 2001)

[DOI: 10.1121/1.1377288]

There should be a note for each theoretical curve and corresponding experimental data points in Fig. 5 as follows:

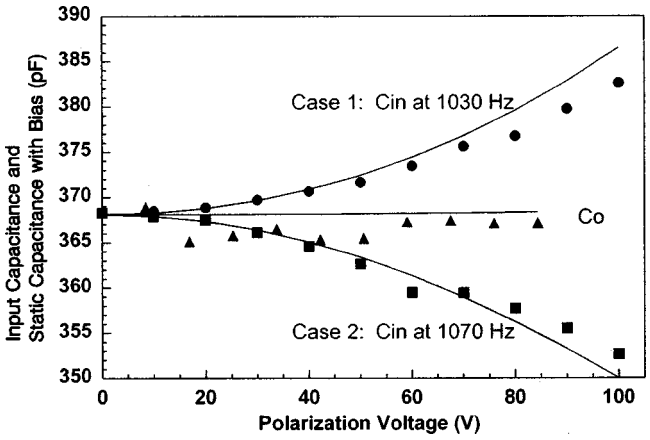


FIG. 5. Input capacitance and static capacitance with bias as a function of polarization voltage for zero air pressure. Solid lines—theoretical, data points—experimental.

[a]Present address: Knowles Electronics, LLC, 1151 Maplewood Drive, Itasca, IL 60143.

